



Molecular Mechanisms of MYC's impact on Transcription Elongation

Molekulare Mechanismen des Einflusses von MYC auf die Transkriptionselongation

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

Section Biomedicine

submitted by

Apoorva Baluapuri

from

Bhopal, India

Würzburg 2021



Submitted on:

Office stamp

Members of the Thesis Committee

Chairperson: Prof. Dr Markus Sauer

Primary Supervisor: Prof. Dr Elmar Wolf

Supervisor (Second): Prof. Dr Martin Eilers

Supervisor (Third): Prof. Dr Alexander Buchberger

Supervisor (Fourth): Prof. Dr Reinhard Lührmann

Date of Public Defence:

Date of Receipt of Certificates:

CONTENTS

SUMMARY	1
ZUSAMMENFASSUNG	2
1 INTRODUCTION	3
1.1 PHYSIOLOGICAL AND ONCOLOGICAL FUNCTIONS OF MYC	3
1.1.1 <i>Expression of MYC and its relation to Oncogenesis</i>	3
1.1.2 <i>The essential nature of MYC proteins</i>	5
1.1.3 <i>MYC as a therapeutic target</i>	5
1.1.4 <i>Cellular and animal models to study MYC functions</i>	7
1.2 MOLECULAR MECHANISMS OF MYC FUNCTION	9
1.2.1 <i>Structure of MYC protein</i>	10
1.2.2 <i>Target Genes of MYC proteins</i>	12
1.2.3 <i>MYC mediated gene regulation</i>	17
1.3 MECHANISMS OF TRANSCRIPTION REGULATION	21
1.3.1 <i>Molecular mechanisms of the transcription cycle</i>	21
1.3.2 <i>Transitions between the steps of the transcriptional cycle</i>	24
1.3.3 <i>Methods to study changes in transcription</i>	26
1.4 AIM OF THE STUDY	31
2 MATERIALS	32
2.1 CELL LINES	32
2.2 CONSUMABLES.....	32
2.3 MANUFACTURERS & DEVELOPERS	33
2.4 EQUIPMENT	34
2.5 CHEMICALS	35
2.6 MEDIA & SUPPLEMENT	36
2.7 KITS AND DETECTION SYSTEMS.....	36
2.8 PRIMARY ANTIBODIES	37
2.9 SECONDARY ANTIBODIES	37
2.10 SOLUTION AND BUFFERS.....	38
2.11 NUCLEIC ACIDS	40
2.12 BIOINFORMATIC TOOLS AND VERSIONS.....	41
3 METHODS	42
3.1 CELL CULTURE METHODS	42

3.1.1	<i>Cell Growth, splitting, freezing and thawing</i>	42
3.1.2	<i>siRNA transfection</i>	42
3.2	CELL BIOLOGY METHODS	43
3.2.1	<i>Indirect Immunofluorescence</i>	43
3.2.2	<i>Proximity Ligation Assay</i>	43
3.3	MOLECULAR BIOLOGY & BIOCHEMISTRY METHODS	44
3.3.1	<i>Determination of Protein Concentration</i>	44
3.3.2	<i>Immunoblotting</i>	44
3.3.3	<i>RNA Extraction and cDNA synthesis</i>	45
3.3.4	<i>Real Time Qualitative PCR (RT-qPCR)</i>	46
3.4	IMAGING METHODS	46
3.4.1	<i>Confocal Microscopy</i>	46
3.4.2	<i>Structured Illumination Microscopy</i>	46
3.4.3	<i>direct Stochastic Optical Reconstruction Microscopy</i>	47
3.4.4	<i>Image Processing</i>	48
3.5	NEXT GENERATION SEQUENCING BASED METHODS	49
3.5.1	<i>mRNA-Seq</i>	49
3.5.2	<i>4sU-Seq</i>	49
3.5.3	<i>DRB 4sU-Seq</i>	50
3.5.4	<i>Illumina NextSeq500 Sequencing</i>	50
3.6	BIOINFORMATIC METHODS	51
3.6.1	<i>References and databases</i>	51
3.6.2	<i>Read Alignment</i>	51
3.6.3	<i>Normalization Procedures</i>	52
3.6.4	<i>RNA Pol II Elongation Rate calculations</i>	52
3.6.5	<i>Density plots and heat maps</i>	53
3.6.6	<i>Read Density-based score calculations</i>	54
3.6.7	<i>Gene Differential Expression Analysis</i>	54
4	RESULTS	56
4.1	4sUDRB-SEQ REVEALS WIDE VARIATION IN ELONGATION RATES	56
4.2	CELLULAR SYSTEM TO STUDY PRIMARY EFFECTS OF MYC	60
4.3	ELONGATION RATES OF POL II ARE REGULATED BY MYC LEVELS	61
4.4	MYC IS REQUIRED FOR ASSEMBLY OF POL II ELONGATION COMPLEX	66
4.4.1	<i>MYC dependent recruitment of SPT5 to Pol II on chromatin</i>	66
4.5	MYC MEDIATED SPT5 HANDOVER IS DRIVEN BY CDK7 VIA TFIIE	68

4.5.1	<i>CDK7 is responsible for transfer of SPT5 from MYC to Pol II</i>	69
4.5.2	<i>TFIIE is responsible for CDK7 sensitivity</i>	71
4.6	POL II PROCESSIVITY AND DIRECTIONALITY IS REGULATED BY MYC LEVELS	74
4.6.1	<i>MYC levels regulate directionality of Pol II</i>	74
4.6.2	<i>Processivity of Pol II is a function of MYC levels</i>	76
4.6.3	<i>MYC enables productive transcription by RNA Pol II</i>	78
4.7	MYC REGULATES GENE EXPRESSION BY ALTERING SPEED, PROCESSIVITY AND DIRECTIONALITY OF POL II	80
4.7.1	<i>MYC mediated defects in elongation rate, processivity and directionality correlate with each other</i>	80
4.7.2	<i>Gene length stratifies for MYC's impact on Pol II behaviour</i>	82
4.7.3	<i>MYC drives the regulation of genes involved in UV response</i>	84
4.8	MYC SHAPES ITS TUMOUR SPECIFIC GENE EXPRESSION PROFILE BY SQUELCHING ELONGATION FACTORS AWAY FROM POL II.....	85
4.8.1	<i>High MYC levels reduce Pol II elongation rates and processivity</i>	85
4.8.2	<i>SPT5 and SPT6 are squelched away from Pol II</i>	88
4.8.3	<i>Squelching by MYC Establishes its Tumour Specific Gene Expression Profile</i>	90
4.9	MYC SHOWS CAPACITY TO FORM CONDENSATE <i>IN VIVO</i>	92
4.10	PROTEASOMAL INHIBITION DRIVES PHASE SEPARATION CAPACITY OF MYC.....	95
4.10.1	<i>MG132 treatment squelches MYC away from chromatin</i>	95
4.10.2	<i>MG132 treatment super-squelches SPT6 and SPT5 with MYC</i>	101
4.11	MG132 TREATMENT CHANGES MYC GENE REGULATION PROFILE.....	102
5	DISCUSSION	106
5.1	MYC AS A REGULATOR OF TRANSCRIPTION ELONGATION.....	106
5.2	ROLE OF OTHER MYC INTERACTORS IN GENE REGULATION.....	108
5.3	MOLECULAR MECHANISMS OF MYC FUNCTION AT CORE PROMOTERS.....	108
5.4	SQUELCHING AS A MECHANISM OF GENE REPRESSION	110
5.5	CONCLUSIONS	111
6	REFERENCES	114
7	APPENDIX	129
7.1	ACKNOWLEDGEMENTS.....	129
7.2	PUBLICATION LIST	131
7.3	CURRICULUM VITAE.....	133
7.4	AFFIDAVIT	134

LIST OF TABLES

TABLE 1.1 COMPARISON OF INHIBITORS USED IN STUDYING TRANSCRIPTION	26
TABLE 1.2 COMPARISON OF TECHNIQUES UTILIZED TO STUDY TRANSCRIPTION.....	27
TABLE 2.1 LIST OF CELL LINES USED.	32
TABLE 2.2 LIST OF CHEMICALS USED.....	35
TABLE 2.3 LIST OF MEDIA AND SUPPLEMENT USED.....	36
TABLE 2.4 LIST OF KITS AND DETECTION SYSTEM USED.....	36
TABLE 2.5 LIST OF PRIMARY ANTIBODIES USED.....	37
TABLE 2.6 LIST OF SECONDARY ANTIBODIES USED.....	37

LIST OF FIGURES

FIGURE 1.1 MYC BOXES AND THEIR FUNCTION.	10
FIGURE 4.1 DRB ASSAY FOR MEASURING ELONGATION RATES.....	57
FIGURE 4.2 4sUDRB-SEQ CAN BE USED TO CALCULATE POL II ELONGATION RATES.	58
FIGURE 4.3 4sUDRB-SEQ TRANSCRIPTION WAVE FRONTS ARE A FUNCTION OF RELEASE TIMES.....	59
FIGURE 4.4 ELONGATION RATES FROM 4sUDRB-SEQ.....	60
FIGURE 4.5 CELLULAR SYSTEM TO ESTABLISH ACUTE CHANGES IN MYC LEVELS.	60
FIGURE 4.6 MYC LEVELS REMAIN UNAFFECTED BY DRB TREATMENT.	62
FIGURE 4.7 MYC IS REQUIRED TO MAINTAIN FAST ELONGATION RATES.....	63
FIGURE 4.8 MYC IS REQUIRED GLOBALLY TO MAINTAIN FAST ELONGATION.....	64
FIGURE 4.9 CHANGES IN ELONGATION RATE BY MYC ARE INDEPENDENT OF RELEASE TIMES.....	65
FIGURE 4.10 POL II ELONGATION RATES ARE A FUNCTION OF MYC LEVELS.....	66
FIGURE 4.11 MYC DRIVEN RECRUITMENT OF SPT5 TO POL II.....	67
FIGURE 4.12 SPT5 AND P ^{SER2} POL II LEVELS ARE UNAFFECTED BY MYC.....	68
FIGURE 4.13 SPT5 AND MYC LEVELS ARE UNAFFECTED BY CDK INHIBITION.....	69
FIGURE 4.14 CDK7 IS RESPONSIBLE FOR TRANSFER OF SPT5 FROM MYC TO POL II.....	70
FIGURE 4.15 CDK7 DEPLETION PREVENTS SPT5 HANDOVER TO POL II BY MYC.....	71
FIGURE 4.16 DEPLETION OF HCE1 DOES NOT DECREASE SPT5-POL II PROXIMITY.....	72
FIGURE 4.17 CDK7 MEDIATED EVICTION OF TFIIE FROM POL II.....	73
FIGURE 4.18 TFIIE RELEASE CONVERTS POL II INTO A RECEPTIVE STATE FOR SPT5.....	73
FIGURE 4.19 MYC LEVELS REGULATE DIRECTIONALITY OF POL II.....	75
FIGURE 4.20 POL II LOSES DIRECTIONALITY GLOBALLY IN ABSENCE OF MYC.....	76
FIGURE 4.21 PROCESSIVITY OF POL II IS A FUNCTION OF MYC LEVELS.....	77
FIGURE 4.22 MYC ENABLES PRODUCTIVE TRANSCRIPTION BY RNA POL II.....	78
FIGURE 4.23 MYC MEDIATED EFFECTS ON POL II BEHAVIOUR CORRELATE AND AFFECT GENE REGULATION.....	80
FIGURE 4.24 GENE LENGTH CORRELATES WITH MYC-MEDIATED GENE REGULATION.....	83
FIGURE 4.25 MYC DRIVES THE REGULATION OF GENES INVOLVED IN UV RESPONSE.....	84
FIGURE 4.26 HIGH MYC CELLULAR SYSTEM.....	86
FIGURE 4.27 HIGH MYC LEVELS REDUCE POL II ELONGATION RATES AND PROCESSIVITY	87
FIGURE 4.28 GENE SPECIFIC CHANGES IN ELONGATION RATES AT HIGH MYC LEVELS.....	88

FIGURE 4.29 MYC AND SPT5 SHARE HIGHER PROXIMITY IN MYC HIGH CONDITION	89
FIGURE 4.30 SPT5 IS SQUELCHED AWAY FROM POL II AT HIGH MYC LEVELS	89
FIGURE 4.31 SPT6 IS SQUELCHED AWAY FROM POL II AT HIGH MYC LEVELS	90
FIGURE 4.32 MYC'S IMPACT ON POL II FUNCTION DRIVES GENE REGULATION IN CANCER	91
FIGURE 4.33 PSER5 POL II CONDENSATES CAN BE VISUALIZED BY <i>d</i> STORM	93
FIGURE 4.34 SINGLE MOLECULE LOCALIZATION OF MYC	94
FIGURE 4.35 SIM REVEALS MYC IS RESTRICTED TO DNA CONTAINING COMPARTMENT IN THE NUCLEUS.....	96
FIGURE 4.36 MYC IS SQUELCHED AWAY INTO DNA FREE COMPARTMENT UPON MG132 TREATMENT	97
FIGURE 4.37 MYC PARTIALLY RELOCATES TO NUCLEOLI UPON MG132 TREATMENT	98
FIGURE 4.38 MYC BEHAVIOUR IS INDEPENDENT OF LEVELS IN RESPONSE TO MG132 ...	99
FIGURE 4.39 MYC PARTIALLY RELOCATES TO PML BODIES UPON MG132 TREATMENT	100
FIGURE 4.40 MG132 DEPENDANT MYC CONDENSATES ARE SENSITIVE TO HEXANE 1,6 DIOL	101
FIGURE 4.41 SPT5 AND SPT6 ARE SQUELCHED INTO PHASE-SEPARATED COMPARTMENTS	102
FIGURE 4.42 MYC LEVELS UPON MG132 TREATMENT AND MYC OVEREXPRESSION ...	103
FIGURE 4.43 MG132 TREATMENT CHANGES MYC GENE REGULATION PROFILE.....	104
FIGURE 4.44 MYC ACTIVATED AND REPRESSED GENES ARE AFFECTED BY MG132	105
FIGURE 5.1 MODEL OF MYC'S EFFECT ON POL II TRANSCRIPTION.....	109

SUMMARY

Expression of the MYC oncoprotein, which binds the DNA at promoters of most transcribed genes, is controlled by growth factors in non-tumor cells, thus stimulating cell growth and proliferation.

Here in this thesis, it is shown that MYC interacts with SPT5, a subunit of the RNA polymerase II (Pol II) elongation factor DSIF. MYC recruits SPT5 to promoters of genes and is required for its association with Pol II. The transfer of SPT5 is mediated by CDK7 activity on TFIIE, which evicts it from Pol II and allows SPT5 to bind Pol II.

MYC is required for fast and processive transcription elongation, consistent with known functions of SPT5 in yeast. In addition, MYC increases the directionality of promoters by stimulating sense transcription and by suppressing the synthesis of antisense transcripts.

The results presented in this thesis suggest that MYC globally controls the productive assembly of Pol II with general elongation factors to form processive elongation complexes in response to growth-factor stimulation of non-tumour cells. However, MYC is found to be overexpressed in many tumours, and is required for their development and progression.

In this thesis it was found that, unexpectedly, such overexpression of MYC does not further enhance transcription but rather brings about squelching of SPT5. This reduces the processivity of Pol II on selected set of genes that are known to be repressed by MYC, leading to a decrease in growth-suppressive gene transcription and uncontrolled tumour growth.

ZUSAMMENFASSUNG

Die Expression des MYC-Onkoproteins, das die Promotoren der meisten exprimierten Gene bindet, wird in gesunden, nicht transformierten Zellen durch Wachstumsfaktoren reguliert und fördert das Zellwachstum und die Zellteilung.

In dieser Arbeit wurde die Interaktion zwischen MYC und SPT5, einer Untereinheit des RNA-Polymerase (Pol II) Elongationsfaktors DSIF gezeigt. MYC ist für die Rekrutierung von SPT5 an Promotoren und die Assoziation mit Pol II notwendig. Der Transfer von SPT5 auf Pol II setzt die Aktivität der Proteinkinase CDK7 voraus, die TFIIE aus dem Pol II Komplex entfernt und es so SPT5 ermöglicht, an Pol II zu binden.

MYC wird für eine schnelle und prozessive Transkriptionselongation benötigt, was mit bekannten Funktionen von SPT5 in Hefe übereinstimmt. Zusätzlich erhöht MYC die Direktionalität von Promotoren, indem es die Sense-Transkription stimuliert und die Synthese der Antisense-Transkripte unterdrückt.

Die in dieser Arbeit vorgestellten Ergebnisse legen nahe, dass MYC in normalen, nicht-transformierten Zellen die produktive Assemblierung von Pol II mit allgemeinen Elongationsfaktoren global steigert, um prozessive Elongationskomplexe als Reaktion auf die Wachstumsfaktorstimulation zu bilden. Die meisten humanen Tumore exprimieren jedoch deutlich erhöhte Mengen des MYC Proteins, das für Tumorentstehung und Progression benötigt wird.

In dieser Arbeit wurde festgestellt, dass eine solche Überexpression von MYC unerwarteterweise keine weitere Steigerung der Expression mit sich bringt, sondern zur Sequestrierung von SPT5 führt. Dies reduziert die Prozessivität von Pol II an ausgewählten Genen, welche durch MYC bekannterweise supprimiert werden, was zu einer Abnahme der wachstumsunterdrückenden Gentranskription und zu einem unkontrollierten Wachstum führt.

1 INTRODUCTION

In multicellular organisms, the DNA sequence remains constant throughout the somatic and germ cells. However, the messenger RNA (mRNA) transcribed from the DNA, which in turn gets translated to proteins, varies from cell to cell, time to time, and also differs between diseased and healthy cells. This regulation not only makes multicellular organism possess variability in cell type, but also gives them ability to respond to external cues from environment (like growth factors).

One way to bring about such variety of mRNA production is to regulate the behaviour of the key enzyme which is responsible. The molecular structure and behaviour of this enzyme, RNA Polymerase II (Pol II) is conserved (partially) across the metazoans and its distribution on chromatin changes in response to a variety of proteins, called transcription factors (TFs). Such factors could operate in many ways, but a key feature of these proteins is the ability to bind to DNA in a sequence specific manner. Among the many TFs which regulate cellular functions involved in cell growth and proliferation, is the widely studied onco-protein MYC. It brings about its action by altering the distribution and behaviour of Pol II, but the exact mechanisms by which this phenomenon is carried out, remains hitherto unknown.

1.1 Physiological and Oncological Functions of MYC

1.1.1 Expression of MYC and its relation to Oncogenesis

Tumour formation in humans and murine experimental systems is usually driven by deregulated protein function (Hanahan and Weinberg, 2011). This deregulation of protein function can be caused due to either a change in amino acid sequence of the proteins, or due to change in its levels. These proteins arise from genes that belong to a certain class of gene called oncogenes. One such example of an oncogene is KRAS. Like other members of this gene family, the protein from this gene is a GTPase and a key player in multiple signal transduction pathways. According to The Cancer Genome Atlas (TCGA), mutations in KRAS gene are found to be present in more than 40% of all patients in the

TCGA database (Cerami et al., 2012). These mutations, occurring from a single nucleotide substitution, are in turn responsible for change in single amino acid resulting eventually in a gain of function mutation of KRAS function.

In contrast, there are certain proto-oncogenes, whose oncogenic function is not resulting from changes in amino acid sequence, but from the levels of protein. One such example of such genes is the MYC family of genes, alterations of which are found in more than 10% of all patients in the TCGA database (Gao et al., 2013) and in 28% of pan cancer occurrences. In published reports, MYC amplifications in breast cancer were found in as many as 48% of cases and up to 78% in osteosarcoma (Schaub et al., 2018).

MYC proteins, which consist of c-MYC (from *MYC* gene), MYCN (from *MYCN* gene) and MYCL (from *MYCL* gene), are known to be expressed at higher levels in tumour compared to the normal tissue (McKeown and Bradner, 2014) and are one of the most highly amplified oncogenes in various cancers types (Dang, 2012). For example, genomic rearrangement events like chromosomal translocation which result in the juxtaposition of the open reading frame (ORF) sequence for MYC from chromosome 8 with sequences from IgH genes' enhancer elements in chromosome 14, are the key cause of Burkitt's Lymphoma (Hayday et al., 1984). Since these enhancer elements are specifically active in mature B-cells, the translocation of MYC to IgH locus drives high levels of MYC mRNA and protein in B-cells (ar-Rushdi et al., 1983) (Schmitz et al., 2014). Similar mechanism is the causative action behind tumour progression in multiple myeloma as well (Shou et al., 2000). Comparably, in case of medulloblastoma, the most common type of childhood brain cancer, the G3 subgroup which is characterized by comparatively poor prognosis, is known to be stratified by high MYC expression as well (Roussel and Robinson, 2013) (Vo et al., 2016).

However, amplification of the gene is just one of the mechanisms by which MYC levels are elevated. Increased levels of MYC family of proteins can, in fact, stem from multiple mechanisms. Since MYC is downstream of key pathways involved in cell growth and proliferation pathways, their deregulation often results in high levels of MYC in the resulting tumours. For example, defects in the WNT-APC pathway found in human colon carcinoma results in overexpression of MYC (He, 1998). MYC levels are also affected in leukaemia arising from deregulation of Notch signalling pathways (Palomero et al., 2006). Multiple E3 ligases and mutation in thereof have been implicated in causing high

MYC levels, thus explaining MYC's frequent overexpression without gene amplification in tumours (Farrell and Sears, 2014).

1.1.2 The essential nature of MYC proteins

Key role of MYC in life cycle of cells and organisms is not restricted to only tumour cells. In normal cells, MYC transfers environmental signals to cellular processes like proliferation, apoptosis, energy metabolism, and differentiation (Eilers and Eisenman, 2008). This essential nature of MYC proteins in development and growth has been shown multiple times using a multitude of experimental systems. In mice where the c-MYC gene has been replaced by null alleles using homologous recombination, the homozygous mutant mice did not complete the gestation cycle, had embryos with smaller size and the yolk sac cells did not show signs of primitive haematopoiesis (Trumpp et al., 2001). However, mice heterozygous for null allele escape embryonic lethality, but achieve body mass which is considerably smaller than their wild type litter mates (Trumpp et al., 2001) (Dubois et al., 2008). In contrast, when the markers of aging like fibrosis, bone density and cholesterol were studied in adult mice heterozygous for null allele, it was found that they show lesser signs of aging when compared to mice expressing wild type alleles (Hofmann et al., 2015), showing the differences in MYC function during embryonic development as compared to mature organisms. Surprisingly, proliferation assays on T-cells from these mice showed that MYC was not required for T-cell activation in response to antigenic stimulation, indicating that MYC's effect on mammalian body size was driven by change in cell number, and not cell size (Trumpp et al., 2001). Finally, in terms of its functions at ectopic levels in mature cells, MYC was also identified as one of four genes, called Yamanaka factors, other being *SOX2*, *OCT4*, and *KLF4*, that could reprogram differentiated cells to a pluripotent stem cell state (Takahashi and Yamanaka, 2006), indicating its key role in stemness and differentiation as well.

1.1.3 MYC as a therapeutic target

In the last 30 years, multiple studies have proven the dependency of tumour initiation and maintenance and MYC levels in many different experimental models (Dang, 2012). A striking example of tumour initiation or development originating from MYC overexpression has been shown in a mouse model for medulloblastoma (Kawauchi et al., 2012), in addition to several other models, like for leukaemia, hepatocellular carcinoma and mammary adenocarcinoma (Gabay et al., 2014).

On the other hand, using mouse models, it had also been shown that a reduction in MYC levels results in ablation of tumours. For example, in a mouse system for acute myeloid leukaemia where MYC was overexpressed in haematopoietic cells, it was found that the tumours established in this manner were actually reversible when MYC level were reduced by a Tet-off system (Felsher and Bishop, 1999) (see 1.1.4 for details). This has also been shown in a mouse model of colorectal cancer, where deletion of MYC in adult small intestines results in abolishment of tumour formation due to repression of the WNT pathway (Sansom et al., 2007).

This addiction of tumours to MYC can be appreciated further in mouse models, where MYC driven tumours can be reversed. Tamoxifen driven MYC activation in adult mice with homozygous alleles for MYC-ER chimeric protein (see 1.1.4 for details) results in transformation of adenomas to adenocarcinomas which can be reversed by subsequent deactivation of MYC (Kortlever et al., 2017). This reversible MYC switch has eventually been shown for other kinds of tumour, like pancreatic ductal adenocarcinoma as well. (Sodir et al., 2020).

A key point that emerges from the above studies is that since the tumours are dependent on MYC, the systemic suppression of MYC expression in an organism could be beneficial in terms of tumour regression, while having a low impact on key physiological functions of MYC. One of the most striking proof of this principle originates from experiments involving systemic conditional expression of a dominant-negative allele of MYC, called OmoMYC (Soucek et al., 2008). In these studies, rapid regression of lung adenocarcinomas and glioblastomas (Annibali et al., 2014) was induced but without impacting proliferation of normal tissues in a major way, at least at early time points. Such promising results have even given way to efforts which have raised OmoMYC from proof of principle to a possibly feasible anti-MYC therapy, by bypassing the requirement of germline engineering and delivering it directly to cells (Beaulieu et al., 2019).

Such studies, in addition to other, have raised the opportunity for a therapeutic potential against MYC driven tumours by targeting either levels of MYC or by inhibiting the DNA-binding capacity of MYC (Massó-Vallés et al., 2020). Several other possibilities have been put forward, not only to target MYC levels directly, but also by targeting its cofactors (see 1.2 for details) and specific dependencies in tumour cells (Wolf and Eilers, 2020).

However, successful application of these efforts into viable clinical drugs remains elusive to date. This impediment originates from the fact that the exact mechanisms by which MYC executes the essential role in a cell, are not fully understood. Hence, inhibition of its physiological function could have deleterious effects on the cell. In fact, this was already seen in mice with long term expression of OmoMYC, where the proliferative intestine tissue was severely affected (Soucek et al., 2008). Additionally, such inhibition of MYC could have extremely organ system specific results depending on the cellular concentration (Lorenzin et al., 2016), and in absence of an understanding of what drives this specificity mechanistically, it is difficult to design a fully viable therapy. This lack of knowledge on MYC's mechanism has been damping the efforts in development of better drugs which can target MYC driven tumours.

Hence, the need for a detailed understanding of how MYC carries out its molecular functions at different levels in a cell, is the underlying rationale behind the work described in this thesis.

1.1.4 Cellular and animal models to study MYC functions

The MYC phenotypes outlined in the earlier sections, both physiological and oncological are a function of its levels in cells (Levens, 2013). In order to dissect its molecular mechanism, many laboratories around the world have made consistent and continuous efforts to modulate MYC levels in a wide variety of systems. While judging the utility of a cellular system, two factors play a critical role. Firstly, the degree of change in MYC levels can impact the phenotype in a major way. Given that low levels of MYC are found in resting cells, while medium levels are indicative of proliferative cells as compared high levels as found in cancer cells, the exact and complete change of MYC levels is a critical requirement to study MYC function. Secondly, the degree of acuteness with which the MYC levels are altered is also important since over long time periods, the possibilities of resulting phenotype stemming from indirect effects of MYC level alteration cannot be positively ruled out.

In terms of the magnitude of change in MYC levels, a system that has proven to be of high utility in many studies involves the application of inducible tetracycline responsive-promoter element (TRE). The TRE promoter, which is the basis of tetracycline resistance in Gram-negative bacteria, can be stimulated either by reverse tetracycline-controlled transactivator (rtTA) in response to doxycycline (dox) ("Tet-On") treatment (Gossen et

al., 1995) or it can be suppressed using tetracycline-controlled transactivator (tTA) (“Tet-Off”) upon dox treatment (Gossen and Bujard, 1992). Mice in which Myc expression was placed under the control of tTA, expression of which was itself under the control of the IgH enhancer and the SR α promoter (E μ SR-tTA), dox treatment resulted in stark reduction of MYC levels with concomitant regression of the tumours which were associated with high MYC levels (Felsher and Bishop, 1999). Similarly, T-cell lymphoma^{MYC-Tet-Off} system has been used to connect the link between MYC and TGF β autocrine signalling in lymphoma formation and maintenance (van Riggelen et al., 2010). Additionally, Tet-on system has been applied in human osteosarcoma cell line resulting in U2OS^{MYC-Tet-On} system to study tumour specific gene expression profiles (Walz et al., 2014) and to account for these profiles by promoter affinities (Lorenzin et al., 2016).

In terms of the acuteness of MYC activation, a ground-breaking cellular system was developed more than 30 years ago which utilized conditional alleles of MYC by fusing the hormone-binding domain of the human oestrogen receptor gene to the human MYC (MYC-ER) (Eilers et al., 1989). In this system, the cells harbour a MYC-ER chimeric protein that can be activated by 4-hydroxytamoxifen (OHT), which in turn shuttles the protein from cytoplasm to nucleus via estradiol binding. Subsequently, this approach was utilized to study MYC function in transgenic mice as well (Lawlor et al., 2006) where a reversibly switchable model of tumour formation and regression was used. In fact, application of this system in neuroblastoma cell lines for MYCN has already been useful to study the molecular mechanisms impacted by MYCN since it generates a gene expression profile that resembles *MYCN* amplified neuroblastoma (Herold et al., 2019).

While the Tet-on/off systems offer the advantage of drastic overexpression and depletion of MYC protein in a cellular system context, it could take many hours before the effect of such changes can be observed, whereas the ER system takes relatively short time to manifest the effects. However, the Tet-on/off system imparts a tight control on MYC expression, while the MYC-ER system is known to allow activation without OHT as well (Sohal et al., 2001).

Additionally, cellular systems offer the advantage of being amenable to manipulation by siRNA and shRNA as well. These methods can work very effectively (Lorenzin et al., 2016) (Popov et al., 2007), but the depletion can be observed only after time period of several hours.

Cellular systems in which MYC levels can be altered within short time offer the advantage that the effects can be interpreted as directly MYC driven without any caveats. Extraordinarily rapid kinetics has been shown while using the auxin-inducible degron (AID) system (Nishimura et al., 2009) to degrade AID-tagged MYC protein in K562 and HCT116 cell lines within less than an hour, thus showing selective MYC dependent amplification of genes involved in biosynthetic pathways (Muhar et al., 2018).

However, in spite of the extensive utility of the above-mentioned systems, they suffer from a number of limitations. Firstly, acute systems like AID can only be used to deplete MYC, and cannot be used to study the effect of its overexpression. Secondly, the same cellular system cannot be used to study MYC functions at low, medium and high levels in tandem. Only few studies have utilized a cellular system to study the phenotypes without MYC while comparing it to normal cells and to cells with high MYC levels. Hence, the resulting gap in our understanding of molecular mechanisms of MYC functions across different levels is still conspicuous.

Here, in this work, it is attempted to bridge this gap by using a combination of siRNA mediated depletion and Tet-on mediated overexpression (see 4.2) of MYC to modulate its concentrations in U2OS cells over two orders of magnitude.

1.2 Molecular Mechanisms of MYC function

MYC belongs to the basic helix-loop-helix family of transcription factors which regulate transcription by binding to DNA at specific sites (Prendergast et al., 1991). Like all transcription factors, MYC contains a trans-activation domain and DNA binding domain. The N-terminus of MYC has been implicated in trans-activation and trans-repression of many target genes, and thus termed as Transregulatory Domain. The amino acid sequence of MYC protein indicates the presence of a dimerization interface comprising of basic helix-loop-helix leucine zipper (bHLH-LZ) domain at the C-terminus, which binds to DNA. Compared to this dimerization domain, the overall sequence homology of the N-terminal is not particularly high among the MYC proteins (MYCN, CMYC and MYCL), but it does include short stretches of amino acids that exhibit remarkable sequence similarity (Fig 1.1).

Introduction

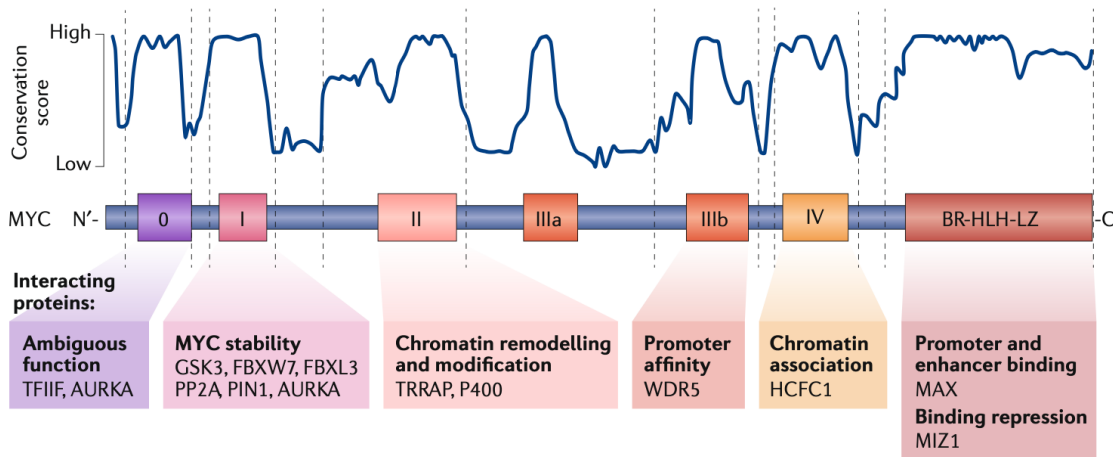


Figure 1.1 MYC boxes and their function.

Diagram showing conserved C-terminus domain responsible for DNA binding along with small stretches of amino acids called the MYC boxes which show a high degree of conservation in the N-terminus. Function definitions of MYC boxes 0–IV are based on deletions and/or point mutations. Examples of proteins that interact with the relevant MYC boxes are indicated. AURKA, Aurora kinase A; BR, basic region; FBXL3, F-box and leucine rich repeat protein 3; FBXW7, F-box and WD repeat domain containing 7; GSK3, glycogen synthase kinase 3; HCFC1, host cell factor C1; HLH, helix-loop-helix; LZ, leucine zipper; MAX, MYC associated factor X; MIZ1, MYC interacting zing finger protein 1; P400, E1A binding protein p400; PIN1, peptidylprolyl cis/trans isomerase NIMA-interaction 1; PP2A, serine/threonine protein phosphatase 2A; TFIIF, general transcription factor IIF subunit 1; TRRAP, transformation/transcription domain-associated protein; WDR5, WD repeat domain 5. This figure was published in similar form in (Balupuri et al., 2020).

These stretches of conserved amino acid sequences, called MYC Boxes, are six in number and some of the functions and interacting proteins of the MYC boxes are well described (Fig 1.1). Multiple studies over the last few decades have resulted in a detailed understanding of the role of MYC's protein structure towards its functions.

1.2.1 Structure of MYC protein

1.2.1.1 DNA-binding dimerization domain

The bHLH-LZ domain confers specificity to dimer partner via the helix-loop-helix-leucine zipper (HLHZ) region and to DNA via interactions between the basic (b) region and the major groove of the double helix. Another member of the same family of transcription factors, MAX, serves as an obligate dimerization partner of MYC, and MYC monomers have not been found. MYC's DNA binding property, thus relies upon heterodimerization of the MYC and MAX bHLH-LZ regions prior to sequence-specific DNA binding (Blackwood and Eisenman, 1991). Such heterodimers bind to a core hexanucleotide element (5'-CACGTG-3') (Blackwell et al., 1993) (Solomon et al., 1993), termed the Enhancer box (E-box). Cocystal structures of such heterodimers on their E-

box targets have revealed how they can mediate specific, high-affinity DNA binding, while indicating that leucine zipper residues specify MYC's heterodimerization with MAX (Nair and Burley, 2003).

However, several variations of the above-mentioned interactions and DNA binding exist. MYC and MAX are just two of the members of a vast network of interacting transcription factors (Conacci-Sorrell et al., 2014). In cells with oncogenic MYC levels, the zinc finger protein MIZ1 (Myc-interacting zinc finger protein 1) dimerizes with MYC and is recruited to the MYC target sites on DNA, thereby causing repression of such genes (Peukert et al., 1997). However, mechanisms of formation of MYC, MAX and MIZ1 complex formation on promoters is not clear yet (Wiese et al., 2013). The E-box DNA sequence is enriched in MYC binding sites on chromatin, but many of the binding sites show non-consensus from this sequence suggesting that other variables regulate chromatin binding of MYC as well (see 1.2.2) (Guo et al., 2014) (Lorenzin et al., 2016).

1.2.1.2 Transregulatory Domain

The N-terminus of MYC proteins consist of long regions of low complexity, which are interrupted by stretches of high conservation, called MYC Boxes (MB), and the current consensus defines six of them (MB0, I, II, IIIa, IIIb and IV)(Meyer and Penn, 2008). Some of the functions of MYC proteins are tightly linked to these MYC boxes. For example, while MBI regulated MYC stability by its interactions of various E3-ligases and kinases (Farrell and Sears, 2014), at the same time it also antagonizes function of FBXW7 via binding to Aurora Kinase A (AURKA) (Dauch et al., 2016).

MBII interacts with proteins involved in chromatin remodelling and histone acetylation such as TRRAP (transformation-transcription domain-associated protein) (McMahon et al., 2000) (McMahon et al., 1998), NuA4 Complex and P400 helicase protein (Kalkat et al., 2018). MBIIIb regulates histone H3Lys4 (H3K4) methylation to make active promoters accessible for MYC by interacting with WDR5 (Thomas et al., 2015). Such examples are just a representative set of interactions which highlight the variety of roles MYC fulfils through the MYC boxes.

However, while considering the role of the MYC boxes in MYC functions, it must be noted that MYC proteins are actually intrinsically disordered except for the bHLHZ domain, and resolve into structured form only when complexed with other proteins.

Evidence for such complexes has been in abundance. For example, crystallographic structures have been resolved for complexes of MYC with WDR5 (Thomas et al., 2015), TBP (Wei et al., 2019) and MYCN with AURKA (Richards et al., 2016).

In contrast to the detailed understanding of MYC boxes and the functions they execute, the role of the regions which are not well conserved among MYC proteins (Fig 1.1) in between the MYC boxes has been of rather limited nature.

1.2.1.3 Disordered Regions of MYC Proteins

Proteins which have disordered regions in their structure have been shown to separate into membrane less regions in the cell in a dynamic fashion (Chong et al., 2018), where they get separated from rest of the adjoining compartment due to the difference in the regions' physical properties (Brangwynne et al., 2009). This phenomenon is termed phase separation of proteins, and its role in gene regulation was suggested for the first time in context of single super enhancer locus being able to regulate transcription of many genes at the same time in proximity of transcription factors (Sabari et al., 2018).

The same concept has now been extended to single transcription factors (TFs) activating genes directly through the phase-separation capacity of their activation domains (Boija et al., 2018). Even though MYC proteins have large regions within MYC boxes which are disordered, the evidence and mechanisms for MYC participating in phase separation is not yet clear. Studies of unperturbed phase separation of MYC with other TFs are restricted only to *in vitro* experiments (Boija et al., 2018), and its relevance in cellular context is not understood.

MYC proteins utilize a wide variety of mechanisms in order to fulfil their role as a transcription factor and oncoprotein. An obvious question that arises in light of this variety is which of these mechanisms can be rate limiting for gene regulation by MYC. In the last few years, immense body of work has come up which describes the impact of MYC on transcription of its target genes.

1.2.2 Target Genes of MYC proteins

As mentioned above, MYC shows enhanced affinity to E-boxes *in vitro* and consequently, initial studies restricted their work on MYC binding to genes which had E-boxes at the binding sites. However, the first genome wide studies showed that surprisingly, MYC can

bind indiscriminately to the majority of promoters (Kim et al., 2008), many of which lacked E-boxes (Fernandez, 2003) (Guo et al., 2014) (Lin et al., 2012) (Nie et al., 2012) (Sabò et al., 2014) (Walz et al., 2014). Even though MYC occupies promiscuously almost all open promoters, its binding largely correlates with expression of genes and trimethylated H3K4 (Guccione et al., 2006) (Walz et al., 2014).

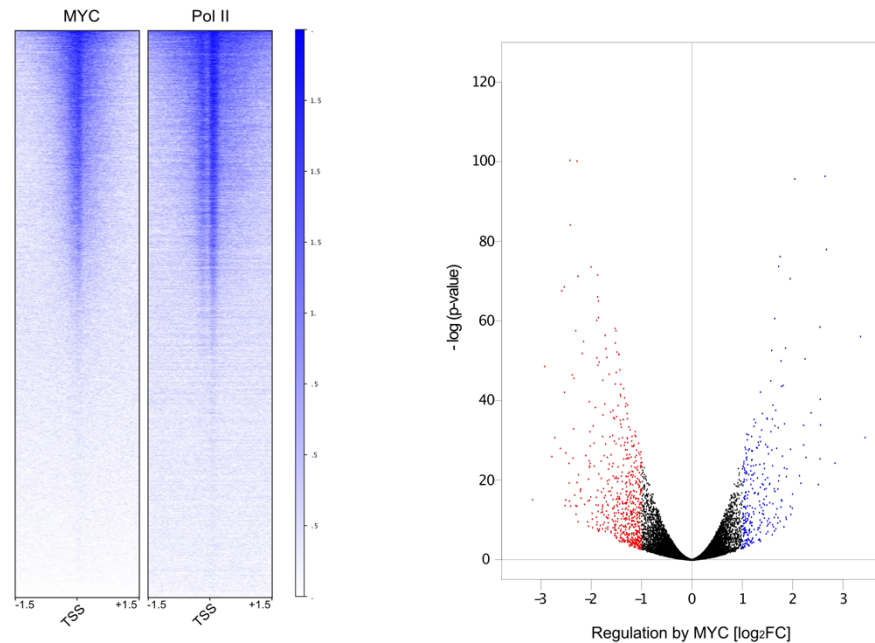


Figure 1.2 Target genes of MYC

Heatmaps (left) of all genes in human genome showing MYC and Pol II binding. Volcano plot (right) showing up-(blue) and down-regulated (red) genes by MYC. Data used from Walz et al., 2014. Part of this figure was published in similar form in (Balupuri et al., 2020).

While the characterisation of the genes that are bound by MYC (i.e., target genes) and their correlation with expression has been established unequivocally, the consensus on exact set of genes regulated by MYC remains a topic of intense debate, and is highly experimental system and level dependent (Fig 1.2). For example, in U2OS^{MYC-Tet-On} system (Walz et al., 2014), MYC is shown to bind to 20,014 sites, of which 8,401 are within 1.5 kb of transcription start sites (TSS). When MYC levels are increased in this system upon addition of Dox, the number of MYC binding sites increases to 45,645, of which 14,926 are close to promoters, representing almost all actively transcribed genes in the genome. However, of these genes, 462 genes were activated, and 896 genes were repressed by MYC, of which only 220 upregulated and 256 downregulated genes were direct target genes (Fig 1.2). This brings to light the paradox that exists between the

seemingly lack of correlation between MYC binding and gene-regulation, which can be explained in terms of various models.

1.2.2.1 The specific gene regulation model

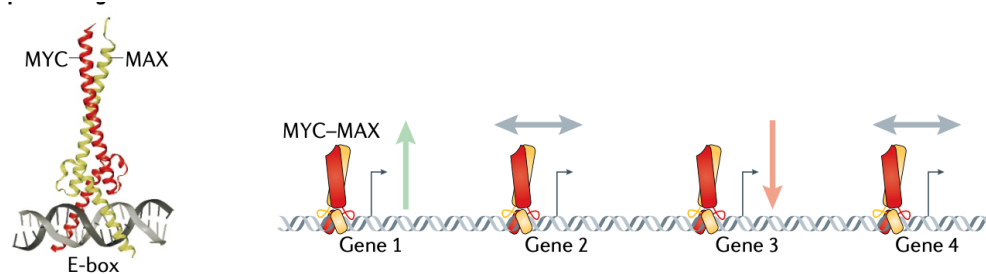


Figure 1.3 The specific gene regulation model

MYC-MAX dimer's structural characterization of enhancer-box (E-box) binding define MYC-induced target genes. Green arrows indicate activation, red arrow indicates repression and grey arrows indicate no regulation of genes. This figure was published in similar form in (Balupuri et al., 2020).

The paradox produced above by comparison of the gene expression data with the chromatin binding data of MYC proteins (Fig 1.2) is not only limited to a single experimental system. Same concept has been shown in B-cell lymphomagenesis in mice, in cultured B cells and fibroblasts (Sabò et al., 2014). Additionally, another study carried out in B cells showed major discrepancy in the number of MYC bound genes and MYC regulated genes (Tesi et al., 2019). Thus, the conclusion from such studies would suggest that MYC binds to all promoters and functions as a specific regulator of genes (Fig 1.3). While it is activating some genes, it is repressing another set of genes. Therefore detection of MYC binding on chromatin should not be equated to a productive transcriptional event. According to this model, the amplification of RNA content in the cells (see 1.2.2.2) can simply be explained by the fact that MYC regulates genes which belong to pathways that drive forward global RNA production (Kress et al., 2015).

1.2.2.2 The global gene activation model

In order to probe the physiological function of MYC, studies have overlaid genome-wide patterns of MYC recruitment, Pol II binding and chromatin modifications to show that MYC does not specifically activate or repress genes of a particular transcriptional programs, but universally amplifies production of mRNA on all genes. For example, when primary B- or T-splenocytes from mouse expressing MYC-eGFP were stimulated with lipopolysaccharide (LPS) or concanavalin A respectively, the cells showed increased eGFP-fluorescence. Moreover, when resting cells were compared with activated cells for

patterns of MYC binding and gene regulation, it was found that mRNA output of genes directly correlated with MYC binding at all active genes (Nie et al., 2012). This model has subsequently been called the ‘global amplifier’ model of MYC function (Fig 1.4), and it suggests that MYC recruitment to all genes is productive since it brings about an increase in the rate of overall transcription, and that, is the key mechanism of MYC function. The very principle was also shown in a B cell system of Burkitt’s lymphoma where the cells harbour a Tet-off MYC transgene, which when taken off tetracycline, show increase in MYC levels from 13,000 to 362,000 molecules per cell after 24 hrs. This increase in MYC levels was found to show a corresponding increase in mRNA production and an associated increase in phosphorylated form of RNA Pol II, thus indicating global amplification in transcription (Lin et al., 2012).

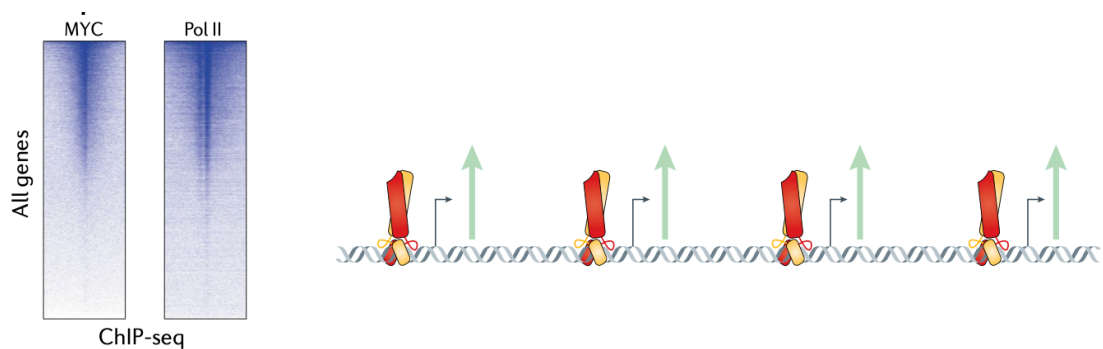


Figure 1.4 The global gene activation model

Genome-wide chromatin immunoprecipitation sequencing (ChIP-seq) studies of MYC identified that MYC binds at the promoters of most RNA polymerase II (Pol II)-bound and expressed genes. Accordingly, MYC induces a global increase in cellular mRNA levels in some biological systems. Green arrows indicate activation. This figure was published in similar form in (Balupuri et al., 2020).

In this model, the gene repression can be explained by either the argument that “repressed” genes are just less activated genes compared to average level of activation, or by invoking indirect mode of regulation enforced by transcriptional or chromatin repressors recruited to the target genes.

The discrepancy with the specific gene regulation reports discussed earlier can be resolved by the fact that in absence of spiked in RNA, the algorithms used to calculate the differential expression obfuscate the global effects. Even when a spike in RNA is used, the internal normalization procedures of such algorithms can again offset the changes in global increase of RNA levels (Lovén et al., 2012). Such explanations seem to be in line with the fact that tumours driven by MYC proteins can be made to regress

by global inhibition of transcription using chemicals that restrict Pol II activity via cyclin dependant kinases (CDKs) (Huang et al., 2014) (Chipumuro et al., 2014). At the same time, it must be noted that such changes in gene expression are highly dependent on cell culture conditions (Lewis et al., 2018).

1.2.2.3 The gene-specific affinity model

At the first glance, the above-mentioned models seem to be in strong contradiction, but the results on which these models are based on, can be accommodated to provide a simple explanation for global binding and specific gene regulation by MYC (Lorenzin et al., 2016). Surprisingly, binding of MYC/MAX to promoters requires additional proteins, like WDR5, which stabilize binding to DNA sequence, even at E-boxes. Additionally, when increasing MYC levels over a broad range of concentrations, the binding changes at promoters are not a linear function of the concentration. At low concentrations, MYC occupies primarily target genes with DNA sequence that have high affinity for MYC (“High Affinity Target Genes”), and this binding is not inconsequential. In fact, this set of genes consists of the ones that drive growth of the cell, and this concept is in alignment with gene-specific model, which states that global changes in RNA levels occur indirectly as a result of MYC-driven cell growth. When the levels of MYC increase in a cell, its binding on these set of genes does not increase any more, and consequently, expression of such genes may not change upon increasing MYC since the promoters are already fully occupied. This holds true, for instance, in case of highly transcribed genes involved in growth (e.g., ribosomal protein genes and genes involved in ribosome biogenesis and translation). However, for a different set of genes, MYC invades the promoters which contain DNA sequence with low affinity for MYC, and as a result, these genes get regulated only at high concentrations of MYC (“Low Affinity Target Genes”). Regulation of such genes, for example the ones involved in extra cellular matrix interaction, can be thus used to explain the oncogenic function of MYC. Recently, this model has also been extended to enhancer invasion by MYCN to explain the transcriptional amplification in neuroblastoma (Zeid et al., 2018) as well.

Thus, the gene-specific affinity model (Fig 1.5) does not require invoking the concept of unproductive MYC binding. In other words, changes in MYC binding govern the changes in gene expression, like how it was proposed in global amplification model. Further, gene repression which occurs at high levels can be partly explained by MYC/MIZ1 complexes which are known to form at high levels, but it is unclear, if this is the only repressive

mechanism. For examples, the mechanisms of repression for genes which do not show MIZ1 binding, has not been clarified in complete detail.

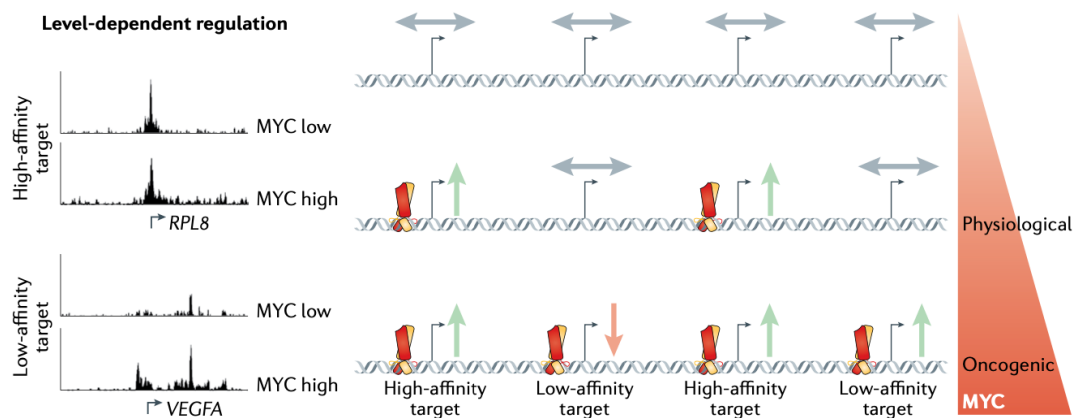


Figure 1.5 The gene-specific affinity model

Genes with high promoter–MYC affinity are bound and upregulated by MYC at physiological levels, but are not further induced at oncogenic levels as binding is already saturated. Instead, oncogenic MYC upregulates low-affinity target genes. The affinity model can reconcile the seemingly opposing specific gene and global gene models. Green arrows indicate activation, red arrow indicates repression and grey arrows indicate no regulation of genes. This figure was published in similar form in (Balupuri et al., 2020).

One key aspect, that holds true universally for the above models, is that even though the studies that have resulted in these models have enhanced our understanding of MYC biology, none of them explain the precise nature and molecular mechanisms which govern MYC mediated gene regulation.

1.2.3 MYC mediated gene regulation

In contrast to the exhaustive and precise cataloguing of target genes of MYC and the enormous efforts in finding transcriptional signatures that are indicative of MYC’s physiological and oncogenic role, it is still not known precisely how MYC regulates Pol II behaviour to dictate the changes in RNA levels. Studies have not only shown global redistribution of Pol II on chromatin as a result on MYC inhibition (Rahl et al., 2010), but also shown that MYC could modulate Pol II behaviour at multiple steps of transcriptional cycle (de Pretis et al., 2017). Additionally, MYC has been reported to interact physically with multitude of Pol II modulators, which could in part explain the mechanism behind MYC’s gene regulation capacity. Such modulators can be studied in two categories. The first category consists of proteins that do not directly modulate Pol II behaviour, but modify histone proteins and remodel chromatin to allow alterations in the transcriptional landscape of the cell. Second category, on the other hand, consists of

proteins that cause direct alterations in Pol II behaviour by either binding to Pol II or by causing post translational modifications, thus changing its distribution and kinetic behaviour on genes.

1.2.3.1 Indirect Pol II Modulators

Pol II behaviour can be modulated indirectly by changing its accessibility to the transcribed regions in the genome. This can be implemented by either modifying the histones on chromatin, or by remodelling the chromatin itself. Among the proteins that are known to interact with MYC, many of them are the enzymes which have histone acetyltransferases (HAT) activity, like GCN5 (McMahon et al., 2000) and TIP60 (Frank et al., 2003) which modify histones 3 and 4, respectively. Additionally, MYC interacts with the acetyltransferases p300 (Faiola et al., 2005), CBP (CREB binding protein) (Vervoorts et al., 2003) and the SWI/SNF subunit BAF47/SNF5 (Cheng et al., 1999) to modulate the epigenetic landscape of the cell. Yet another protein which modifies histones and interacts with MYC is WD40 domain protein WDR5, which not only brings about the assembly of histone modifier complexes but aids in the recruitment of MYC/MAX to chromatin in permissive chromatin environment (Thomas et al., 2015).

On the other hand, MYC interacts with protein having chromatin remodelling capacity as well. Also, these proteins themselves are known to interact with HAT activity proteins. For example, via MB II, MYC interacts with transformation/transcription domain-associated protein (TRRAP) (McMahon et al., 1998), which in turn is a scaffold protein for other complexes like as the transcription-promoting histone acetyltransferase complex NuA4 (Zhang et al., 2014). More recently, this was also shown using high throughput proteomics approach as well, that MBII interacts with HAT complexes like TRRAP and Tip60, and drives MYC driven tumour initiation (Kalkat et al., 2018). TRRAP is found to interact with proteins with yet other kinds of enzymatic activity. An example in this regard includes the helicase P400 E1A-binding protein which is devoid of HAT activity (Fuchs et al., 2001).

1.2.3.2 Direct Pol II Modulators

Regulation of the transcriptional process can be carried out not only by modifying the chromatin and histone proteins, but also by executing post translational modification of the Pol II molecule itself and/or of the proteins that directly bind to it. For example, CDK7, which is part of the holoenzyme TFIIH, is reported to interact with MYC

(Bouchard et al., 2004) (Cowling and Cole, 2007), showing that MYC could affect phosphorylation of Pol II CTD and other transcription factors like TFIIE (see section 1.3). Yet another modulator of Pol II activity is CDK9, which is a component of positive elongation factor (p-TEFb) (Marshall and Price, 1992).

Highlighting the principle that MYC could regulate several steps in the transcription cycle, it has been shown that the p-TEFb, in addition to CDK7, can be recruited by MYC at single gene level (Bouchard et al., 2004) (Eberhardy and Farnham, 2002). This interaction has been reported multiple times (Kanazawa et al., 2003) as an explanation of MYC's role in cell proliferation and differentiation, but one of the most exhaustive work which highlighted MYC's role in embryonic stem cells of mice was using chromatin binding redistribution of Pol II (Rahl et al., 2010) where the interaction of MYC with CDK9 was reported to have rate limiting role in regulation of transcription (Price, 2010). Although this redistribution of Pol II has been reported to be driven by MYCN as well (Büchel et al., 2017) (Herold et al., 2019), conclusive evidence that this behavior is solely via CDK9 interaction remains elusive.

However, when the proteins in the proximity of MYC were biotinylated by a biotin ligase fused to it, and subjected to mass spectrometry following streptavidin affinity capture, it could be shown that MB0 mediates interactions with CDK9 via direct binding to the general transcription factor TFIIF, and this interaction was indeed critical for tumour growth (Kalkat et al., 2018). Given that the mediator complex is also found to interact with MYC (Adhikary and Eilers, 2005), and is also required for enabling sense and antisense transcription in conjunction with TFIIF (Bernecky and Taatjes, 2012), Pol II modulators certainly offer strong candidature in being MYC's key interacting partner(s) to drive its functions in the cell.

Over the last few decades, multiple reports have shown that Pol II behaviour is actually dependent on proteins that directly interact with Pol II and can act as negative and positive regulators of transcription (Marshall and Price, 1992). It is also worthy to note that not only the pause-release of Pol II during transcription cycle, but also the steps preceding and following it (Kouzine et al., 2013) also seems to regulate the total output of the process (Guenther et al., 2007) (Gilchrist et al., 2008). The interaction of CDK9 with MYC alone doesn't clarify, which of the step(s) in the transcription cycle is/are rate-limiting in nature in context of MYC. Additionally, using pause-release itself as a proxy

for elongation, and associating it as rate-limiting with respect to transcription regulation by MYC doesn't describe the effects of MYC on Pol II in full detail.

It remains entirely open that MYC could affect the step of transcription cycle in which Pol II spends most of its time in the transcription cycle, i.e., productive elongation. This brings focus on the role of those Pol II modulators, which regulate its behaviour by directly attaching to specific binding sites (Cramer, 2019a) on the Pol II holoenzyme during the elongation phase of transcription cycle specifically (Noe Gonzalez et al., 2020).

1.2.3.3 Handover mechanism for regulation of transcription by MYC

Interestingly, among the proteins that directly interact with Pol II and regulate its behaviour on chromatin, MYC seems to be interacting with proteins which assist Pol II in elongation of nascent RNA. One such example is the PAF complex (Jaenicke et al., 2016), which binds to Pol II and aids in transcription by stabilizing the elongation complex of Pol II (Vos et al., 2018a). Furthermore, it has been shown that MYC itself is recruited to chromatin in a PAF dependent manner as well (Gerlach et al., 2017).

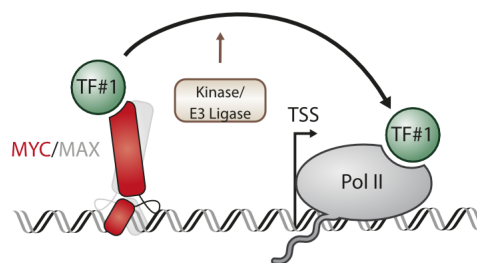


Figure 1.6 Handover mechanism for regulation of transcription by MYC

Possible elongation factors (TF#1) handed over to Pol II by MYC. The handover can happen by help of either an E3 ligase or CDK.

The fact that association of MYC with individual elongation factors like PAF complex (Jaenicke et al., 2016) can occur, indicates the possibility of a handover mechanism (Fig 1.6) where MYC binds to the promoters and hands over associated proteins to Pol II in order to make Pol II processive.

However, the details explaining this suggested mechanism remain incomplete in several aspects. Firstly, it has not been shown if this handover actually occurs with PAF complex, nor has it been shown that in absence of this handover, Pol II behaviour is affected.

Secondly, mechanistic details on how CDKs mediate this handover were not studied either. In fact, interaction of MYC with PAF complex was found to be E3-ligase dependent, since it was enhanced in presence of proteasomal inhibitor (Jaenicke et al., 2016). Thirdly, it was not explored if the interaction of MYC with PAF complex drives the function of MYC in terms of gene regulation. Finally, since there hasn't been a systemic study exploring the handover, it's not known if PAF complex is only protein which is handed over from MYC to Pol II or several proteins undergo this handover.

Thus, in the current study, to fill the gaps mentioned above, it has been attempted to find the protein(s) that are handed over to Pol II via MYC, and the dependency on CDKs for this handover is further explored. Additionally, attempts have also been made to find which is the rate limiting step on transcription in absence of this handover.

1.3 Mechanisms of Transcription Regulation

1.3.1 Molecular mechanisms of the transcription cycle

During the process of transcription, Pol II undergoes transition from one stage to next by virtue of interaction with various transcription factors and kinases. Classically, *in vitro* biochemical assays have revealed plethora of the proteins that bind to Pol II along each step of the transcription cycle. Also, recent advances in cryo-electron microscopy have offered clear insights into the structural mechanisms of regulated transcription initiation and elongation (Cramer, 2019a), which can be studied in the order of occurrence.

Firstly, the open chromatin offers Pol II possibility to scan the promoters (Dyran and Tjian, 1983) (McSwiggen et al., 2019) (Kadonaga, 2019) and some initiation factors, like TATA binding protein (TBP) bind to chromatin (Goodrich and Tjian, 1994) in order to stabilize the melted DNA double strand at promoters, which allows binding of Pol II to chromatin, and this binding is modulated on the basis of interaction with set of general transcription factors of Pol II (TFII). The first of such factors to bind Pol II during initiation and stabilize it on chromatin is TFIID.

Secondly, during initiation of transcription, Pol II dissociates with TFIID and associates with TFIIIE (Conaway et al., 1991) (Fig 1.7). Next, the Pol II transcribes 30-300 bp from the transcription start site (TSS), before it associates with DSIF (DRB sensitivity inducing

factor, consisting of SPT5 & SPT4) and NELF (Negative Elongation Factor) causing it to undergo pausing (Price et al., 1989).

Next, a variety of elongation factors like TFIIS, ELL and PAF (Shilatifard et al., 2003) (Chen et al., 2015) (Conaway and Conaway, 2019) get associated with Pol II in a stepwise manner, thus eventually modulating Pol II activity further while ensuring processive movement of Pol II to transcribe full length nascent mRNA. The factors involved in early stage, like DSIF/NELF have capacity to modulate the factors in later stages of transcription, like TFIIS, as well (Palangat et al., 2005).

Elongation is also assisted by proteins like SPT6 and FACT which act as chromatin modulators at the same time (Bortvin and Winston, 1996) (Orphanides et al., 1998) (Jeronimo and Robert, 2014). This processive movement continues till Pol II encounters a poly-adenylation site in genome, which when transcribed attracts multiple termination factors to the nascent RNA (Proudfoot, 2016) and results in its cleavage from transcribing Pol II, which in turn slows down considerably before being evicted from chromatin by XRN2 (Cortazar et al., 2019).

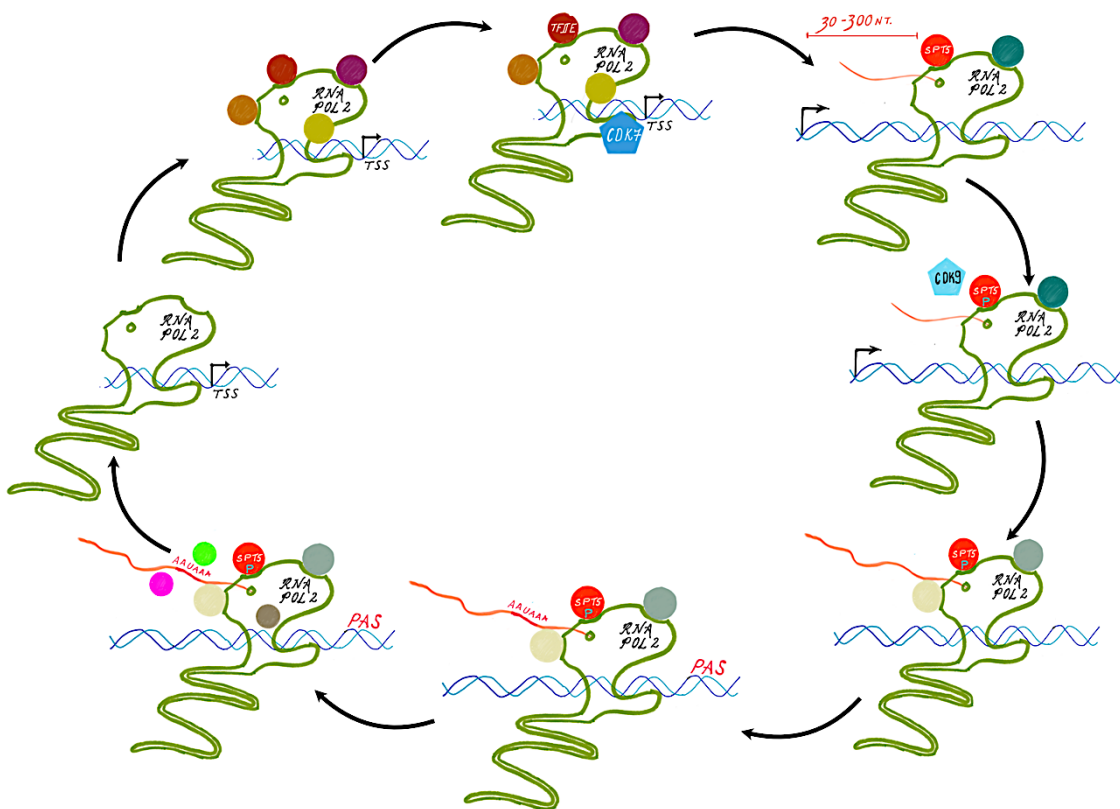


Figure 1.7 The transcription cycle and its potentially regulated steps.

Binding of GTFs (red, maroon, yellow and orange circles) facilitate Pol II (green) binding to chromatin. CDK7 (blue pentagon) binds to Pol II and phosphorylates TFIIE (maroon), allowing Pol II to break contact with it, while transcribing few hundred bases downstream of the TSS, producing an RNA (red line) and pauses, mediated by SPT5 (red circle). CDK9 (light blue pentagon) is recruited directly or indirectly and phosphorylates SPT5. Pol II escapes from the pause, either terminating or entering productive elongation. After the Pol II complex transcribes the gene, it encounters termination factors (grey and white circles), and the RNA is released. The freed Pol II can eventually reinitiate transcription.

Importantly, the changes in these set of proteins are accompanied by the changes in post translational modifications of these proteins as well as of Pol II. The modifications that are associated with Pol II during these transitions are restricted to the post-translational modifications of the heptad repeat (YSTPSPS) in the C-terminal domain (CTD). Among these post translational modifications, phosphorylation is the most well studied one. Broadly, phosphorylation of Ser5 by CDK7 is a mark of initiated, but paused Pol II; whereas Ser2 phosphorylation is associated with pause released Pol II (Fuda et al., 2009).

However, the precise nature of the phosphorylation status of these residues is much more complicated, and a Pol II CTD code has been proposed (Schüller et al., 2016) which indicates a signature of double (or more) phosphorylation might exist which behave as a landing pad for elongation factors to bind to Pol II CTD and assist in the transcription process (Bentley, 2014). However, the post transcriptional modifications are not restricted to just Pol II. The associated CDKs which regulate CTD phosphorylation, also impact other proteins like TFIIE and SPT5 which are bound to Pol II (see section 1.3.2).

In the last few years, a key paradigm which has appeared in the field of Pol II regulation is the fact that clustering of the above-mentioned proteins takes place with Pol II to facilitate their interactions and reactions. Among the first of such proteins reported were BRD4 and the mediator complex, which form liquid-like condensates with Pol II (Sabari et al., 2018) due to weak interactions driven by long disordered regions. Additionally, it was shown that these condensates containing Mediator and Pol II are sensitive to certain transcription inhibitors and their dynamic interactions regulate transcription elongation (Chong et al., 2018), thus indicating that the phase separated condensates can be modulated in response to chemical perturbations.

These results eventually paved way to the reports which indicated that phosphorylation of Pol II CTD had a major impact on clustering and activity of Pol II in these condensates as well. CTD condensates can get bigger until they are dissolved by phosphorylation at Ser5 CDK7 (Boehning et al., 2018), thus freeing up Pol II molecules for processive

elongation. Moreover, both *in vivo* and *in vitro* experiments suggest that while CTD phosphorylation by both CDK7 and CDK9 causes a reduction in CTD incorporation into Mediator condensates, it increased the incorporation of CTD into condensates rich in splicing factors (Guo et al., 2019). Thus, phase separation of Pol II also seems to be a key driving factor behind the transition from one stage to another during transcription (Cramer, 2019b). As mentioned earlier in this work, even though MYC proteins contain long stretches of disordered regions, it is not yet shown if MYC is localized to such Pol II rich condensates or not.

1.3.2 Transitions between the steps of the transcriptional cycle

The previously mentioned steps in the transcription cycle actually offer points of gene regulation, and thus are catalysed by well-defined set of kinases, which consist mostly of CDKs, and act in a step wise manner upon the factors bound to Pol II and on the CTD of Pol II itself.

Firstly, the transition of initiated to paused Pol II is associated with eviction of TFIIE from Pol II and the very same binding site is then occupied by SPT4/5 (Grohmann et al., 2011). This transition is associated with production of nascent RNA which is few hundred base pairs long and considered to be catalysed in part by CDK7 (Fig 1.7). It has been shown that inhibition of CDK7 increases TFIIE retention and prevents DSIF binding to Pol II (Larochelle et al., 2012). However, in spite of the indications that CDK7 might be driving the handover of SPT4/5 to Pol II, it is not clear what are the substrates of this kinase activity, nor it is clear if the handover is mediated by another chromatin bound transcription factor.

Next, the nascent RNA transcribed moves out of Pol II from the exit channel and serves as an identification site for SPT5 which binds selectively with the Pol II early elongation complex (Missra and Gilmour, 2010). SPT5 with its partner SPT4 give rise to a complex called DSIF (dichloro-1- β -D-ribofuranosyl-benzimidazole [DRB] sensitivity-inducing factor) and then SPT5 binds to Pol II near the RNA exit channel (Bernecky et al., 2017) (Vos et al., 2018a) (Vos et al., 2018b), where it helps in nascent RNA capping (Pei and Shuman, 2002). The negative elongation factor (NELF) complex then recognizes the Pol II–SPT5 complex and binds to it (Marshall and Price, 1992). The paused state is stabilized by both SPT5 and NELF (Yamaguchi et al., 1999) since NELF ensures that reactivation of the Pol II catalytic site does not take place (Vos et al., 2018a) (Vos et al., 2018b). The

transition from paused to elongating Pol II is catalysed by CDK9 (Lis et al., 2000). CDK9 phosphorylates SPT5 (Kim and Sharp, 2001) and NELF, which converts SPT5 into a positive elongation factor and releases NELF from Pol II, thus further increasing the processivity of Pol II by SPT5 mediated re-winding of transcribed DNA and thereby allowing the inhibition of aberrant backtracking of Pol II by TFIIS (Bernecky et al., 2017) (Ehara et al., 2017).

In order to probe the exact steps and drivers of the regulators of each step, small molecule inhibitors have been used which act on kinases and on other factors which drive these transitions. For example, flavopiridol, which binds to pTEFb and inhibits its activity (Chao and Price, 2001), is used extensively to probe Pol II behaviour (Bensaude, 2011). Similarly, THZ1, a selective covalent CDK7 inhibitor, was initially reported to have anti-tumour activity (Kwiatkowski et al., 2014) but has been since then used in mechanistic studies extensively (Sampathi et al., 2019). Eventually, chemical compounds with higher specificities for their respective kinases have been developed (Brägelmann et al., 2017) (Kelso et al., 2014), and different compounds have been utilized in studying molecular mechanisms in transcription based on the application involved. However, due to the shared structural properties among the CDKs, there exists a redundancy in the selectivity of these inhibitors (Table 1.1), and thus, caution must be maintained while interpreting results based on use of these inhibitors. However, for studying gene regulation, inhibitors have been used not only for modulating CDK activity, but to modulate MYC activity as well (Yin et al., 2003).

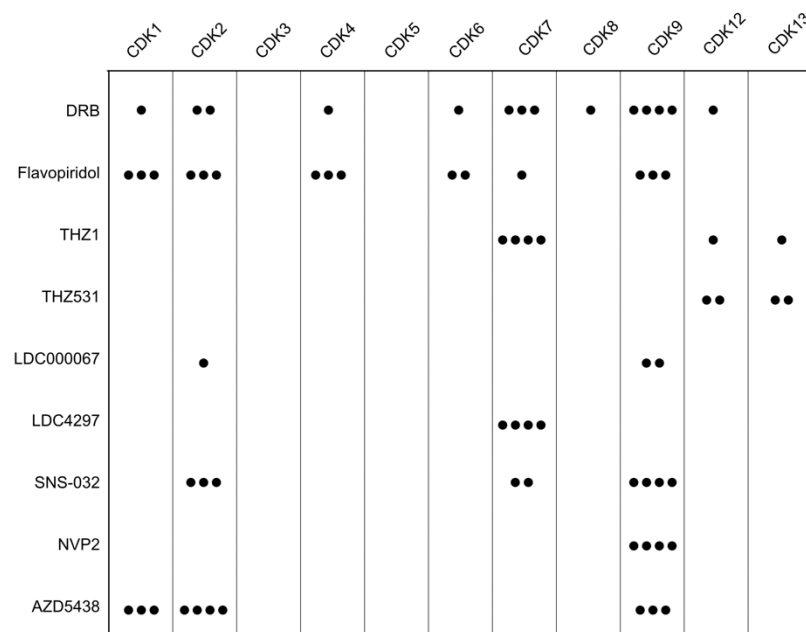


Table 1.1 Comparison of inhibitors used in studying transcription

Number of black spots indicate inhibitory activity specific to the mentioned CDK. The table is based on data from <https://www.selleckchem.com/CDK.html>.

One of the more strikingly examples of use of inhibitors in the field of MYC biology comes from treatment of embryonic stem cells with 10058-F4, an inhibitor of MYC/MAX heterodimerization (Yin et al., 2003). This caused a reduction of Ser 2 phosphorylated Pol II but had no effect on Ser5 phosphorylated Pol II (Rahl et al., 2010). This affected not only the levels, but the distribution of Pol II on the chromatin, as determined by ChIP-Seq analysis as well. 10058-F4 treatment decreased the RNA Pol II binding in the gene body, where elongating Pol II is found, but did not change the Pol II binding at the promoter. Similar results were obtained when MYC was depleted or when cells were treated with flavopiridol, a small molecule compound that inhibits CDK9 (Chao and Price, 2001) indicating role of MYC in transcription elongation.

However, ChIP-Seq of Pol II reveals only the distribution on the chromatin and not its activity. Also, utilizing phopho-specific antibodies for immunoprecipitation relies on the quality of antibody and comparison of the results between the antibodies requires certain assumptions. In fact, if the differential impact of MYC on various stages of transcription has to be taken into account (Michel and Cramer, 2013), than Pol II ChIP-Seq falls short of the required discrimination between defects in elongation and/or pause release.

In other words, Pol II molecules transcribing a smaller number of base pairs per minute will be cross linked more on the chromatin, thus increasing the read density in the region, giving false impression of increased Pol II activity. Hence, other methods, which measure Pol II activity are instead required while studying impact of MYC on transcription elongation.

1.3.3 Methods to study changes in transcription

In the recent years, due to rapid advances in next generation sequencing (NGS) technologies, deep sequencing methods have become relatively affordable to use and this has resulted in a wide variety of NGS based techniques to study the process of transcription on all genes globally (Wissink et al., 2019). Such techniques can be broadly divided into techniques which study the distribution of Pol II on chromatin and other techniques which study the actual enzymatic activity of Pol II. Among the techniques that

study the distribution of Pol II on chromatin, the most widely used are briefly touched upon in the next few paragraphs.

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) can be used to map RNA Pol II binding in a genome-wide manner (Park, 2009) by sequencing the DNA fragments which are bound to Pol II. While total and phospho-specific Pol II ChIP-Seq provide genome-wide maps of Pol II localization, they are blind for kinetic Pol II molecules. Thus, paused, stalled and slowly transcribing Pol II are captured in higher density as compared to moving and transcribing Pol II, and this confounding factor prevents accurate read out of transcription elongation and splicing (Table 1.2).

	Stages of Transcription					Properties of transcription					
	Initiation	Pausing	Pause Release	Elongation	Termination	TSS mapping	Pol II CTD Modification	RNA Synthesis	Elongation Rates	Processivity	Directionality
Total Pol II ChIP-Seq	•	•	•							•	
Ser5 Pol II ChIP-Seq	•					•					
Ser2 Pol II ChIP-Seq			•			•			•		
mNET-Seq	•	•	•		•	•	•				•
GRO-Seq	•		•	•		•					•
PRO-Seq		•		•	•						
Co-PRO-Seq	•	•	•								
4sU-Seq	•			•	•			•		•	•
4sUDRB-Seq	•			•		•		•	•	•	•
TT-Seq	•			•	•			•		•	•
SLAM-Seq					•			•			

Table 1.2 Comparison of techniques utilized to study transcription

Black spots indicate if a particular stage or property of transcription can be studied by the given technique or not. While multiple properties can be studied with the same technique, the ones used in this thesis are highlighted in green box, thus allowing most properties and stages of transcription to be captured.

Mammalian native elongating transcript sequencing (mNET-seq) reduces the reliance on Pol II position on chromatin alone, by extracting chromatin-bound RNA from immunoprecipitated Pol II, and subjecting it to deep sequencing after fragmentation (Nojima et al., 2015). This technique is based on native elongating transcript sequencing (NET-Seq) (Churchman and Weissman, 2011) and can capture genome wide nascent

RNA transcription and its processing events (Nojima et al., 2015). However, since mNET-Seq is still an IP based technique, the data from it still depends on antibody quality, crosslinking and fragmentation time.

On the other hand, several techniques are used to study the enzymatic activity of Pol II itself, and some them have been widely used in the last few years.

Gene run on sequencing (GRO-Seq) uses NTP-biotin run on followed by biotin affinity purification. Such techniques, or its variation have been successfully used to describe novel phenomenon like divergent transcription in mammals (Core et al., 2008), but remain dependent on *in vitro* run on where novel initiation events are prevented by addition of sarkosyl to the run on buffer. This might result in unintended release of paused Pol II and may disengage regulatory factors bound to the polymerase (Gregersen et al., 2020).

In 4sU-Seq, the metabolic labelling of nascent RNA by short pulse of 4-thiouridine (4sU) allows biotinylation of nascent RNA, thus allowing utilization of streptavidin based pull down. This enriched nascent RNA can then be studied by subjecting to cDNA library preparation and deep sequencing (Windhager et al., 2012). In this way, the kinetics of transcription elongation, molecular mechanisms of Pol II activity and their contribution to the regulation of gene expression can be studied in a genome wide manner. In 4sU-Seq, the nascent RNA labelling is carried out *in vivo*. This doesn't require nuclei isolation, thus reducing variability or cellular stress that might be introduced during the transcription reaction. Furthermore, the labelling reaction can be stopped by the direct addition of Qiazol or Trizol solution on the cells, thus controlling the exact duration of labelling in a very precise manner. This point become more relevant when experiments on many replicates and conditions are conducted at the same time.

4sUDRB-Seq is a technique, which is in technical terms similar to 4sU-Seq. A key feature of 4sU-Seq is that it allows measurement of only the processivity of Pol II (defined as nucleotide additions per initiation event). Although this knowledge already enables the differentiation between effects on Pol II pause-release as compared to premature termination of Pol II, it doesn't indicate if the actual nucleotide additions per min (elongation rate) are affected as well. While low elongation rates could be concurrent with low pausing (Saponaro et al., 2014), it has also been shown that, reduced elongation

rates lead to premature eviction of Pol II from the chromatin, and these effects could happen independently of changes in processivity (Mason and Struhl, 2005).

In other words, slowly transcribing Pol II could still be processive, and conversely low processivity can be a result of low elongation rates as well. Thus, to differentiate between these scenarios, Pol II molecules can be synchronized at the start site by treating cells with CDK9 inhibitors (table 1.1) and released with concomitant labelling of nascent RNA. This technique was originally developed by Oren lab (Fuchs et al., 2015) and since then, variations of this technique have been performed by Svejstrup and Shilatifard labs (Saponaro et al., 2014) (Gregersen et al., 2019) (Liang et al., 2018) and applied to differentiate regulation of elongation with pausing and/or other stages of transcription.

However, in order to account for the possibility that multiple stages could be regulated at the same time, data from two or more such techniques could be utilized to make a comparative account instead of relying on absolute changes. Also, since 4sU-Seq would not be a readout for paused Pol II, ChIP-Seq data would still be required to rule out changes in pausing or pause release.

It is a strong possibility that the effects on nascent transcription result from changes in Pol II processivity and cannot be caused by changes in pause release. Changes in Pol II initiation, pause release and processivity induce distinct alterations in 4sU-Seq and Pol II ChIP-Seq profiles, as ChIP-Seq measures the spatial distribution of Pol II molecules (regardless of elongation that is fast, slow or does not occur), while 4sU-Seq measures transcriptional activity during the 4sU-pulse period.

Therefore, changes in pause release result in a redistribution of reads between the pause site and the gene body in Pol II ChIP-Seq but do not cause a relative redistribution in the 4sU signal. Instead, the whole curve is “shifted” (Figure 1.8).

Introduction

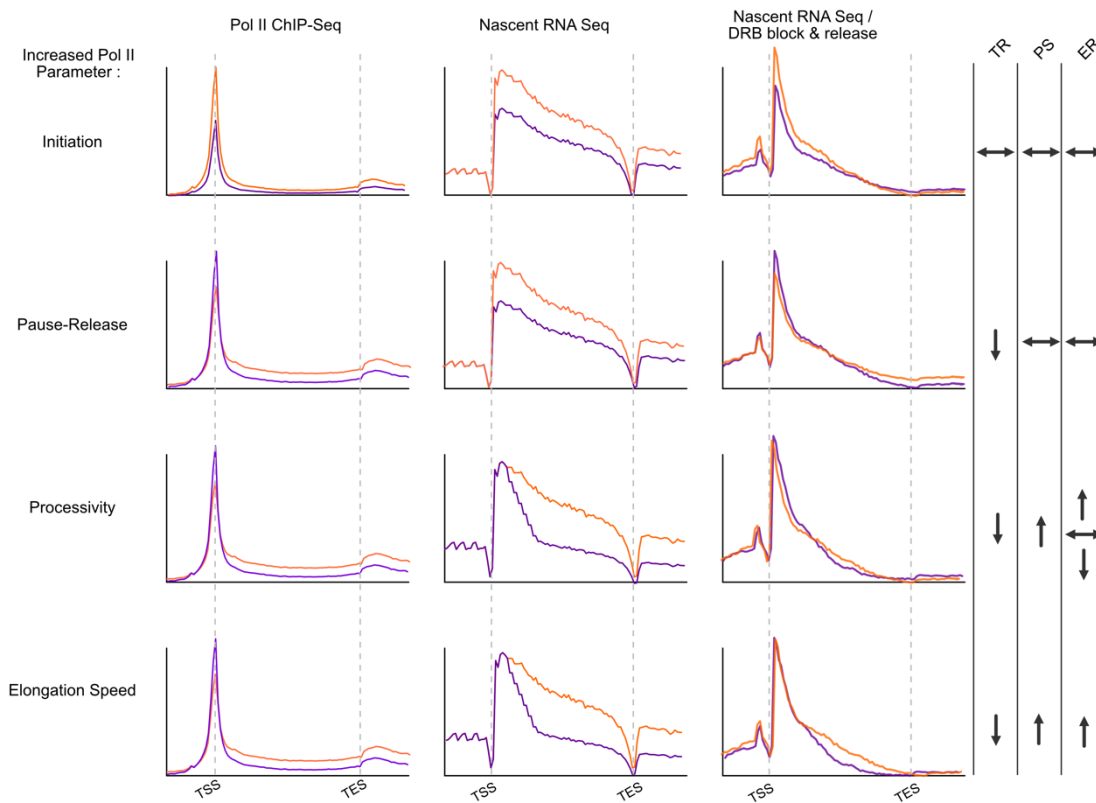


Figure 1.8 Orthogonal techniques to study Pol II behaviour.

Orange and purple lines indicate unperturbed and perturbed conditions causing change in Pol II behaviour. The curves indicate the possible extrapolated virtual average density profiles generated by various techniques and their interpretations. TR = Travelling Ratio, PS = Processivity Score and ER = Elongation rate.

Consequently, pause release affects the traveling ratio (TR, ratio of reads in paused regions and gene body of genes (Rahl et al., 2010)) calculated from ChIP-Seq but not the processivity score (PS, measured as ratio of reads in distal and proximal regions of the gene) calculated from 4sU-Seq. Additionally, the changes in processivity might either stem from changes in elongation speed (indicated by the point at which read density approaches background in 4sUDRB-Seq) or utilization of promoter proximal alternative polyadenylation sites.

In other words, since TR would still be increased if processivity of Pol II is reduced, it would be erroneous to conclude the key stage in transcription cycle which gets affected is pause release based on Pol II ChIP-Seq data alone, since the nascent RNA based techniques would add orthogonal dimension to the study, and would add crucial information on pol II behaviour which cannot be learnt from binding information of Pol II. These points highlight the need to study regulation of transcription by multiple

orthogonal techniques which can dissect changes in Pol II distribution from changes in Pol II activity.

1.4 Aim of the study

The main aim of this study is to find the exact molecular mechanism by which MYC regulates transcription. This aim would be pursued by working on two distinct, but related lines of investigation.

Firstly, it will be examined precisely which step in the transcription cycle is regulated by MYC. In order to find the key rate limiting step of transcription, metabolic labelling and purification of nascent RNA would be carried out in a cellular system where MYC levels can be acutely changed. These results would be used to test the hypothesis if MYC has an impact on transcription elongation or not, and more importantly, if these changes in elongation are rate limiting for MYC mediated gene regulation.

Secondly, description of the mechanism behind this regulation would be explored in terms of which transcription factor is responsible for changing Pol II behaviour via MYC. Additionally, which CDK is mediating the handover of such molecules from MYC to Pol II would be probed as well. This would be performed by a screen involving chemical inhibition of various kinases to find which of them is required by MYC to regulate transcription elongation.

While attempting to answer the above questions, it would be attempted to answer in which cellular compartments within the nucleus does the above processes take place, and if the formation of Pol II associated phase separated compartments are relevant for MYC or not.

More importantly, the degree of change in the precise step of transcription affected by MYC would be utilized to explain how MYC generates its specific gene expression profile, specifically with the hypothesis whether the changes in the gene regulation by MYC correlate with these changes or not.

Finally, the above outlined work carried out at low, medium and high cellular levels of MYC will be used to explain how MYC establishes tumour specific gene expression profiles, ultimately assisting in the functional understanding and eventually targeting of this important oncogene.

2 MATERIALS

2.1 Cell Lines

Table 2.1 List of cell lines used.

Name	Source	Reference
U2OS ^{MYC-Tet-On}	Wolf Lab	Walz et al, 2014
SH-EP MYCN-ER	Eilers Lab	Herold et al., 2019
HMLE	Weinberg Lab	N/A

2.2 Consumables

Table 2.2 List of consumables used.

Consumable	Manufacturer
Cell Culture Well Plates (12 Well, 24 Well)	Greiner
Cell Culture Well Plates (15 cm)	Greiner
CellView Cell Culture Dish, PS, 35/10 MM	Greiner
Filter Tips	Sarstedt
Glass Pasteur Pipettes	Hartenstein
Glasspipetts 5ml, 10ml, 25ml	Hartenstein
Microscope cover glasses (Ø 10 mm)	Hartenstein
Microscope slides (ca. 76x26 mm)	R. Langenbrinck
Parafilm	Bemis
PCR tubes	Sarstedt
Pipette Tips	Sarstedt
Reaction tubes (1.5 ml, 2 ml)	Sarstedt

2.3 Manufacturers & Developers

Table 2.3 List of manufacturers and developers.

Manufacturer/Developer	Location
Abcam	Cambridge, UK
Adobe	San José, CA, USA
AppliChem	Darmstadt, Germany
Beckman Coulter	Brea, CA, USA
Bio-Rad	München, Germany
Biologend / Biozol	San Diego, CA, USA
Biorad	Munich, Germany
Bioviz	Charlotte, NC, USA
Carl Roth GmbH	Karlsruhe, Germany
Carl Zeiss AG	Oberkochen, Germany
Cell Signaling	Danvers, MA, USA
Cellvis	Mountain View, CA, USA
Eppendorf AG	Hamburg, Germany
GE Health Care Life Sciences	Chicago, IL, USA
Gibco	Karlsruhe, Germany
Gilson	Middleton, WI, USA
Greiner Bio One	Frickenhausen, Germany
Hartenstein	Würzburg, Germany
Hartmann	Heidenheim
Heraeus Instruments	Langenselbold, Germany
Horizon	Cambridge, UK
Illumina	San Diego, CA, USA
Leica Microsystems	Wetzlar, Germany
LI-COR Bioscience	Lincoln, NE, USA
Life Technologies GmbH	Carlsbad, CA, USA
Merck Millipore	Billerica, MA, USA
Microsoft	Redmond, WA, USA
New England Biolabs GmbH	Frankfurt, Germany
Nikon Instruments Europe B.V	Amstelveen, Netherlands
Qiagen	Hilden, Germany
Roche Diagnostics	Rotkreuz, Switzerland
Santa Cruz Biotech	Santa Cruz, CA, USA
Sarstedt AG and Co.	Nümbrecht, Germany
Serif Europe	West Bridgford, UK
Sigma	St. Louis, MS, USA

Materials

StarLab	Hamburg, Germany
VWR Int.	Darmstadt, Germany
Wayne Rasband (NIH)	Bethesda, MD, USA

2.4 Equipment

Table 2.4 List of equipment used.

Equipment	Type	Manufacturer
Blotting System	PerfectBlue Tank Electro Blotter Web S	Peqlab
Cell culture incubator	BBD 6220	Heraeus
Cell counter	CASY cell counter	Innovatis
Centrifuge	Galaxy MiniStar	VWR Interantional
Centrifuge	5417 R	Eppendorf
Centrifuge	542	Eppendorf
Centrifuge	Multifuge 1S-R	NA
Centrifuge	Avanti J-26 XP	Beckman Coulter
Chemiluminescence imaging	LAS-4000 mini	Fujifilm
Fluorescence readers	Odyssey® CLx Infrared Imaging System	LI-COR
Fluorescence readers	Infinite 200 PRO Microplate Reader	Tecan
Heating block	Dry Bath System	STARLAB
Heating block	Thermomixer® comfort	Eppendorf
Microscope	Axiovert 40CFL	Zeiss
Microscope	DMI 6000 B	Leica
Microscope	Ti-Eclipse	Nikon
Microscope	SP8 (DM6000) upright microscope	Leica
Microscope	Elyra S.1 SIM - Super Resolution Microscope	Zeiss
Nucleic acid analysis	Experion Automated Electrophoresis	Biorad
Nucleic acid analysis	Fragment Analyser	Agilent
PCR thermal cycler	Mastercycler Pro S	Eppendorf
Photometers	Ultrospec™ 3100 pro UV/Visible	Amersham Biosciences
Photometers	NanoDrop 3000	Thermo Scientific
Power supply	PowerPac HC	Biorad
PVDF transfer membranes	Immobilion P and FL Transfer Membranes	Millipore
SDS-PAGE system	Mini-PROTEAN Tetra Cell	Biorad
Sequencing equipment	NextSeq500	Illumina

Materials

Vortex mixer	Vortex-Genie 2	Scientific Industries
Water bath	ED-5M heating bath	Julabo
Magnetic Plate	MAGNUM™ EX Universal Magnet Plate	Alpaqua
Magnetic Rack	DynaMag™-2 Magnet	Thermo Fisher Scientific

2.5 Chemicals

Table 2.2 List of chemicals used.

Chemical	Source
Doxycycline	Sigma
Protease Inhibitor Cocktail	Sigma
Phosphatase Inhibitor Cocktail 2	Sigma
Phosphatase Inhibitor Cocktail 3	Sigma
Hoechst 33342	Thermo Fisher Scientific
Alexa 488-conjugated phalloidin	Thermo Fisher Scientific
SYTO-RNA Stain	Thermo Fisher Scientific
DRB (5,6-Dichlorobenzimidazole 1-β-D-ribofuranoside)	Sigma
THZ1 Hydrochloride	MedChem
LDC000067	Selleckchem / Biozol
LDC4297	Selleckchem / Biozol
4-thiouridine (4sU)	Sigma
Actinomycin D	Sigma
Dynabeads MyOne Streptavidin T1	Thermo Fischer Scientific
Pierce DTT (Dithiothreitol), No-Weigh Format	Thermo Fischer Scientific
MG132	Selleckchem / Biozol
16% Paraformaldehyde	Science Services GmbH
Poly-L-Lysine Solution 0.01%	Sigma
EZ-Link™ HPDP-Biotin	Thermo Fischer Scientific
Pierce™ Dimethylformamide (DMF), Sequencing grade	Thermo Fischer Scientific
Bicinchoninic acid solution	Sigma

2.6 Media & Supplement

Table 2.3 List of media and supplement used.

Media/Supplement	Source
DMEM	Thermo Fisher Scientific
RPMI	Thermo Fisher Scientific
Opti-MEM	Thermo Fisher Scientific
FBS	Sigma
Penicillin/Streptomycin	Sigma
FluoroBrite™ DMEM	Thermo Fisher Scientific

2.7 Kits and Detection Systems

Table 2.4 List of kits and detection system used.

Kits/Detection System	Source
PLA Duolink Kits	Sigma
Quant-iT Ribogreen RNA Assay Kit	Sigma
Quant-iT PicoGreen dsDNA Assay Kit	Sigma
NEBNext ChIP-Seq Library Prep Master Mix Set	NEB
NEBNext Ultra Directional RNA Library Prep Kit	NEB
NEBNext rRNA Depletion Kit (Human/Mouse/Rat)	NEB
RNeasy MinElute Cleanup Kit	Qiagen
miRNeasy MiniKit	Qiagen
Dynabeads MyOne™ Streptavidin T1	Thermo Fisher Scientific
NextSeq 500/550 High Output Kit v2 (75 cycles)	Illumina
SYBR Green qPCR Master Mix	Thermo Fisher Scientific
Agencourt AMPure XP	Beckman Coulter
Fluoromount Aqueous Mounting Medium	Sigma
NEBNext Multiplex Oligos for Illumina	NEB
Duolink Probemaker	Sigma

2.8 Primary Antibodies

Table 2.5 List of primary antibodies used.

Antibodies	Source	Application	Concentration
Rabbit monoclonal anti-MYC (Y69)	Abcam	IF, IB, PLA	1:400, 1:5000
Mouse Monoclonal anti-MYC (C33)	Santa Cruz Biotechnology	IF, PLA	1:100
Mouse monoclonal anti-vinculin	Sigma	IB	1:5000
Rabbit polyclonal anti-HA (Y-11) X	Santa Cruz Biotechnology	IF, PLA	1:1500
Rabbit monoclonal anti-SUPT5H	Abcam	IF, PLA	1:500
Rabbit polyclonal anti-SUPT6H	Novus Biologicals	IF, PLA	1:200
Mouse monoclonal anti-Pol II (F-12)	Santa Cruz Biotechnology	IF, PLA	1:800
Mouse polyc. Pol II (H5) (pSer2)	Biolegend	IF, PLA, IB	1:400, 1:1000
Rabbit monoclonal anti-CDK2 (78B2)	Cell Signaling Technology	IB	1:1000
Mouse monoclonal anti-TF2EB (A-1)	Santa Cruz Biotechnology	IF, PLA, IB	1:100
Rabbit monoclonal K48-Ub	Abcam	IF, PLA	1:600
Mouse monoclonal anti- Pol I	Santa Cruz Biotechnology	IF	1:2500

2.9 Secondary Antibodies

Table 2.6 List of secondary antibodies used.

Fluorophore/Tag	Species	Source	Application	Concentration
Alexa 488	rabbit	Thermo Fisher Scientific	IF	1:600
Alexa 488	mouse	Thermo Fisher Scientific	IF	1:600
Alexa 568	rabbit	Thermo Fisher Scientific	IF	1:600
Alexa 568	mouse	Thermo Fisher Scientific	IF	1:600
Alexa 647	rabbit	Thermo Fisher Scientific	IF	1:600
Alexa 647	mouse	Thermo Fisher Scientific	IF	1:600
ECL- IgG Horseradish Peroxidase	rabbit	GE Healthcare	IB	1:7500
ECL- IgG Horseradish Peroxidase	mouse	GE Healthcare	IB	1:7500

2.10 Solution and Buffers

4-Thiouridine (4sU), 100 mM

- Dissolve 250 mg of 4sU in 9.6 ml of nuclease free water. Use 200 μ l per 10ml.
- Make 600 μ l aliquots

5,6-Dichlorobenzimidazole 1-b-d-ribofuranoside (DRB), 50 mM

- Dissolve 50 mg of DRB in 3135 μ l of nuclease free water. Use 20 μ l per 10 ml.
- Make 100 μ l aliquots
- Freeze at -20°C

Biotin labelling buffer, 2.5 \times

- Mix 25 mM Tris (pH 7.4) with 2.5 mM EDTA. (1.25 mL Tris (1 M) + 250 μ l 0.5 M EDTA + 48.5 ml nuclease free water)
- Use 100 μ l per sample.
- Make 1 ml aliquots.
- Freeze at -20°C

Dynabeads washing buffer, 2X

- Mix 2 M NaCl with 10 mM Tris (pH 7.5), 1 mM EDTA and 0.1% (vol/vol) Tween 20.
- E.g. (20 ml 5M NaCl + 500 μ l Tris 1M + 5 ml 1% Tween 20 + 24.5 ml nuclease free water)
- Use 400 μ l per sample.
- Make 5 ml aliquots
- Freeze at -20°C

Dynabeads washing solution A

- Mix 100 mM NaOH with 50 mM NaCl.
- E.g. (2.5 ml 2N NaOH + 500 μ l 5M NaCl + 47 ml Nuclease free water)
- Use 100 μ l per sample.
- Make 1 ml aliquots
- Freeze at -20°C

Dynabeads washing solution B

- Prepare 100 mM NaCl. (1 ml 5N NaCl + 49 ml Nuclease free water)
- Use 50 μ l per sample.
- Make 1 ml aliquots
- Freeze at -20°C .

Materials

Biotin HPDP–DMF stock

- Dissolve 50 mg of EZ-Link biotin HPDP in 50 ml of DMF.
- Heat the mixture to 37 °C for ~1 h until dissolved.
- Use 50 µl per sample.
- Make 1 ml aliquots
- Freeze at -80°C. DMF corrodes some classes of plastics. Do not use cell-culture plastic pipettes. Use only plastic tips which are of the PP grade for aliquoting.

PBS

- 137 mM NaCl
- 2.7 mM KCl
- 10.1 mM Na₂HPO₄
- 1.76 mM KH₂PO₄
- autoclaved

RIPA buffer

- 50 mM HEPES (pH 7.9)
- 140 mM NaCl
- 1mM EDTA
- 1% Triton-X-100
- 0.1% Sodium deoxycholate 0.1% SDS

SDS sample buffer 6X

- 1.2 g SDS pellet
- 6 mg Bromophenol blue
- 4.7 ml Glycerol 86%
- 2.1 ml water
- 0.93 g DTT

SDS running buffer

- 25 mM Tris Base
- 250 mM Glycine
- 0.1% SDS

TBS-T

- 0.2% Tween-20
- 25mM Tris
- 140mM NaCl
- Adjust to pH 7.4

Trypsin solution

- 0.25% Trypsin

Materials

- 5 mM EDTA
- 22.3 mM Tris (pH 7.4)
- 125 mM NaCl

200X Reduction agent for Bis-Tris system

- 1M Sodium bisulfite
- 3.5X Bis-Tris buffer
- 1.25 M Bis-Tris
- Adjust pH 6.7 with HCl

20X Transfer buffer Bis-Tris system

- 25 mM Bicine
- 25 mM Bis-Tris
- 1 mM EDTA pH 7.2

2.11 Nucleic Acids

Type	Source	Reference	Sequence/Target
siRNA	Horizon Discovery Group	L-003282-02-0050	ON-TARGETplus Human MYC
siRNA	Horizon Discovery Group	D-001810-10-50	ON-TARGETplus Non-targeting
siRNA	Horizon Discovery Group	L-006531-00-0005	ON-TARGETplus Human GTF2E1
siRNA	Horizon Discovery Group	L-017725-00-0005	ON-TARGETplus Human GTF2E2
siRNA	Horizon Discovery Group	L-003241-00-0005	ON-TARGETplus Human CDK7
siRNA	Horizon Discovery Group	L-009782-00-0005	ON-TARGETplus Human HCE1
RTqPCR	Fuchs et al., Nat.Prot.,2014	OPA1+2kb_F	ACCATGGATGCCATTGAGTCA
RTqPCR	Fuchs et al., Nat.Prot.,2014	OPA1+2kb_R	TGTGCCATCACCAGGAGACAT
RTqPCR	Fuchs et al., Nat.Prot.,2014	OPA1+22kb_F	AGATGTGAAAGTCAGGTGGCAT
RTqPCR	Fuchs et al., Nat.Prot.,2014	OPA1+22kb_R	CAGACCTGCCAAGAAGTGCTT
RTqPCR	Fuchs et al., Nat.Prot.,2014	OPA1+28kb_F	GTGGCCAGGCTTTAGCTGAAT
RTqPCR	Fuchs et al., Nat.Prot.,2014	OPA1+28kb_R	CAGCCCTCAGACTGAACCAGTT
RTqPCR	Walz et al., Nature, 2014	b2MGF	GTGCTCGCGCTACTCTCTC
RTqPCR	Walz et al., Nature, 2014	b2MGR	GTCAACTTCAATGTCGGAT

2.12 Bioinformatic Tools and versions

Tools	Version	Source
Bowtie	v1.2	http://bowtie-bio.sourceforge.net/index.shtml
Bowtie	v2.3.2	http://bowtie-bio.sourceforge.net/index.shtml
Bedtools	v2.26.0	https://github.com/arq5x/bedtools2/releases
SAMtools	v1.3	http://samtools.sourceforge.net
Deeptools	v2.4.2	https://deeptools.readthedocs.io/en/develop/index.html
Ngsplo	v2.61	https://github.com/shenlab-sinai/ngsplo/

3 METHODS

3.1 Cell Culture Methods

3.1.1 Cell Growth, splitting, freezing and thawing

U2OS MYC-Tet-On were cultured in DMEM supplemented with 10% FBS and 1% penicillin/streptomycin solution. Human mammary epithelial cells (HMLE) were cultivated in advanced DMEM/F12 supplemented with 20 ng/ml EGF, 0.5 mg/ml Hydrocortisone, 10 mg/ml Insulin (human), 1% penicillin/streptomycin solution, 1% glutamine solution and 15 mM HEPES (pH 7.2). Human Neuroblastoma SH-EP cells were grown in RPMI medium supplemented with 10% FBS and 1% penicillin/streptomycin solution. All cell lines were cultured at 37°C, 5% CO₂ and routinely screened and found negative for mycoplasma contamination in a PCR-based assay.

For splitting of cell cultures, the media was removed, and cells were washed once with PBS. 1 ml of warm trypsin solution was added per 15 cm dish and incubated for 2-3 mins at 37°C. Trypsin activity was stopped by adding fresh medium containing FBS, followed by collection in falcon tubes and centrifuged at 1200 rpm for 4min. The media was then aspirated, and the cell pellet was resuspended in fresh medium and a portion of cell suspension was plated in new dishes with medium.

For long-term freezing storage, cell pellet was resuspended in 1ml freezing medium, transferred in cryo-vials and slowly frozen at -80°C. After 24 hours, cells were stored in a liquid nitrogen storage tank. To thaw frozen cells, the cryo-vials was quickly warmed up at 37°C in a water bath and then the cell suspension was transferred in a dish containing fresh medium.

3.1.2 siRNA transfection

2.2 million U2OS cells were seeded over 15 cm dish, approximately 24 hrs before transfection. For each transfection, 16 µl of respective siRNA and Lipofectamine RNAiMAX were added to separate tubes of 1.15 ml OptiMEM. After brief vortexing,

they were allowed to be incubated for 5 min at RT. This was followed by mixing both of them and incubated for another 20 min at RT. This mix was then added to 8 ml transfection medium (DMEM with 2% FBS), and added on the cells which were washed at least 2 times with PBS. On the next day (18 hrs post transfection), cells were washed multiple times with PBS and 10 ml full media was added, 6 hrs after which cells were split and then harvested 24 hrs later.

3.2 Cell Biology Methods

3.2.1 Indirect Immunofluorescence

20,000 cells per 10 mm were grown directly on cover slips placed in each well of 24 well plate. Before fixing, the cells were washed twice with ice cold PBS and fixed by incubating them in 4% paraformaldehyde for 15 minutes at room temperature. This was followed by washing in TBST and permeabilized with 0.3% Triton-X 100 solution in PBS, followed by washing in TBST. Blocking was carried out subsequently with 10% goat serum, 2% BSA, 5% sucrose in PBS solution for 45 minutes at 37°C. The cover slips with the cells were moved into a wet chamber and incubated with 80 µl primary antibody diluted in blocking solution. After overnight at 4°C, cells were washed 3 times with TBST and then incubated for 45 minutes at 37°C with 40 µl secondary antibody in blocking solution, Hoechst nuclear stain (1:1000) and Phalloidin, where applicable. After washing 3 times with water, the cover slips were mounted on a glass slide using a small drop of aqueous mounting medium. The slides were stored at 4°C in the dark scanned under confocal microscope.

3.2.2 Proximity Ligation Assay

Cells were plated at a density of 20,000 - 35,000 cells per 10 mm² coverslip in each 24 well plate and left to adhere overnight before washing with ice cold PBS and fixing with 4% paraformaldehyde. After washing away the excess PFA with PBS, the cells were permeabilized with 0.3% Triton X-100, followed by 3X wash in TBST, and treated with blocking solution (10% goat serum, 2% BSA, 5% sucrose in PBS) for 45 min. After being transferred to wet chamber, the coverslips were incubated overnight at 4°C with 80 µl primary antibodies in blocking solution and then used in *in situ* proximity ligation assays as per manufacturer's instructions. Briefly, cells were treated for 1 hour at 37°C with plus and minus probes directed at rabbit and mouse antibodies, respectively. Then probes were

ligated for 30 minutes at 37°C. Next, *in situ* PCR amplification was done with Alexa 568- or Alexa 488-conjugated oligonucleotides for 2.5 h at 37°C. Samples were stained with Hoechst 33452 and Alexa 488- or Alexa 568-conjugated phalloidin mounted on microscope slides using Fluoromount (Sigma), and imaged under a confocal microscope.

3.3 Molecular Biology & Biochemistry Methods

3.3.1 Determination of Protein Concentration

For protein isolation, cells were washed twice with ice cold PBS, scraped in a 1.5 ml tube and then pelleted at 1200 rpm for 5 minutes at 4°C. The cell pellet was lysed by adding 300 µl ice cold RIPA buffer with proteinase and phosphates inhibitors (1:1000) and left for 20 minutes on ice. Subsequently, the cellular debris was removed by centrifuging the tubes at 13600 rpm for 5 minutes at 4°C followed by aspirating the supernatant in a new tube. This was subjected to flash freeze in liquid nitrogen and stored at -80°C. For protein concentration determination, 1.5 µl of protein sample was mixed with 100 µl bicinchoninic acid and CuSO₄ (50:1) solution in a 96-well plate in duplicates. The solution was incubated 37°C till the colour change was observed, which was followed by measurement of absorption at 550 using an appropriate reference. The measured values were then compared with a BSA calibration curve to estimate the protein concentration.

3.3.2 Immunoblotting

Proteins were separated in size by SDS-PAGE, followed by transfer on a PVDF membrane by means of a tank blot system. Briefly, the PVDF was first activated for 30 seconds in methanol and then kept in in the transfer buffer. Next, blotting sandwich was assembled while being flooded in transfer buffer in the following order: sponge, Whatman papers, gel, membrane, Whatman papers, sponge. In order to remove air bubbles, each layer of sandwich was rolled over with a plastic roller. Finally, the transfer was performed by exposing the sandwich to 400 mA electric current for 3 hrs at 4 °C.

Following the transfer, membrane was washed once in TBST before blocking at room temperature with a 5% solution of milk in TBST. Next, the blocked membrane was incubating with primary antibody overnight at 4 °C. Afterwards, the membrane was washed TBST before being treated with HRP-labelled secondary antibody for 1 hour,

followed by washing with. Immobilon Western HRP Substrate was used according to the manufacturer's instructions and detected by LAS- 4000 mini imager (Fujifilm).

3.3.3 RNA Extraction and cDNA synthesis

Specifically, for RT-qPCR, the cells from 10 cm plate were harvested with 700 μ l peqGOLD TriFast solution after aspirated of media and the lysate was transferred into a 1.5 ml reaction tube. 200 μ l chloroform was added and the suspension was vortexed for 10 seconds, followed by centrifugation for 10 minutes at 20000 g at 4°C. The upper aqueous phase containing the RNA (~350 μ l) was transferred into a new reaction tube, and treated with 350 μ l isopropanol and 1 μ l glycogen. Next, the samples were left on ice for 10 minutes followed by centrifugation for 10 minutes at 20000 g at 4°C. The supernatant was aspirated, and the RNA pellet was subsequently washed twice with 75% EtOH. The washed RNA pellet was air-dried and solubilized in 20 μ l nuclease free water. The concentration of RNA was measured with NanoDrop 1000 and its purity was estimated via ratio of absorbance at 260nm and 280nm, and eventually stored at -80°C.

Total RNA was transcribed into complementary DNA (cDNA) by taking 4 μ l of random hexanucleotides and adding them 1 μ g of RNA diluted in 15 μ l of water. The mix was denatured by heating at 65°C for 1 minute and transferred to ice for 2 minutes.

The following master mix was set up for each cDNA mix per sample:

20 μ l 5X MLV buffer

0.4 μ l Ribo-lock RNase inhibitor

2.5 μ l 10 mM DNTPs

2 μ l M-MLV Reverse Transcriptase (200U/l, Promega)

55.1 μ l nuclease free water

80 μ l of the above master mix was added to 20 μ l of RNA sample with random hexamers prepared earlier, and subjected to the following order temperatures:

23°C - 10 minutes

37°C - 50 minutes

70°C - 5 minutes

For measuring pre-mRNA, the cDNA was diluted in 150 μ l of RNAase free water and further used for RT-qPCR.

3.3.4 Real Time Qualitative PCR (RT-qPCR)

For each sample technical replicates were performed, and the reactions were set up with 1:9 diluted 10 μ M primer-pairs in SYBR Green mix, mixed with equal volume of cDNA. The quantification of the amplified DNA was measured by fluorescence monitoring in every cycle after the end of the elongation step. Calculation of the relative transcript amount or DNA enrichment was performed using the $\Delta\Delta$ -CT method (Applied Biosystems User Bulletin 2). For normalization of RNA samples, the housekeeping gene β 2MG was used.

3.4 Imaging Methods

3.4.1 Confocal Microscopy

For obtaining images with *Nikon Ellipses TE* confocal line scanning microscope system, slides were scanned at constant scanning rate with 300 nm stack size and fixed digital zoom at 60X object under oil (Plan Apo, NA 1.42, WD 0.15 mm) immersion. The confocal lasers used were 405 nm UV diode, 488 nm Argon ion, 568 nm Green/Yellow Diode and 647 nm red diode with minimum pinhole size possible. For image acquisition settings, aspect ratio was ensured to be 1:1 and oversaturated pixels were checked with intensity (Hi-Lo) mode. Stacks of 0.3 μ m spaced sections were taken using with 1200 \times 1200-pixel resolution with 12-bit depth range at 800 scanning speed. For using *Leica SP8* confocal line scanning microscope system, white laser with AOBS (Acousto-Optical Beam Splitter) along with UV diode laser was used. 2x PMT, 1x PMT (Trans) and 2x HyD detectors were used to capture excited emissions, while the image acquisition settings were set to 1064 x 1064 pixels, with a depth range of 8 bits.

3.4.2 Structured Illumination Microscopy

The cells for treated in the same fashion as for confocal microscopy, except high precision glass coverslips were used to seed the cells. For obtaining images with *Zeiss Elyra S.1* scanning system, the coverslip mounted on the slides was placed on the holder scanned with Zeiss inverted microscope using Immersol 30°C Oil (NA = 1.518). *Xcite 120* PCQ excitation lamp was used for multichannel fluorescent illumination to locate cells with

oculars. *ZEN 2012* coupled with sCMOS pCO-EDGE camera was used to drive the scanning system. For all images, The Plan-Apochromat 63X/1.4 Oil DIC was used. The grating size was calculated automatically during acquisition by *ZEN 2012* software depending on the laser. Exposure time, laser intensity and number of grating rotations was kept constant for individual experiment and comparisons. After raw image acquisition, the super-resolution image was processed using 3D-SIM setting for all channels for whole sized image. Laser excitation light (672, 561 and 488 nm) was shuttered using a tilt mirror filter *FSet77 HE* with grating size calculated separately each individual laser. Unprocessed image stacks were composed of 25 images per z-section (five 72-degree phase-shifted images per angle at each of five interference pattern angles). The z-sections were scanned at a spacing of every 250 nm for a total raw data image count of 525 images per 1.5 μm section. Full super-resolution 1 μm wide stacks with 78.35 x 78.35 μm field of view at 2048 x 2048 pixels resolution with 16-bit dynamic range was captured.

3.4.3 *direct* Stochastic Optical Reconstruction Microscopy

20,000 cells per well were grown on 8 well CellVue High Precision glass chamber, and treated in the same manner as for confocal microscopy, with an additional end-fixation step with 2% paraformaldehyde and 0.2% glutaraldehyde.

*d*STORM (direct Stochastic Optical Reconstruction Microscopy) was performed by Patrick Eiring and Marvin Jungblut (Department of Biotechnology, University of Würzburg) on a single molecule sensitive widefield setup. The setup was built around an inverted microscope equipped with an oil-immersion objective (60x, NA 1.45). Excitation lasers used were 641 nm (Cube 640-100C; Coherent) and a 532 nm diode laser (Nano250; Linos), with irradiation intensities of 1–5 kW/cm². Laser lines were used in combination with by dichroic mirror and coupled into the microscope objective using an polychromatic mirror which also separated the excitation light from the fluorescence. Pair of lenses (L1 and L2) placed in the excitation path to direct the laser on the back focal plane of the oil-immersion objective ensured homogeneous widefield excitation. The experiments were performed in internal reflection fluorescence (TIRF)-illumination or by a highly inclined and thin beam that allows axial sectioning (highly inclined and laminated optical sheet, HILO). The fluorescence light in the detection path was filtered

using bandpass filters and imaged with an EMCCD with quantum yields of 80–90% in the visible range.

To minimize the drift, a nosepiece stage to decouple the specimen and the objective from the microscope stative was used. The switching buffer containing 100 mM MEA (β -mercaptoethylamine) was added to the fixed cells before data acquisition. The reconstruction was done with the opensource software rapidSTORM which was developed in the Sauer laboratory (Wolter et al., 2012). The reconstructed images were set to 10 nm per pixel and display the data colour-coded whereas the colour intensity is directly proportional to the localization density per pixel weighted with the photon number per localization. Before all the actual measurements, calibration measurements to create a transformation matrix were performed with 200 nm fluorescent microspheres.

3.4.4 Image Processing

Images from confocal systems were processed using ImageJ (<http://imagej.nih.gov/ij/>). The z stacks were converted to maximum intensity projections, processed with 0.8 factor Gaussian blur in relevant channels and were used for intensity quantification, counting of PLA signals and co-localization quantification. For display purposes, the images were normalized to same value of brightness and contrast within a single experiment using ImageJ. During 3D-SIM image processing, the output of raw data from the ZEN 2012 was processed at the same time using 3D-SIM algorithm in SR mode of ZEN 2012. The microscope was routinely calibrated to calculate both the lateral and axial limits of image resolution under our experimental conditions. Super resolved images were rendered in ImageJ for conversion to merged channel format and the final image for display and co-localization/intensity measurement were normalized for brightness to same value in a single set of experiment using ImageJ. After normalization, an ImageJ plug-in, Coloc2 was used for calculation of Pearson's co-efficient and Overlap co-efficient after Manders. During calculation, manual thresholding was necessary to be carried out as the images were normalized already.

3.5 Next Generation Sequencing Based Methods

3.5.1 mRNA-Seq

RNA was isolated using RNeasy kit with DNAase digestion on columns, following the manufacturer's instruction manual and eluted in 20 μ l of nuclease free water. After extraction, RIN assessment using Standard RNA Sensitivity Kit on Fragment Analyzer was carried out. Only samples with RIN > 8 were processed further. Following which mRNA was isolated with NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB) and the subjected to library preparation using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina, following the manufacturer's instructions. After the final purification, the concentration and size distribution of the library was checked with Fragment Analyzer using the NGS Fragment High Sensitivity Analysis Kit (1-6,000 bp). The libraries were sequenced on a NextSeq500 Illumina platform for 75 cycles.

3.5.2 4sU-Seq

4-8 million cells per 15 cm plate were treated with 1-2 mM 4-thiouridine (4sU) for 10 or 15 min before being harvested in QIAzol reagent. Total RNA was subjected to extraction using miRNeasy kit with on-column DNAase digestion implemented. The resulting RNA was tested for quality on a Bioanalyzer or Fragment Analyzer, and quantified on a Nanodrop spectrophotometer. 50- 120 mg of resulting total RNA in 100 μ l nuclease free water was biotinylated using EZ-Link Biotin-HPDP in 0.2 mg/ml dimethylformamide in biotin labelling buffer for 2 h with rotation at 25°C. Resulting biotinylated RNA was then subjected to clean up using chloroform-isoamylalcohol (24:1) extraction in MaXtract (high density) tubes. The top aqueous phase was taken and mixed with 1:10 volume of 5 M NaCl and an equal volume of isopropanol along with 1 ml of Glycoblu coprecipitant. Resulting mixture was then placed on room temperature for 5 min followed by centrifugation at 20,000 g for 20 min while cooling maintained at 4°C. The supernatant from the centrifuge aspirated and the remaining pellet was washed twice by adding 75% ethanol and centrifugation at 20,000 g for 10 min while cooling maintained at 4°C. After the second centrifuge, the pellet was air dried and resolved in 100 μ l nuclease free water.

The resulting biotinylated RNA was subjected to pull down using 50 μ l of Dynabeads MyOne Streptavidin T1 beads which were suspended in 100 μ l wash buffer for 15 min at 25°C with rotation. The beads were washed with biotin wash buffer while implementing

magnetic separation. Ultimately, 4sU labelled RNA was eluted with 100 μ l of freshly prepared 100 mM DTT in nuclease free water, and cleaned using RNeasy MinElute Spin columns. The 4sU labelled RNA was quantified with the RiboGreen Assay subjected to cDNA library preparation. cDNA libraries were prepared using Ultra RNA Library Prep kit in combination with NEBNext rRNA Depletion Kit (Human/Mouse/Rat) while using 16-19 PCR cycles, depending on the input. Subsequent libraries were checked for size distribution and concentration on a Bioanalyzer or on the Fragment Analyzer using the NGS Fragment High Sensitivity Analysis Kit (1-6,000 bp).

3.5.3 DRB 4sU-Seq

8-10 million cells per 15 cm plate were seeded before carrying out 2.5 h of DRB treatment after which the cells were washed with PBS and released from transcription block by adding culture medium containing 1-2 mM 4sU for indicated times for 10 or 15 minutes for the respective release time. The labelling was subsequently terminated by harvesting the cells in 2.1 ml QIAzol reagent and storing in 3 separate tube, which were flash frozen in liquid nitrogen. After thawing, the lysates were processed in the same way as for 4sU-Seq.

3.5.4 Illumina NextSeq500 Sequencing

Once individual libraries with compatible index combinations were pooled together, they were again tested for library size and concentration, while running a previously calibrated library mix on the same lane on fragment analyser. Following the correction for variation based on the previously calibrated library mix, the new library mix was diluted to 4 nM concentration and mixed in equal volume with 200 mM NaOH for denaturation. The denaturation was halted by adding 200 mM Tris-Cl following by dilution in Illumina supplied hybridisation buffer to a final concentration of 1.8 pM. In order to check for base calling errors, 1-3 μ l of 1.8 pM pre-denatured PhiX library was added to the mix and loaded on high throughput Illumina flowcell for NextSeq500 and subjected to 75 cycles of sequencing by synthesis followed by 8+8 cycles of index sequencing.

3.6 Bioinformatic Methods

3.6.1 References and databases

The human reference genome build 37 released in 2009 (hg19) was used throughout the study. RefSeq genomic coordinates for all genes, exons, introns and rRNA cluster were downloaded from UCSC (University of California, Santa Cruz) table browser. Pre-built chromosome sizes and index for Bowtie were downloaded from Illumina iGenomes suite. For gene set enrichment analysis, version 5.2 of C2, C5 and Hallmark databases were downloaded from Molecular Signature Database hosted by Broad Institute.

3.6.2 Read Alignment

FASTQ files from Illumina's BaseSpace environment were downloaded and checked for indexed combinations post catenation in the following way:

```
# Checking for number of occurrences of each index combination in the merged file
cat sample.fastq | grep '^@NB500931' | cut -d : -f 10 | sort | uniq -c | sort -nr > sample.txt
```

In order to find the coordinate location of the short reads on the reference genome, *bowtie2* was used with one mismatch allowed. Before alignment, the FASTQ files downloaded, catenated and unzipped on the high-performance cluster (HPC) and were checked for quality with FASTQC.

For 4sU-Seq data, this was implemented with *bowtie* in the following manner:

```
bowtie -t -S --norc -p 64 hg19 sample.fastq sample_norc.sam
bowtie -t -S --nofw -p 64 hg19 sample.fastq sample_nofw.sam
```

The command to align to reference genome was as follows:

```
bowtie2 -x hg19 sample -S sample.sam -N 1 -p 75 2> sample.txt
```

In case of stranded alignment, two files, one for each strand was generated using *-norc* and *-nofw* parameters of *bowtie2*.

The resulting *sam* file was transformed into a binary bam file using *samtools*, while being sorted during the conversion as follows:

```
samtools view -bS sample.sam | samtools sort -T t1/aln.sorted - -o sample.sorted.bam
```

3.6.3 Normalization Procedures

The mapped reads from each resulting *bam* file was collected by `-F 4` flag using *samtools*, counted and stored into a single text file:

```
samtools view -c -F 4 $sample.sorted.bam >> Total_hg19_Mapped_Reads.txt
```

For 4sU-Seq and 4sUDRB-Seq, the exons from contaminating mature RNA confound the counts of *bam* files. Hence, only the reads falling into the intronic regions was kept. Additionally, in order to ensure the normalization was carried out only based on Pol II transcribed reads, rRNA reads were also removed from the *bam* fields before normalization:

```
bedtools intersect -abam sample.sorted.bam -b Exons.bed -v > sample.ex_cl.sorted.bam
bedtools intersect -abam sample.ex_cl.sorted.bam -b 3pUTRexons.bed -v > sample.ex_3p_cl.sorted.bam
bedtools intersect -abam sample.ex_3p_cl.sorted.bam -b 5pUTRexons.bed -v > sample.ex_5p_cl.sorted.bam
bedtools intersect -abam sample.ex_5p_cl.sorted.bam -b rRNA.bed -v > sample.hg19.cleaned.bam
```

The number of mapped reads post cleaning were again counted using the approach mentioned above, and all files were then subsampled to the minimum number of reads after shuffling the reads, followed by converting the files into *bedGraph* format in order to allow visualization in Integrated Genome Browser :

```
BAMS=./*.hg19.cleaned.sorted.bam
for sample in $BAMS
do
echo "$sample now being normalized to 9348088 reads"
samtools view -H $sample > ./sample.header.txt
samtools view -F 4 $sample | grep -v '^@' - | shuf - | head -n 9348088 | cat $sample.header.txt - | samtools view -bS - > $sample.norm.bam"
samtools sort $sample.norm.bam -o $sample.read.normalized.bam"
samtools index $sample.read.normalized.bam"
echo "$sample now has the following number of reads:"
samtools view -c -F 4 $sample.read.normalized.bam"
genomeCoverageBed -ibam $sample.read.normalized.bam -g hg19.chrom.sizes -bg > $sample.read.normalized.bedGraph"
done
```

3.6.4 RNA Pol II Elongation Rate calculations

RNA Pol II elongation rates from 4sUDRB-Seq were computed using *TERate* (Zhang et al., 2016). The intronic regions of each gene was divided into 300-bp non-overlapping bins, and the read density in each 300-bp bin was calculated. Such bins within the promoter region were randomly selected and treated as “expressed bins.” Another set of such bins within the region around the TES of long genes (>60 kb) was randomly selected for “background bins”. Yet another class of bins, called “transcribed bin” was selected as the bin for which the read density of expressed bins was higher than that of the background bin. For each gene on which the elongation rate was calculated, the position of the last two consecutive “transcribed bins” was defined as the transcription “wave

front”. Thus, the distance transcribed within the specific release time for a gene was calculated from its TSS to the transcription “wave front”, and the Pol II elongation rate was predicted by dividing the transcribed distance by the release time of the specific experiment.

An example of this implementation for 10-minute release time is shown below:

```
cp sample.bedGraph TERate
cd TERate
./gene_to_window refFlat.txt 300 > refFlat_bins.txt
mkdir split
cd split/
sh ../split_bedgraph.sh ../sample.bedGraph
sh ../split_refFlat.sh ../refFlat_bins.txt
ls |grep "bin" |awk -F"_" '{print "nohup ../bedgraph_to_hits "$1"_bedgraph.txt "$1"_bin.txt > "$1"_hits.txt &"}' |sh
cd ../
cat split/*_hits.txt > sample.txt
sort -k4,4 -k1,1 -k2,2n -k3,3nr sample.txt > sorted_hits_sample.txt
./calculate_TER sorted_hits_sample.txt 10 300 |sort -k1,1 -k4,4nr |awk '{a[$1,++b[$1]]=0}END{for(i in b)print a[i,1]}' > TERate_output_sample.txt
```

3.6.5 Density plots and heat maps

Average density plots were generated from strand-based normalized *bam* files with *NGStools* (Shen et al., 2014) with the gene lists as specified for each graph. Next, the density files for sense and anti-sense reads were collected and averaged for all genes in the gene group over both strands and re-plotted in *R*. When replicates were normalized to individual number of reads between each condition, the resulting densities as mentioned here were then re-normalized to same read density for all curves under comparison. The same rule was implemented to generate metagene plots.

For heatmaps, the sorting parameter used was either gene length or Pol II elongation rates as measured by independent 4-min and 8-min release 4sUDRB-Seq experiments in U2OS cells (unperturbed for MYC levels), and only the genes for which the elongation rates could be measured in all experiments were considered ($n = 2,163$). Heatmap generation was carried out using *deeptools*. BAM files were converted to *bigwigs* with a bin size of 50 using *bamcoverage*. *computeMatrix* was utilized with *reference-point* as TSS and bin size of 50 to generate a read and position based matrix in a sorted manner; *-sortRegions* parameter was set to *keep* and the sorted reference file (sorted for gene length or speed) was used as the input for *-R* parameter. The resulting matrix was then used for *plotHeatmap* command with same *-zmin* and *-zmax* values for the set of conditions under comparison. A representative command for the above description is as follows:

```

/home/user/anaconda3/bin/computeMatrix reference-point --referencePoint TSS \
  -a 120000 -b 1500 \
  -bs 5 \
  -p max \
  --sortRegions no \
  -R references.bed \
  -S \
  sample.bw \
  -out sample.txt.gz

/home/user/anaconda3/bin/plotHeatmap -m sample.txt.gz \
  -out sample.pdf \
  --zMin 0 \
  --whatToShow 'heatmap and colorbar' \
  --colorMap Blues \
  --sortUsing region_length \
  --regionsLabel 'Expressed_Genes' \
  --samplesLabel Control \
  --labelRotation 45 \
  --zMax 1

```

3.6.6 Read Density-based score calculations

Gene based scores were calculated by dividing reads in a specific region of the gene with other. Specifically, processivity scores were calculated from stranded *bam* files from 4sU-Seq data, where the sense strand *bam* files were converted to *bed* files:

```
bamToBed -i sample.plus.bam > sample.plus.bed
```

coverageBed from *bedtools* was used to calculate read densities falling in proximal and distal gene regions:

```
coverageBed -a proximal.txt -b sample.plus.bed > sample.plus.proximal.txt
coverageBed -a distal.txt -b sample.plus.bed > sample.plus.distal.txt
```

The proximal region was demarcated as the region from TSS+1 kb to TSS+2 kb, whereas the distal region was marked to be from TSS+5 kb to TSS+7 kb. The processivity ratio was defined as the \log_2FC of pseudo-count added reads in the distal and proximal regions. Only genes longer than 8 kb were considered for this calculation. In case of calculation for directionality score, the same *bed* files as above were used. Sense reads were retrieved similarly with *coverageBed* and the pseudo-counted reads were divided by gene length to overcome bias towards long genes, while for antisense reads the region 1 kb upstream of TSS, but for the reads from the opposite strands was used. Directionality ratio was defined as the \log_2FC of sense reads to antisense reads. Very small genes (< 0.3 kb) were excluded from the analysis.

3.6.7 Gene Differential Expression Analysis

In order to calculate gene regulation (in terms of \log_2FC), stranded BAM files were used to calculate read counts per gene. For each replicate, these reads were normalized for gene

length and sequencing depth. Reads in exonic regions were not considered. Expression filter was set based on reads per kilobase per million (RPKM) averaged over all samples. Fold change was calculated as \log_2 ratio of averaged reads from replicates, and significance was calculated with Benjamini-Höchberg algorithm, implemented via limma package in R.

4 RESULTS

4.1 4sUDRB-Seq reveals wide variation in elongation rates

In order to measure the impact on MYC on elongation, precise measurement of elongation rates was required. 4sUDRB-Seq is a technique in which all the Pol II in cells are synchronized to stay paused at the promoters by reversible CDK9 inhibition using DRB. Release of this pause and allowing the Pol II to transcribe for a specific time period can be achieved by washing the cells with PBS, and during this time, the nascent RNA is concomitantly labelled with 4sU, thus allowing precise measurement of distance transcribed on the genes in the specific release time. Since the concentration and time for which the cells need to be treated with DRB in order to achieve this reversible Pol II block varies from cell line to cell line, it was required to carry out a pre-experiment to test these factors. Additionally, the Pol II elongation rates also differ between cell lines (Fuchs et al., 2015), and thus the release time also needs to be checked in a way that it allows Pol II to travel without transcribing the whole gene.

To test these parameters, gene-based DRB assay employing qPCR with intronic primers to capture nascent RNA was performed on the *OPAI* gene. This was performed on intronic regions close to exons which were 2, 22 and 28 kb downstream of the TSS and the release was performed for 4, 8, 16, 24 and 32 minutes (Fig 4.1A). Expression of nascent RNA was calculated as fold changes (FC) over non-DRB treated (NT) condition. In case of promoter proximal primer pair (2 kb), all release time points showed higher expression, whereas in case of primers for 22 kb exon, higher expression than NT was shown only after 24 minutes of release. Similarly, for primers against 28 kb exon, higher expression than NT was achieved after 32 minutes, indicating Pol II speed of around 1 kb/min. This experiment allowed the deduction of several points with respect to U2OS cell line.

Firstly, 2.5 hours of DRB treatment is sufficient to block the transcription since the relative expression was least in 0 min release time point. Secondly, the DRB induced block of transcription is reversible and release time of as less as 4 min can be used in

Results

U2OS cells to deduce Pol II elongation rates. Finally, it can be seen that Pol II advances into gene body at a more or less constant rate, and thus, this technique can be implemented genome-wide to measure Pol II elongation rates.

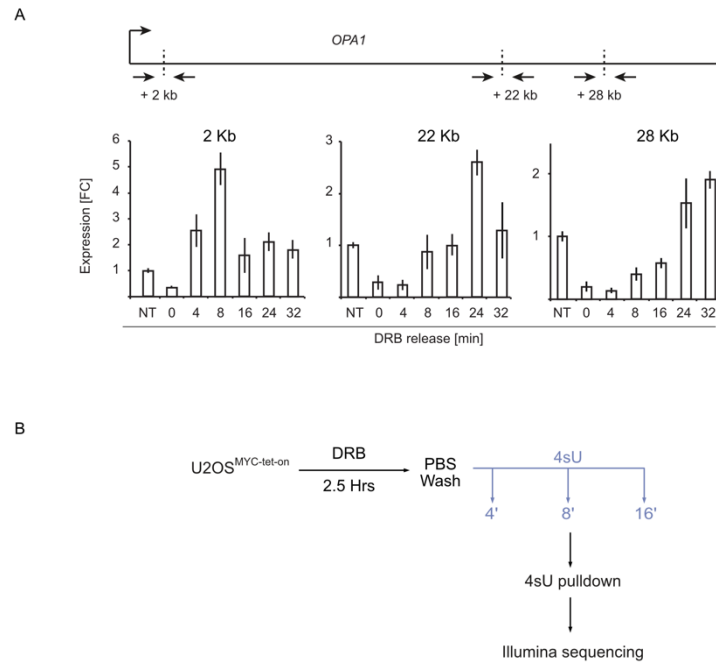


Figure 4.1 DRB assay for measuring elongation rates.

(A) Analysis of OPA1 nascent RNA in U2OS cells, immediately after 2.5 hrs DRB treatment and at the indicated release time points after DRB removal. NT = non-treated. Pre-mRNA levels were determined by qRT-PCR, employing intronic primers. All values were normalized to B2MG in the same sample. Levels in NT samples were set as 1. Bars indicate averages of three technical replicate experiments; error bars represent standard deviation. (B) Scheme depicting the 4sUDRB-Seq experiment carried out to measure elongation rates in unperturbed cells.

In order to confirm if the release times ranging from 4 min to 16 min were sufficient to capture the elongation rates genome-wide, 4sUDRB-Seq was carried out at 4 min, 8 min and 16 min release time points (Fig 4.1B). Newly synthesized RNA was labelled by 4sU during these release times, and purified after biotinylation using streptavidin-based capture, and subjected to Illumina sequencing. The sequences base-pairs when aligned to the human genome, reveals the pile-up from density of reads. These reads thus represent the region of the gene transcribed within the time-interval of 4sU treatment and DRB release.

The location along the gene where read density approaches background levels can be considered as the “transcriptional wave front”, which indicates the genomic distance that Pol II transcribed during the release interval. Inspection of individual genes on the

Integrated Genome Browser (IGB) showed that the distance travelled by Pol II after DRB washout could be clearly measured and this distance increased with the longer release times (Fig 4.2). This distance, when divided by the time of DRB release, gives elongation rate of the respective gene in that experiment. For example, for *TOP1*, since the transcription wave front can be approximated at 11 kb, 18kb and 35 kb for 4 min, 8 min and 16 min release experiments, the elongation rate in these experiments can be deduced as 2.75, 2.25 and 2.19 kb/min, thus showing a high degree of correlation between the experiments.

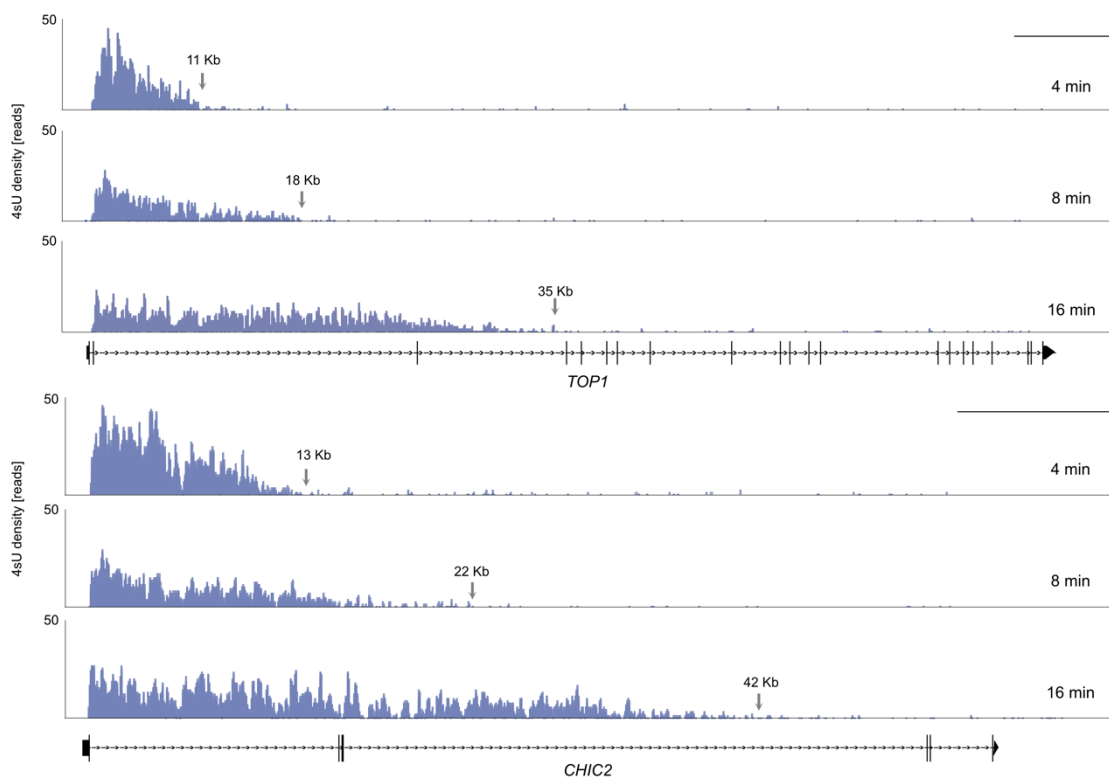


Figure 4.2 4sUDRB-Seq can be used to calculate Pol II elongation rates.

IGB (Integrated Genome Browser) track of 4sU labelled RNA indicating the transcriptional wave front for *TOP1* and *CHIC2* genes with the indicated DRB release times. Horizontal lines above gene names show the genomic location from TSS to TES, with vertical lines representing the exon locations. Scale bars represent 10 kb, and arrows with values indicate the approximate distance from TSS. Parts of this figure appear in a similar form in (Balupuri et al., 2019).

Additionally, these results were not restricted to single genes. In order to visualize the transcriptional wave fronts for multiple genes, reads from all expressed was averaged and displayed (Fig 4.3A) for the given experiments. Here, only long genes were considered, since in short genes, the transcriptional wave front would already cross the TES for the given release time points. Such averaged density plot showed that as the release time

increased, the wave front averaged over all genes reached background levels at a later position in the gene body, showing that the genomic distance which Pol II can transcribe is directly proportional to the time of release from DRB block. However, such effects could be driven by few highly expressed genes. To rule out such a possibility, heat maps showing the wave front were generated and sorted for gene length, with the longest genes on top and shortest at the bottom (Fig 4.3B). Such an analysis revealed that on all expressed genes, the transcription wave front covered longer distances into the genes for 8-min release compared to the 4-min release time.

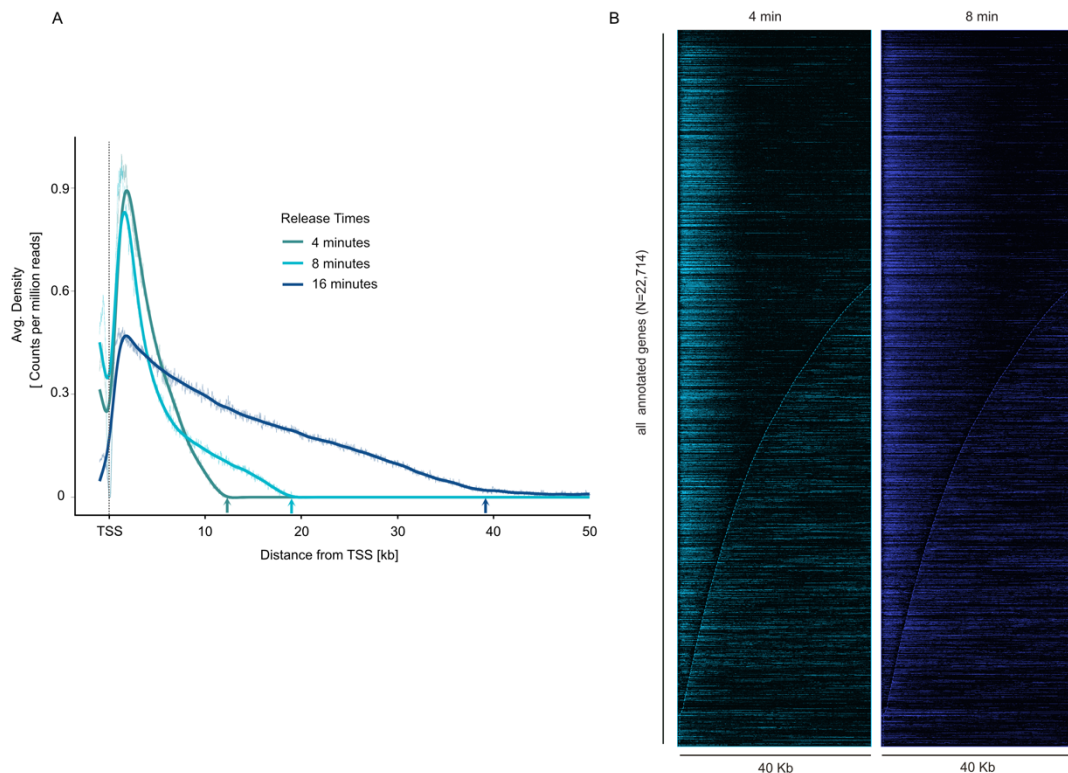


Figure 4.3 4sUDRB-Seq transcription wave fronts are a function of release times.

(A) Averaged reads' density plot of 4sUDRB-Seq reads comparing various release times for all expressed genes. Arrows indicate the location where wave-front approaches background levels after the indicated DRB release times. (B) Heatmap sorted for gene length showing 8 mins release times cover greater distance as compared to 4 mins. The dotted line marks the gene ends.

In order to quantify the absolute Pol II elongation rates in terms of kilobases travelled per minute (Fig 4.4) a previously published algorithm (Zhang et al., 2016) was utilized and implemented on this 4sUDRB-Seq experiment. These speeds showed marked variations in Pol II elongation rates among the genes, and this variation was found to be consistent among the various release experiments. In other words, the genome consists of genes

Results

which can be transcribed in terms of bases/min faster than the other, and the classification of such genes remains consistent throughout the different release times.

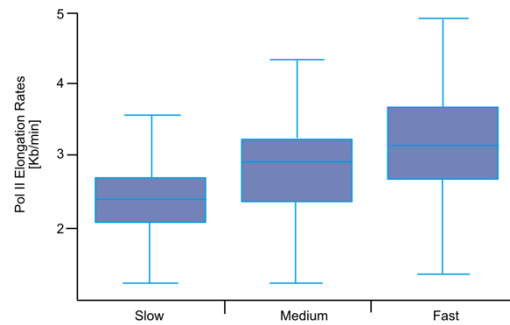


Figure 4.4 Elongation rates from 4sUDRB-Seq

Pol II elongation rates from TERate algorithm 4 min after release from DRB inhibition. Genes were classified as “Slow”, “Medium” and “Fast” based on elongation rate measurements from an independent experiment (8 min after DRB release). This figure appears in a similar form in (Balupuri et al., 2019).

These experiments show that 4sUDRB-Seq can be used to generate reliable and reproducible elongation rates in U2OS cells to measure changes in the elongation rates of Pol II transcription.

4.2 Cellular system to study primary effects of MYC

Further, in order to find out if the elongation rates as calculated by 4sUDRB-Seq are a function of MYC levels, a combination of siRNA mediated depletion and Tet-on mediated overexpression in U2OS cell line was utilized to acutely manipulate MYC levels in these cells.

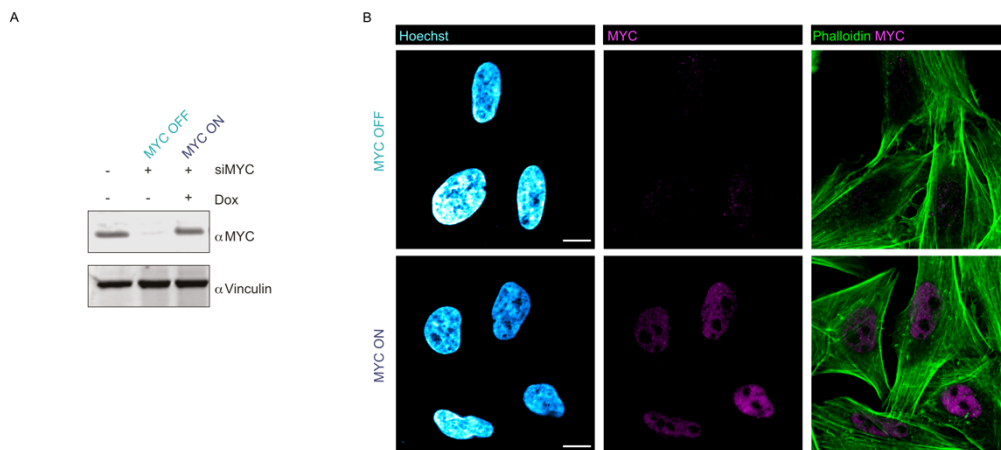


Figure 4.5 Cellular system to establish acute changes in MYC levels.

(A) Immunoblot of MYC in U2OS^{MYC-Tet-On} cells before (MYC OFF) and after re-establishment of MYC concentration at physiological levels (MYC ON) using combination of Dox and siRNA against MYC (see “Methods”). Vinculin is shown as loading control. (B) Immunofluorescence images captured by confocal microscope from indirect immunostaining against MYC in MYC ON and MYC OFF conditions. (blue: Hoechst-stained nuclei, magenta: MYC, green: phalloidin; scale bar: 5µm). This figure appears in a similar form in (Balupuri et al., 2019).

In this experimental set-up, the U2OS^{MYC-Tet-On} cells were used, and the addition of doxycycline (Dox) was preceded by transfection of cells with siRNA against MYC. Thus, MYC concentration in these cells were reduced to very low levels first, followed by re-establishment of endogenous MYC levels within 12-16 hours by addition of Dox. The cells in this condition were termed as “MYC ON” and were compared to cells in which Dox was not added after siRNA treatment (“MYC OFF”).

Upon comparing the levels of MYC through immunoblotting, it was found that the concentrations of MYC in doxycycline-treated cells matched endogenous levels (Fig 4.5A). This cellular system thus brings about acute changes in MYC levels, and counterbalances the off-target effects of siRNA treatment, if any. To confirm the cellular distribution of MYC, indirect immunostaining against MYC in paraformaldehyde fixed cells was performed followed by imaging under confocal microscope. Similar to earlier noted cellular distribution (Lorenzin et al., 2016), MYC showed uniform nuclear localization and the changes in its level were not driven by just a few cells, but were found to be evenly distributed across the culture (Fig 4.5B).

These results demonstrate that this new experimental set-up with MYC-ON and MYC-OFF conditions can be used to manipulate MYC levels acutely and uniformly in the U2OS cells, and is thus suitable for analysing the elongation rates of Pol II.

4.3 Elongation Rates of Pol II are regulated by MYC levels

MYC being a protein with fast turnover rate of synthesis and degradation, it was possibly that MYC levels could be affected after DRB treatment since DRB blocks mRNA synthesis by Pol II. Therefore, to rule out the possibility that MYC transcription itself was hampered by DRB treatment, it was required to check if MYC levels are unperturbed by short term DRB treatment or not.

Results

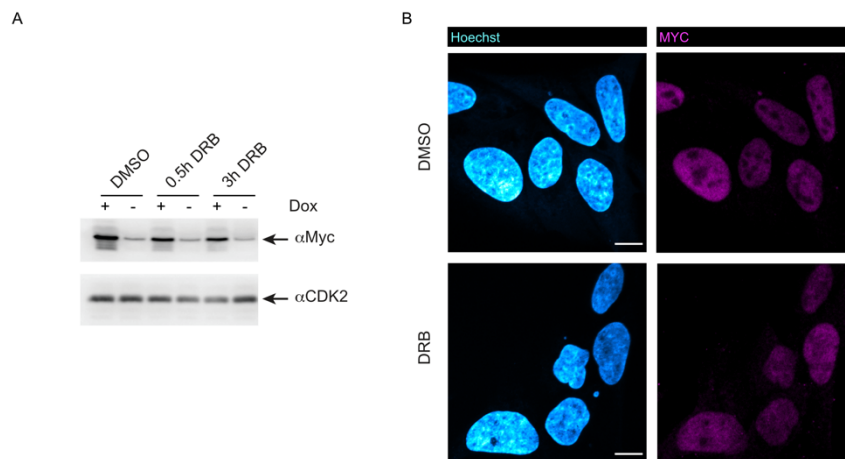


Figure 4.6 MYC levels remain unaffected by DRB treatment.

(A) Immunoblot showing comparison of MYC levels in various durations of DRB treatment along with 16 hours of Dox treatment. CDK2 is shown as loading control. (B) Immunofluorescence images captured by confocal microscope from indirect immunostaining against MYC in MYC ON and MYC OFF conditions. (blue: Hoechst-stained nuclei, magenta: MYC; scale bar: 5 μ m). This figure appears in a similar form in (Balupuri et al., 2019).

Towards this end, immunoblotting in U2OS^{MYC-Tet-ON} cells was carried out against MYC to check for MYC levels after DRB treatment for 2.5 hours in presence and absence of Dox, and compared to levels in cells in which the DRB treatment was carried out only for 0.5 hour. The results showed that the change in MYC levels upon Dox treatment was found to be similar among the cells which were treated with DRB as compared to the ones which were treated with DMSO (Fig 4.6A). In order to check the distribution of MYC within the nuclei, indirect immunostaining was performed and viewed under confocal microscope. Images from this experiment neither showed any redistribution of MYC within the nuclei, nor showed major variations of MYC levels among the cells (Fig 4.6B).

Since the cellular system defined above allowed to study primary effects of MYC, 4sUDRB-Seq was performed at two different release times after DRB treatment on cells from MYC ON and MYC OFF conditions. The cells were released from DRB block for 10 minutes with a concomitant labelling of nascent RNA, and the transcription wave front of Poll II was compared between MYC ON and MYC OFF conditions.

Results

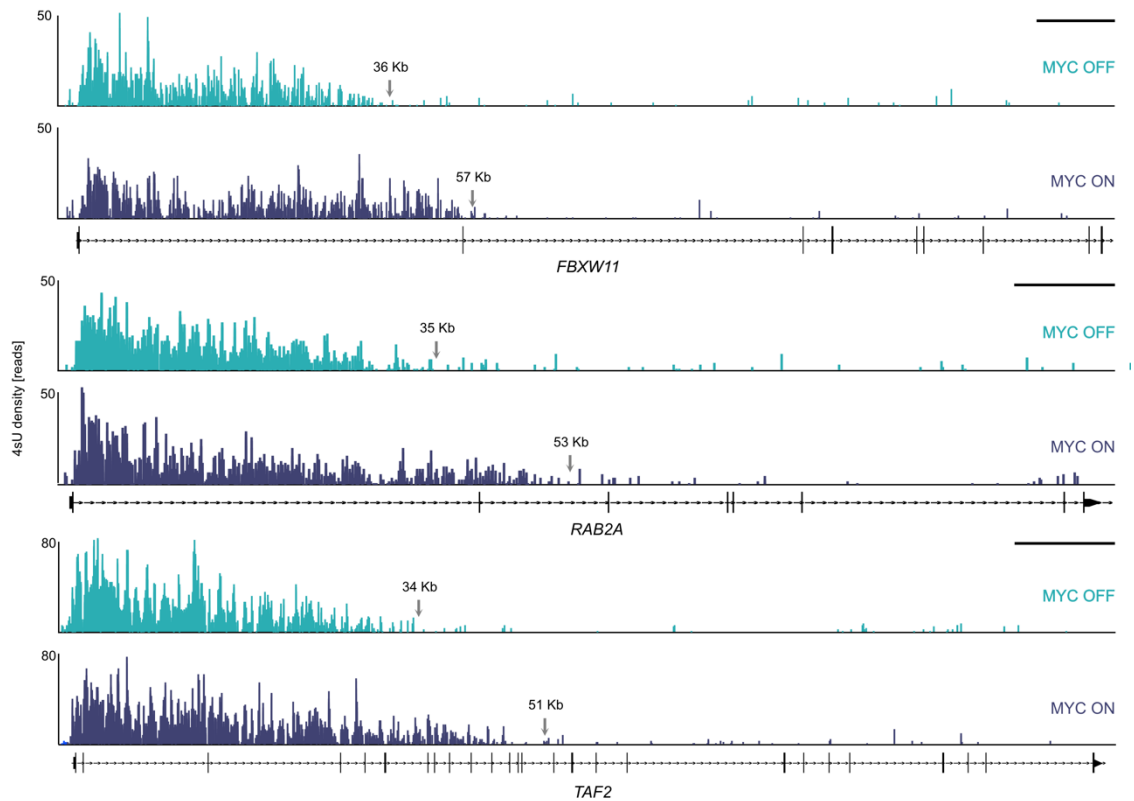


Figure 4.7 MYC is required to maintain fast elongation rates.

Genome browser tracks of *FBXW11*, *RAB2A* and *TAF2* genes indicating the transcription wave front in MYC ON (dark blue) condition compared MYC OFF (light blue) condition from 10 mins release DRB4sU-Seq. Scale bars represent 10 kb, and arrows with values indicate the approximate distance from TSS. This figure appears in a similar form in (Balupuri et al., 2019).

Upon inspection of multiple individual genes (Fig 4.7), on one hand, the density of nascent RNA was found to be unaltered near the TSS. On the other hand, the transcriptional wave front in MYC OFF cells was found to be significantly behind when compared to MYC ON cells. In order to confirm if this was a global effect, 4sU read densities were averaged and plotted for all expressed genes between the length of 50 kb to 100 kb.

Results

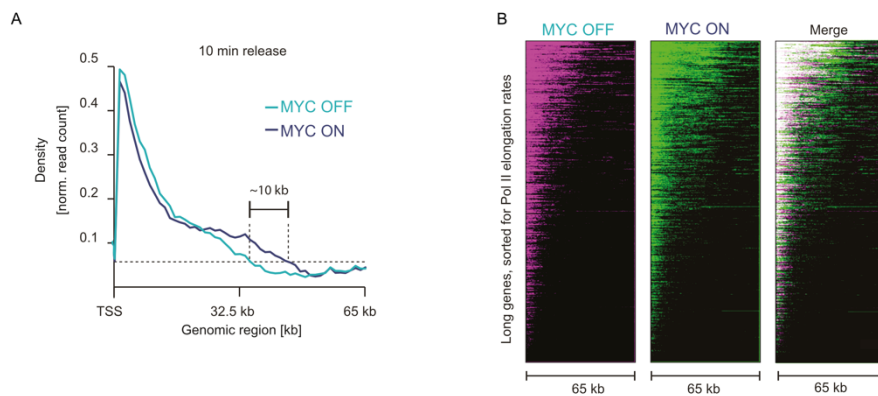


Figure 4.8 MYC is required globally to maintain fast elongation.

(A) Density of reads from 4sUDRB-Seq followed by 10 mins release time, averaged for 3,732 genes (50–100 kb long). (B) Heatmaps of reads from same experiment, aligned to TSS and the genes were then sorted for Pol II velocity calculated from an independent 4 min unperturbed condition experiment shown in Fig 4.4. This figure appears in a similar form in (Balupuri et al., 2019).

This comparison of average densities (Fig 4.8A) revealed that the transcription wave front in absence of MYC lagged by 10 kb in the 10 mins release time. However, to rule out the possibility that such an effect was driven by just a few highly expressed genes, heat maps depicting the transcriptional wave-fronts were generated. Comparison of the wave fronts in the heat map showed that the genes that showed the maximum lag in wave front in absence of MYC, were the ones which had highest elongation rates as well (Fig 4.8B). Importantly, the decrease in rate of elongation was not driven by a few genes, but found to be all across the heatmap.

Next, an additional experiment with a DRB release time of 15 mins was carried out to test if this effect was restricted only to a specific release time. Upon inspection of browser tracks, similar results were noted as in case of 10 mins release (Fig 4.9). Whereas the transcription wave front was found to be ahead by approx. 19 kb on *PPP1R12A* gene, the comparison of averaged densities revealed a difference of approx. 11 kb among the wavefronts from MYC OFF and MYC ON conditions.

Results

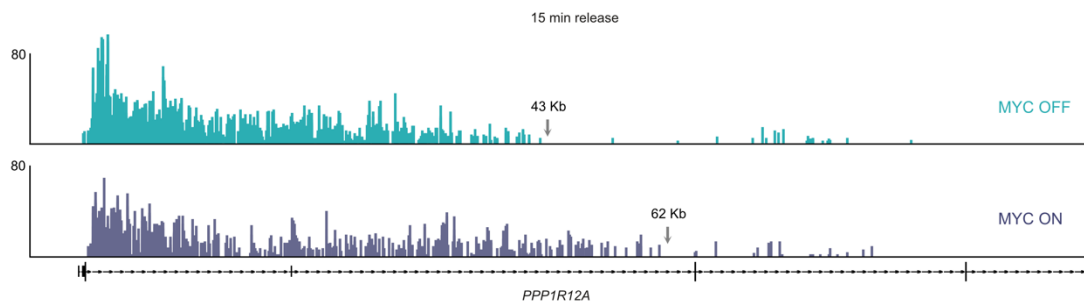


Figure 4.9 Changes in elongation rate by MYC are independent of release times.

Genome browser tracks of *PPP1R12A* gene indicating the transcription wave front in MYC ON condition compared MYC OFF condition in 15 mins release DRB4sU-Seq. Scale bars represent 10 kb, and arrows with values indicate the approximate distance from TSS.

In order to quantify the reduction in elongation rates upon MYC depletion over each gene, TERate algorithm was used to calculate the rates in terms of bases/minute. This quantification for the 10 min release experiment revealed that in presence of MYC, the average elongation rate was found to be 3200 bases/minute, which dropped down to 2700 bases/minute in absence of MYC. The elongation rates captured over all genes when compared for 10 minutes release experiment showed significant reduction in elongation rate in MYC OFF condition (Fig 4.10A). Importantly, when a ratio between the rates from MYC ON and MYC OFF cells was calculated, it was found to correlate significantly to a high degree between the 10- and 15-minutes release experiments (Fig 4.10B) indicating a dependence of Pol II elongation rates on MYC levels. The results from the above experiments suggest that the reduction in the ability of Pol II to transcribe at a fast rate happens far away (Fig 4.8A, 4.9B) from the promoter, where MYC is known to be bound (Rahl et al., 2010) (Walz et al., 2014).

Since DRB treatment synchronizes all Pol II at the start sites, changes in pause release cannot account for the reduction in elongation rate in MYC OFF condition. Hence, a deeper investigation into the Pol II elongation complex was required to understand how MYC maintains fast elongation rates of Pol II.

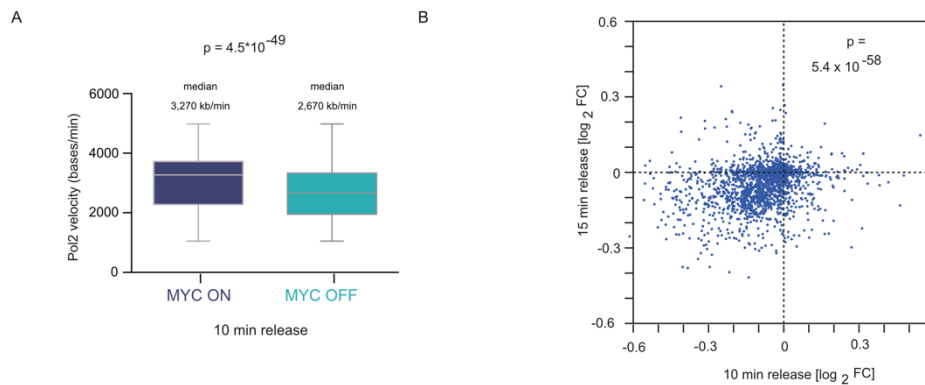


Figure 4.10 Pol II elongation rates are a function of MYC levels.

(A) Boxplot showing a comparison of elongation rates in MYC ON and MYC OFF conditions for 10 minutes release experiment (B) Scatterplot showing the correlation between the ratios of elongation rates (in MYC ON and OFF) for the 10- and 15-minute release experiments. Statistical significance measured by two-sided Wilcoxon test. Parts of this figure appear in a similar form in (Balupuri et al., 2019).

4.4 MYC is required for assembly of Pol II Elongation Complex

One of the possibilities how MYC could alter the elongation rates would be if it assists in the assembly of factors on Pol II before the elongation phase of transcription cycle begins. However, the number of such factors is quite high, and required a systematic unbiased study.

4.4.1 MYC dependent recruitment of SPT5 to Pol II on chromatin

In order to check if MYC affects the interactome of Pol II while bound to chromatin, a series of mass spectrometry-based experiments were carried out by my colleague where the interactome of MYC and Pol II was compared (Fig 4.11A) (experiments performed and analysed by Julia Hofstetter, Department of Molecular Biology and Biochemistry, University of Würzburg). These experiments suggested that in absence of MYC, SPT5 and SPT6 proteins cease to interact with chromatin bound Pol II.

In terms of techniques used to confirm interaction of other proteins with MYC as predicted by mass spectrometric analysis, techniques like co-immunoprecipitation (Co-IP) and proximity ligation assay (PLA) have been used (Jaenicke et al., 2016) (Gerlach et al., 2017). While Co-IP is a definitive proof of interaction between proteins, PLA utilizes *in situ* PCR between oligomers ligated to antibodies against the two proteins in paraformaldehyde fixed cells. Since the fluorophores added in a subsequent step are linked to oligos hybridizing to sequences in the amplicons, bright fluorescent dots can be

seen under the microscope with proper excitation, and counter-staining with nuclear or cytoplasmic dyes (Hoechst and phalloidin, respectively) offers spatial information within the cell while being able to give quantitative comparisons between different conditions.

Based on the above-mentioned arguments, PLA with antibodies against SPT5 and transcriptionally engaged form of Pol II (pSer2 Pol II) was performed in U2OS MYC ON and MYC OFF system. Upon inspecting confocal images of cells from MYC ON condition, it was found that PLA foci, which are indicator of proximity between the respective antibodies, were limited to nucleus in the cells (Fig 4.11B). Since the transcriptionally engaged Pol II is restricted to nucleus, presence of foci in nucleus was not surprising, but showed the specificity of the antibodies and the technique.

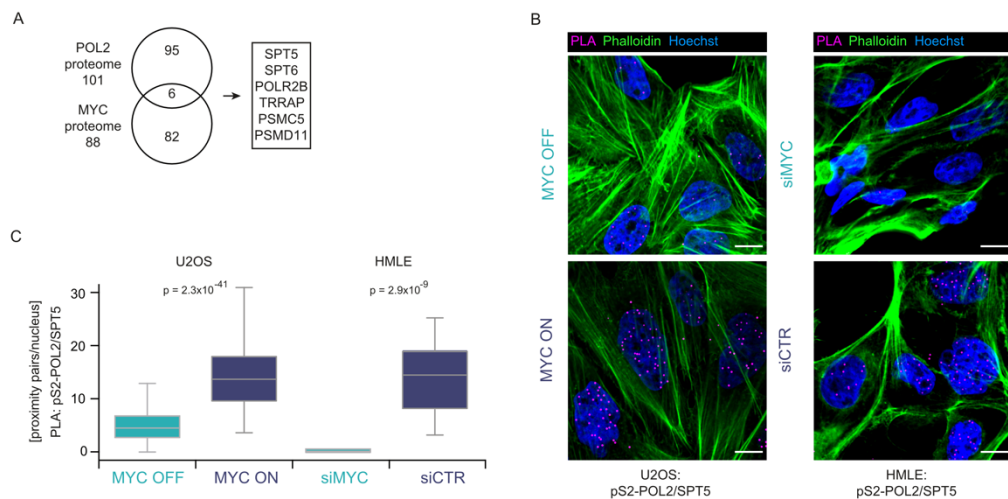


Figure 4.11 MYC driven recruitment of SPT5 to Pol II

(A) Venn diagram depicting the intersection between the Pol II interactome and the MYC interactome. Interactome experiments and analysis were performed by Julia Hofstetter. (B) Immunofluorescence images taken using confocal microscope between pSer2 Pol II and SPT5 in cells treated with (MYC ON) or without MYC (MYC OFF) for U2OS and HMLE cell lines (C) Quantification of PLAs shown in (B). Statistical significance measured by two-sided Wilcoxon test. This figure appears in a similar form in (Balupuri et al., 2019).

However, the number of these foci decreased drastically in images of cells from MYC OFF conditions, and were found to be significantly lower while comparing the quantifications from multiple experiments (N=3) and cells (approx. 100 cells) (Fig 4.11C). Additionally, these results were not restricted to U2OS MYC ON and OFF system, but were found in HMLE cells as well where MYC levels were altered using siRNA targeted against MYC and compared to siRNA with non-targeting sequences (Fig

4.11 B, C). These results show that interaction of SPT5 and Pol II is a function of MYC levels, and this holds true across multiple cellular systems.

However, one potential factor that could explain the above results, would be the reduction in the protein levels of SPT5 and/or pSer2 Pol II. To rule out this possibility, pSer2 Pol II and SPT5 levels in the cells fixed with paraformaldehyde were checked using indirect immunostaining (Fig 4.12A) as well as by quantification of immunoblots from multiple experiments and repetitions (Fig 4.12B). In both experimental setup no significant change in the respective proteins was detected. Since the levels of SPT5 and pSer2 Pol II were constant, it can be concluded that it is actually the levels of MYC that regulate SPT5-Pol II proximity in cells.

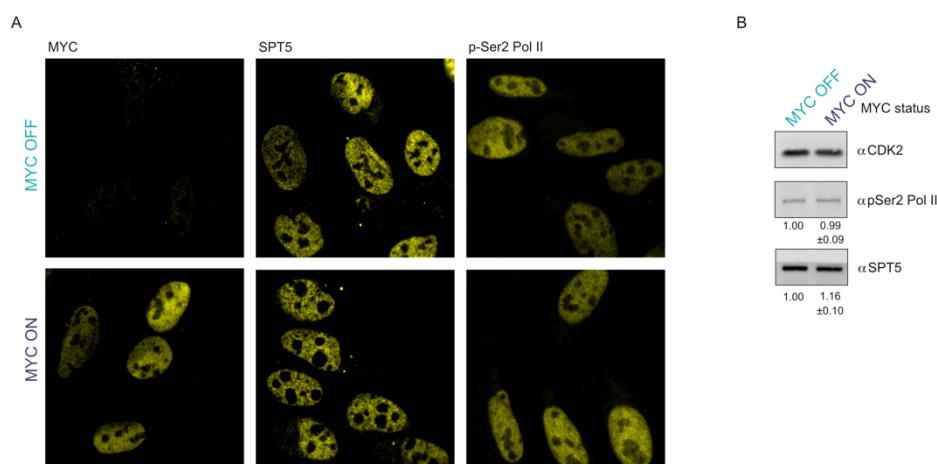


Figure 4.12 SPT5 and pSer2 Pol II levels are unaffected by MYC

A) Immunofluorescence images taken on confocal microscope from indirect immunostaining of SPT5, MYC and pSer2 Pol II in presence (MYC ON) and absence of MYC (MYC OFF). (B) Immunoblot of MYC, SPT5 and p-Ser2 Pol II from cells shown in (A). Numbers below indicate the ratio of each protein to MYC OFF condition from at least three experiments. CDK2 was used as a loading control. This figure appears in a similar form in (Balupuri et al., 2019).

4.5 MYC mediated SPT5 handover is driven by CDK7 via TFIIIE

The fact that MYC interacts with SPT5, and this interaction is instrumental in loading SPT5 to Pol II raises the question how this can occur mechanistically. One possibility is that when MYC binds to the promoters, it recruits SPT5 along with it and hands SPT5 over to Pol II upon transcription initiation. Since the transcription process is heavily regulated by CDKs (see 1.3.1), it is possible that they also mediate this transfer of SPT5 from MYC to Pol II.

Since PLA can be used to measure the proximity between SPT5 and Pol II, this can be used as a readout to check whether Pol II elongation complex contains SPT5 or not. This readout can be used to study the transfer of SPT5 to Pol II and its dependence on CDKs.

4.5.1 CDK7 is responsible for transfer of SPT5 from MYC to Pol II

In order to probe which CDK was involved in handing over SPT5 from MYC to Pol II, a screening approach with a small library of kinase inhibitors was used (see 1.3.2). In this approach, first, PLA between MYC and SPT5 was carried out and then compared with results of PLA between SPT5 and Pol II carried out separately after inhibition of individual CDKs.

However, a key requirement for such an assay to work accurately would be similarity in the protein levels of MYC and SPT5 upon CDK inhibition. To test this, paraformaldehyde fixed cells were subjected to indirect immunofluorescence imaging against SPT5 and MYC, and the resulting confocal images were compared between cells treated either with pan-CDK inhibitor DRB or DMSO (4.13A). Also, the levels of both proteins were also measured by quantifications of immunoblots under same conditions (Fig 4.13B). These experiments did not show any significant changes in levels or nuclear distribution of MYC and SPT5 upon pan-CDK inhibition, thus showing that the changes noted PLA signals are caused by changes in proximity of respective proteins and are not a function of proteins levels in the cells

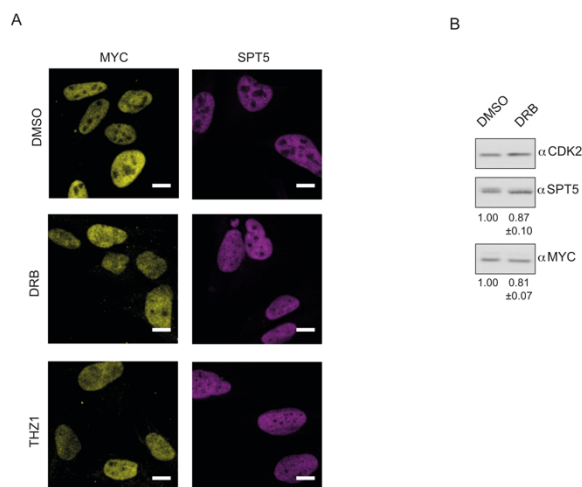


Figure 4.13 SPT5 and MYC levels are unaffected by CDK inhibition

Results

A) Immunofluorescence images taken on confocal microscope from indirect immunostaining of SPT5 (magenta) and MYC (yellow) after treatment of U2OS cells with various transcription inhibitors (scale bars: 5 μ m). (B) Immunoblot of MYC and SPT5 in U2OS cells treated with pan CDK inhibitor DRB. CDK2 is shown as a loading control. This figure appears in a similar form in (Balupuri et al., 2019).

To test CDK involvement, the above-mentioned PLA experiments were performed in biological triplicates on paraformaldehyde fixed U2OS cells which were treated with various CDK inhibitors for 4 hours and compared to each other, along with U2OS cells which were treated with DMSO. If the handover of SPT5 from MYC to Pol II would be hampered by a certain inhibitor, increased PLA signal between MYC and SPT5 with a concomitant decrease in signal between SPT5 and Pol II would be noted.

Upon observing the changes in PLA signals induced by DRB (a pan-kinase inhibitor), PLA signal for MYC/SPT5 was found to be increased (Fig 4.14 A,C) whereas the signal for SPT5/Pol II antibody pair decreased (Fig 4.14 B,D), indicating that indeed one of the CDK was responsible for the handover of SPT5 from MYC to Pol II. When more specific CDK inhibitors THZ1 and LDC4297 (which inhibit CDK7) were used, similar results were seen. However, when using a CDK9 inhibitor (LDC67), no increase of PLA signal between MYC and SPT5 was noted (Fig 4.14 C). These results indicate that specifically, CDK7 was responsible for handover of SPT5 from MYC to Pol II.

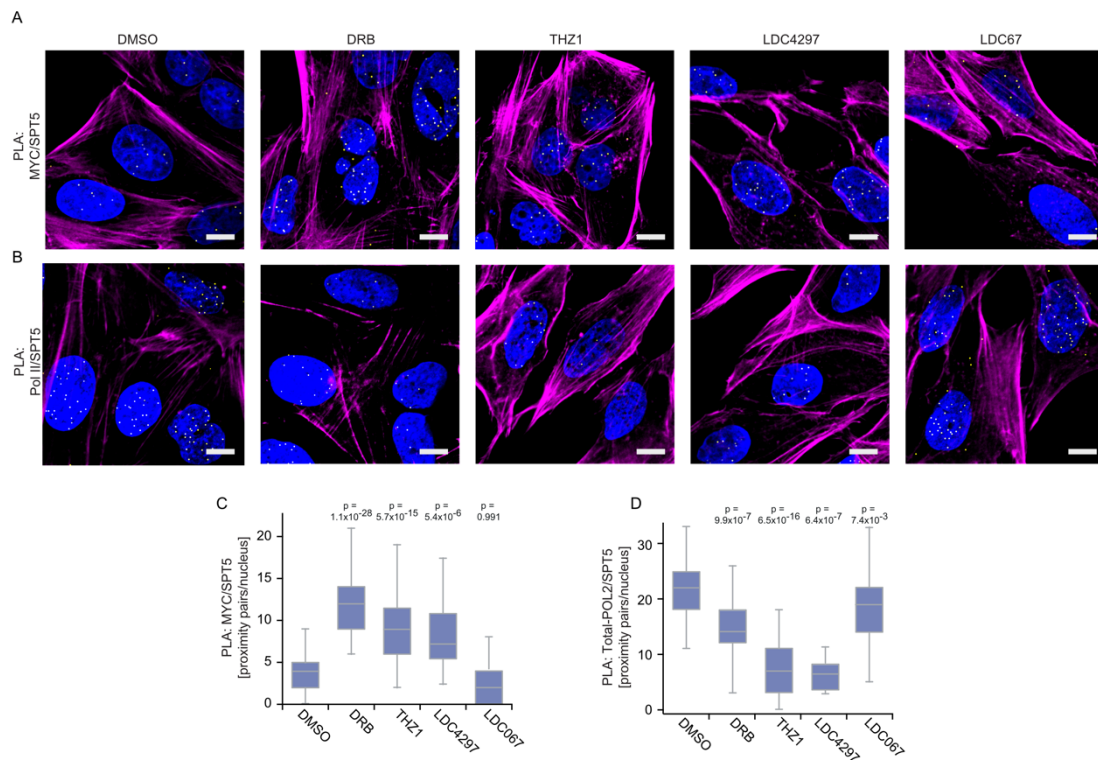


Figure 4.14 CDK7 is responsible for transfer of SPT5 from MYC to Pol II

Results

(A and B) Immunofluorescence images taken from confocal microscope between MYC and SPT5 (A) or SPT5 and total Pol II (B) in cells treated with CDK inhibitors (DRB, THZ1, LDC4297, LDC067) or DMSO as control. (C) Quantification of PLAs shown in (A). The quantification of proximity pairs upon inhibitor treatment was calculated in independent experiments and normalized to the DMSO condition. (D) Quantification of PLAs shown in (B). The quantification of proximity pairs upon inhibitor treatment was calculated in independent experiments and normalized to the DMSO condition. Statistical significance measured by two-sided Wilcoxon test. This figure appears in a similar form in (Baluapuri et al., 2019).

In order to confirm this results in an experiment where CDK7 levels are genetically manipulated instead of chemically inhibiting it, siRNA against CDK7 (siCDK7) was used to deplete the levels of the protein in the U2OS cells. The same set of PLA experiments as above was performed in U2OS cells which have been treated with siCDK7 for 48 hours and compared with cells which were treated with non-targeting siRNA (siCTR). The MYC/SPT5 signals in siCDK7 treated cells were again found to be higher while SPT5/Pol II PLA signal were lower as compared to those in siCTR treated cells (Fig 4.15 A-C), thus showing that CDK7 indeed mediates the transfer of SPT5 from MYC to Pol II.

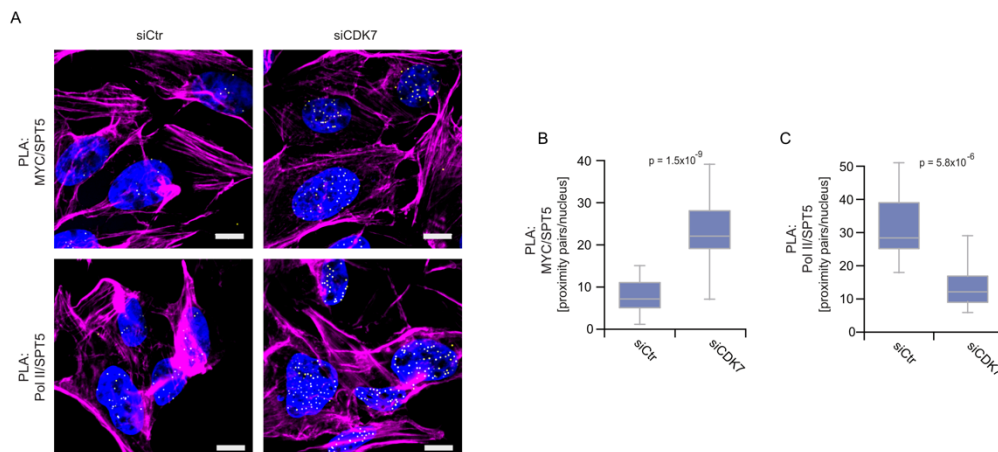


Figure 4.15 CDK7 depletion prevents SPT5 handover to Pol II by MYC

Immunofluorescence images taken from confocal microscope between SPT5 and MYC (top) or between MYC and SPT5 (top) and SPT5 and total Pol II (bottom), observed without CDK7 (siCDK7) or with CDK7 (siCTR) in U2OS cells (Yellow dots: intensity centres of proximity pairs, blue: Hoechst stained nuclei, magenta: Phalloidin staining, scale bar: 5 μm). (B), (C) Quantification of respective PLAs shown in (A). Statistical significance measured by two-sided Wilcoxon test. This figure appears in a similar form in (Baluapuri et al., 2019).

4.5.2 TFIIE is responsible for CDK7 sensitivity

The experiments described above point towards the dependency on CDK7 activity for SPT5 binding to Pol II. This finding is in line with current literature (see 5.2), but the exact molecular details of this dependency are not known. While it is described that CDK7 phosphorylates Pol II CTD and SPT5 (Larochelle et al., 2006), we could not

identify any phospho-site on SPT5 which is altered by CDK7 inhibition (data not shown). However, CDK7 is additionally known to phosphorylate HCE (Human Capping Enzyme, also known as RNGTT (RNA Guanylyltransferase and 5'-Phosphatase)) for enabling mRNA-capping, and TFIIE (Larochelle et al., 2012) for converting Pol II into a receptive state for SPT5 binding (Nilson et al., 2015). To test which of the substrates of CDK7 were critical in transfer of SPT5 from MYC to Pol II, it was important to deplete either of these above-mentioned proteins, i.e., HCE and TFIIE and check if transfer of SPT5 from MYC onto Pol II is perturbed. First, HCE was depleted in U2OS cells by treating them with siRNA targeted against *RNGTT* (*HCE*) gene for 48 hours and PLA between Pol II and SPT5 was performed after fixation and compared with cells which were treated with siCTR. Since this experiment did not show any decrease in the number of proximity pairs between Pol II and SPT5 (Fig 4.16), it can thus be concluded that that HCE doesn't plays a critical role in SPT5 handover, at least in this cellular system.

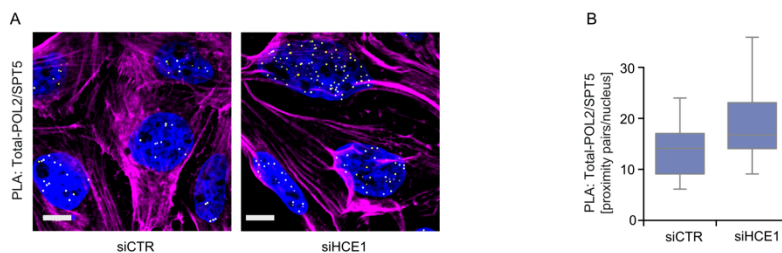


Figure 4.16 Depletion of HCE1 does not decrease SPT5-Pol II proximity

Immunofluorescence images taken from confocal microscope (A) and their quantification (B) of PLAs between Pol II and SPT5, observed in the absence (siHCE1) or presence (siCTR) of the human capping enzyme in U2OS cells (Yellow dots: intensity centres of proximity pairs, blue: Hoechst stained nuclei, magenta: Phalloidin staining, scale bar: 5 μ m). This figure appears in a similar form in (Balupuri et al., 2019).

This pointed towards the hypothesis that TFIIE prevents association between SPT5 and Pol II, which could be rate-limiting for handover of SPT5 from MYC. To confirm this, similar PLA was performed after depletion of one of the TFIIE subunits via siRNA (Fig 4.17A). Additionally, the same PLA was also performed in presence and absence of CDK7 inhibitor THZ1. The expectation in this experiment was that a higher fold change between siTFIIE and siCTR would be noted for SPT5/Pol II proximity pairs in THZ1 treated cells if CDK7 really helps in evicting TFIIE from Pol II complex. Even though the number of proximity pairs decreased upon THZ1 treatment, the fold change upon

depletion of TFIIE was indeed higher (FC=6) when compared to uninhibited condition (FC=1.46) (Fig 4.17 B, C).

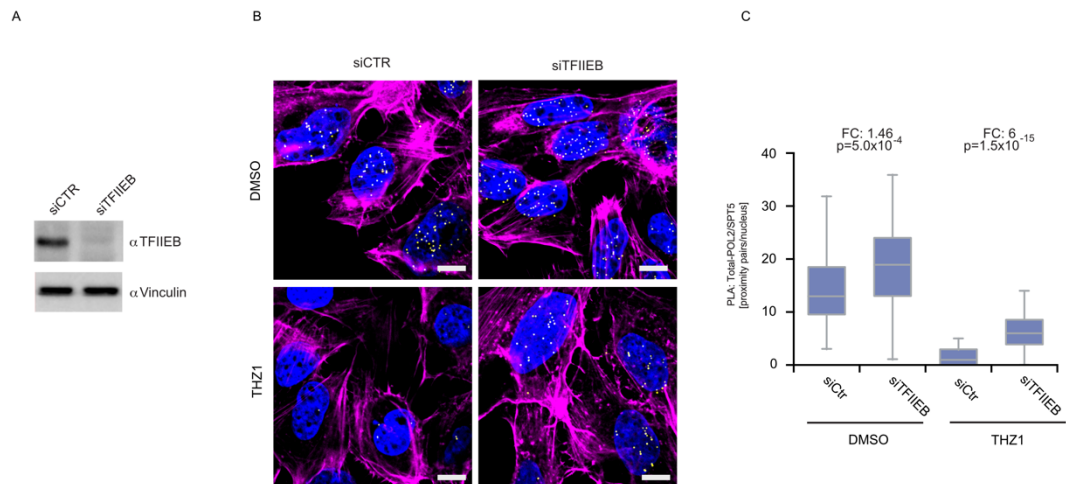


Figure 4.17 CDK7 mediated eviction of TFIIE from Pol II

(A) Immunoblot showing depletion of TFIIE-beta, vinculin is shown as loading control. Immunofluorescence images (B) and quantification (C) of PLAs between SPT5 and total Pol II in cells shown in (A), combined with or without THZ1 treatment. The proximity pairs upon inhibitor and depletion of TFIIE were quantified and displayed as boxplots (yellow dots: intensity centres of proximity pairs, blue: Hoechst stained nuclei, magenta: Phalloidin staining, scale bar: 5 μ m). Statistical significance measured by two-sided Wilcoxon test. This figure appears in a similar form in (Balupuri et al., 2019).

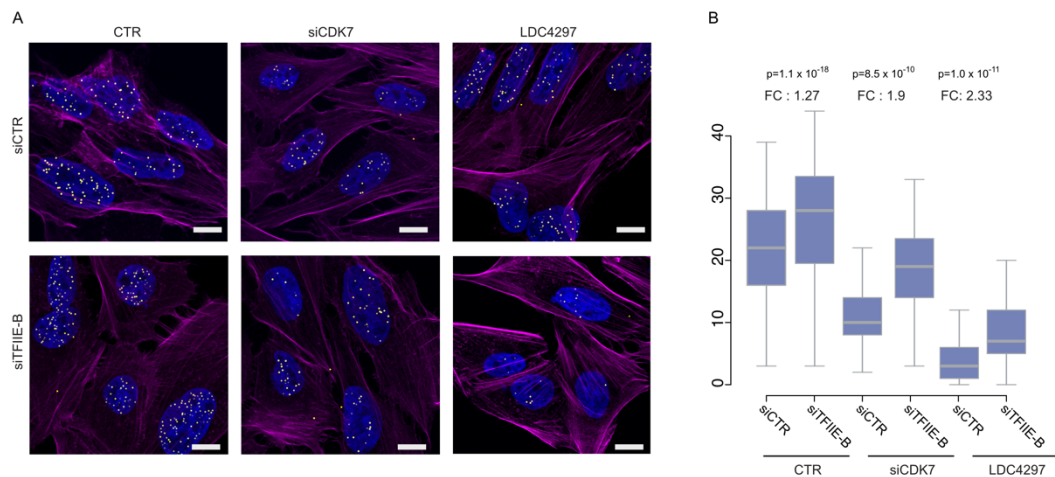


Figure 4.18 TFIIE release converts Pol II into a receptive state for SPT5

Immunofluorescence images taken from confocal microscope (A) and their quantification (B) of PLAs between SPT5 and total Pol II. TFIIE was depleted in cells with CDK7 inhibition (via LDC4297) or depletion (via siCDK7) and compared to controls cells. The number of proximity pairs was quantified and the ratio between TFIIE-depleted cells to control cells was calculated as fold change (FC). Statistical significance measured by two-sided Wilcoxon test. For (A), yellow dots: intensity centres of proximity pairs; blue: Hoechst stained nuclei; magenta: Phalloidin staining; scale bar: 5 μ m. This figure appears in a similar form in (Balupuri et al., 2019).

Similar results were obtained when CDK7 was either depleted using siRNA or inhibited by a more specific inhibitor LDC4297 (Fig 4.18). When the fold changes between the number of proximity pairs (siTFIIE vs siCTR) for SPT5/Pol II PLA was calculated, it was always found to be higher in CDK7 inhibited or depleted conditions (Fig 4.17, Fig 4.18), thus indicating that the release of TFIIE could allow Pol II to bind SPT5.

It can thus be concluded from these experiments that CDK7 phosphorylates TFIIE in order to generate a SPT5 receptive Pol II complex, thus enabling the handover of SPT5 from MYC to Pol II.

4.6 Pol II Processivity and Directionality is regulated by MYC levels

SPT5 is known to affect multiple aspects of Pol II behaviour in yeast. Depletion of SPT5 in yeast using an auxin-inducible degron allele resulted in increase of antisense transcription and decrease of sense transcription. In other words, in absence of SPT5, Pol II complex loses directionality and processivity during transcription (Shetty et al., 2017).

If SPT5 indeed gets transferred to Pol II by MYC, then similar changes in Pol II behaviour can be expected when MYC levels are altered in cells as well. To estimate changes in promoter directionality, metabolic labelling of nascent RNA with 4-thiouridine (4sU) was performed for a short time (15 mins) followed by biotinylation of RNA and pull-down using streptavidin beads. The resulting RNA was then converted into cDNA libraries in a fashion which preserved the information on which strand do the reads originate from, and subjected to deep sequencing. This technique, called 4sU-Seq, was performed in triplicates in U2OS cells corresponding to MYC ON and MYC OFF conditions.

4.6.1 MYC levels regulate directionality of Pol II

Upon inspection of 4sU-Seq reads on individual genes, it was noted that the absence of MYC resulted in decreased reads in the direction of genes itself, suggesting lower transcription in sense orientation (Fig 4.19). However, on the anti-sense strand at upstream promoter, the reads showed behaviour opposite to that of the sense strand. In MYC OFF conditions, the 4sU-Seq signal was found to be increased at the upstream antisense promoters as compared to the reads in MYC ON condition in the same region and gene.

Results

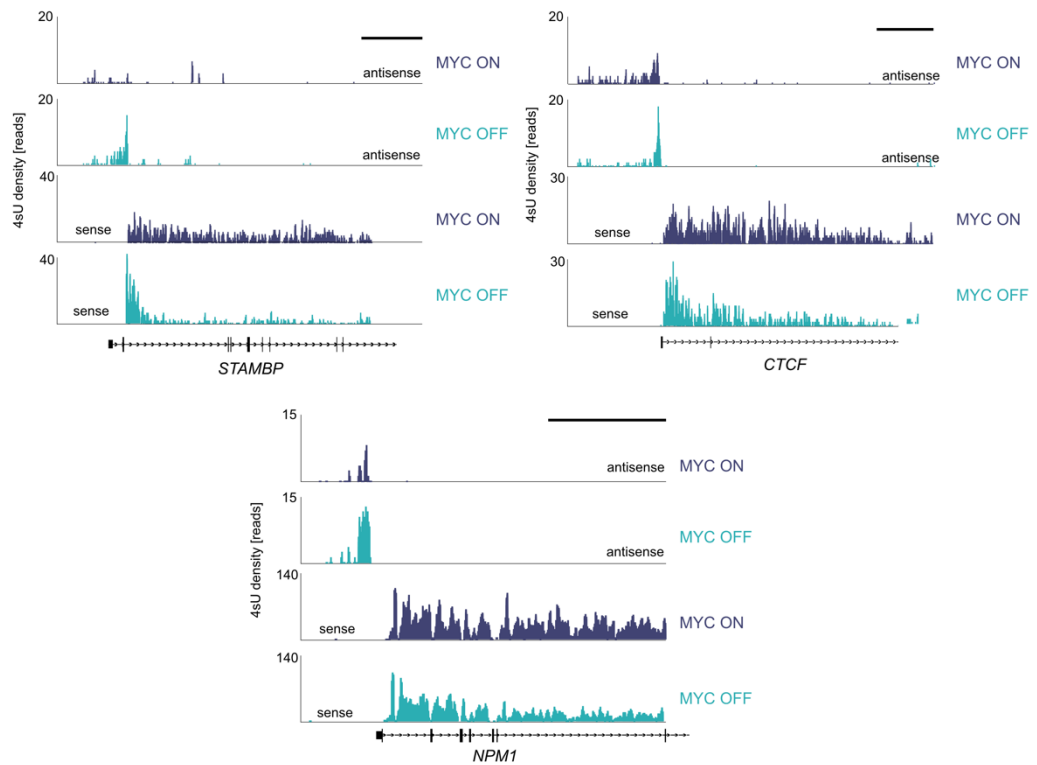


Figure 4.19 MYC levels regulate directionality of Pol II

Genome browser picture of nascent RNA. Example of normalized directional 4sU signal at the *STAMBP*, *CTCF* and *NPM1* gene from U2OS cells in the absence (MYC OFF) and presence of MYC (MYC ON). Scale bars represent 10 kb. This figure appears in a similar form in (Balupuri et al., 2019).

This pointed towards a defect in ratio of Pol II undergoing sense and antisense transcription, which is in line with the function ascribed for SPT5 in yeast (Shetty et al., 2017). In order to test if this was a phenomenon happening on all genes, read densities were averaged across the length of all expressed genes and compared between MYC ON and MYC OFF condition. Such a comparison (Fig 4.20 A) revealed that in MYC OFF condition, nascent RNA signal decreased along the genes while showing an increase in reads in upstream anti-sense direction, while in MYC ON condition, the sense transcription showed high level of reads across the genes with less degree of antisense transcription.

In order to quantify this behaviour, ratios of reads in sense and antisense direction, termed as directionality score (Core et al., 2008), were calculated for each gene. Comparison of directionality scores between MYC ON and MYC OFF showed consistent effects across the replicates (Fig 4.20 B) indicating that MYC levels affect the directionality of Pol II across all genes (Fig 4.20 C).

Results

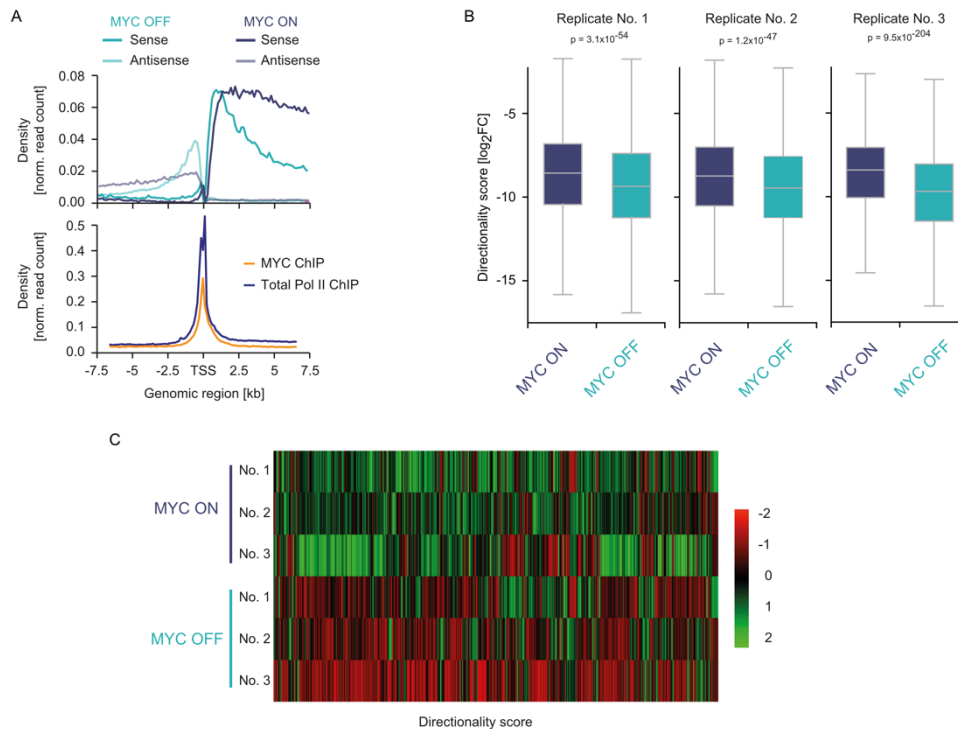


Figure 4.20 Pol II loses directionality globally in absence of MYC

(A) Normalized average read density of 4sU-seq experiments (upper panel) in U2OS cells in the absence and presence of MYC. Curves showing spatial distribution of reads independently aligned to sense and antisense strands around ± 7.5 kb of TSS for genes longer than 8 kb. Comparison to MYC and RNA Pol II binding in the same region originating from ChIP-seq data is shown (lower panel) as average read density (Walz et al., 2014). (B) Box plot of directionality scores. Directionality scores in U2OS cells in the absence and presence of MYC (approximately 10,400 expressed genes, p-value, two-tailed unpaired Wilcoxon test). Directionality scores were calculated by dividing reads from TSS-TES by TSS-1.5 kb gene regions for all genes. (C) Heat map with normalized directionality scores for all 3 replicates calculated in the absence and presence of MYC (approximately 10,400 expressed genes, TSS: Transcriptional start site). This figure appears in a similar form in (Balupuri et al., 2019).

4.6.2 Processivity of Pol II is a function of MYC levels

Upon closer inspection of genes in browser tracks, it was evident that in some cases, more reads were found in TSS proximal ends of the genes in MYC OFF condition as compared to MYC ON condition. However, this behaviour changed after a few kb from the promoter, and into the gene bodies. In fact, less reads were found the very same genes in the distal ends of the genes in MYC OFF condition (Fig 4.21 A).

This loss of processivity is distinct from changes in elongation rate (Mason and Struhl, 2005), and indicates a ratio of all Pol II that successfully completed transcription as compared to the ones that initiated the transcription. Low processivity of Pol II would indicate premature termination of the transcription cycle. To test this, processivity scores

(i.e. ratio of distal to proximal 4sU-Seq reads) were calculated. This analysis revealed that MYC levels regulate Pol II processivity across all replicates (Fig 4.21 B) and over all the expressed genes in the genome (Fig 4.21 C).

While this cellular system allows acute manipulation of MYC levels, utilization of estrogen receptor fused to MYC results in even faster activation (Eilers et al., 1989) of MYC's activity. Also, since MYCN is also known to affect Pol II behaviour (Büchel et al., 2017), it was also possible the MYCN regulates processivity of Pol II in the cellular systems where *c-MYC* is expressed to lower degree.

In order to test if the Pol II processivity is dependent on MYCN, SH-EP neuroblastoma cells that express a MYCN-estrogen receptor chimeric protein were used, in which upon addition of 200 nM OHT for 4 hours, MYCN is activated. 4sU-Seq carried out in this system in biological triplicates revealed a similar phenotype where MYCN activation increased the processivity of Pol II with mild effect on suppression of antisense transcription (Fig 4.21 D).

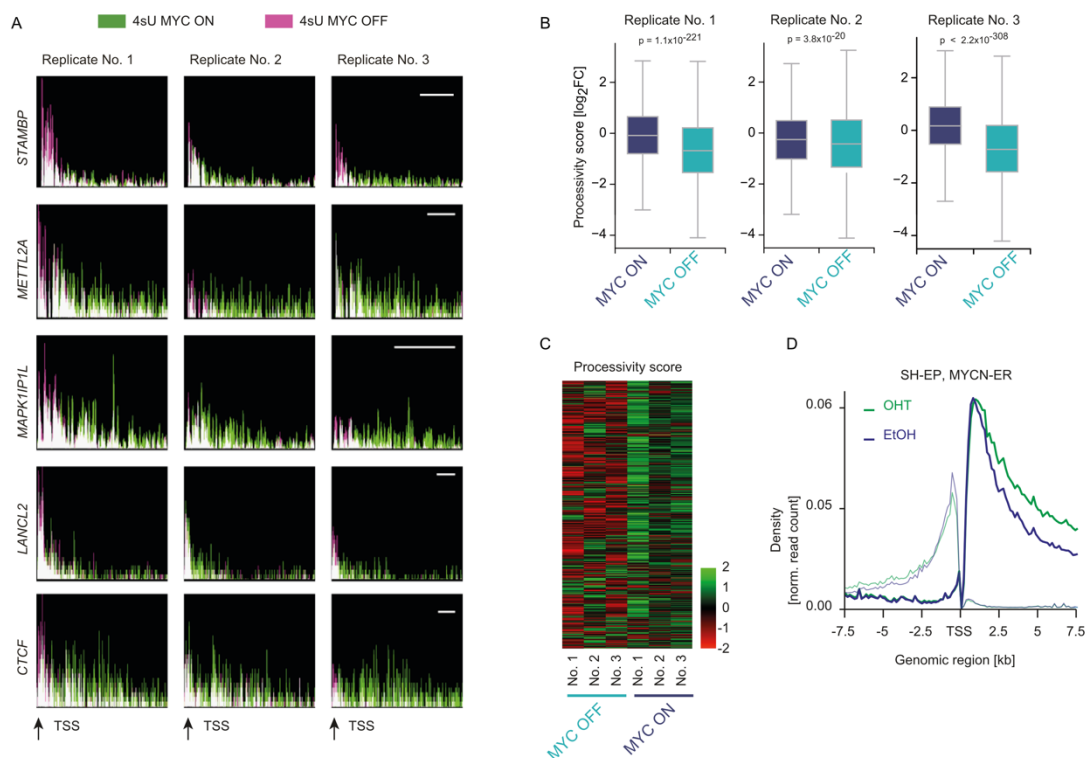


Figure 4.21 Processivity of Pol II is a function of MYC levels

(A) Genome browser pictures of various genes. Normalized 4sU-sequencing reads from U2OS cells visualizing transcriptional activity in the absence (magenta) and presence (green) of MYC in U2OS cells. Scale bar represents 5 kb. (B) Box plot of processivity scores. Processivity scores were calculated by

dividing reads from distal by proximal gene regions for approximately 10,400 expressed genes in U2OS cells. (p-value, two-tailed unpaired Wilcoxon test). (C) Heat map with normalized processivity scores for all 3 replicates calculated in the absence and presence of MYC (approximately 10,400 expressed genes). (D) Combined normalized average read density of 4sU-sequencing experiments (triplicates) in SH-EP-MYC-ER cells with (OHT) and without (ETOH) MYC activation. Curves showing spatial distribution of reads independently aligned to sense and antisense strands around ± 7.5 kb of TSS for genes longer than 8 kb. This figure appears in a similar form in (Baluapuri et al., 2019).

4.6.3 MYC enables productive transcription by RNA Pol II

The above noted dependency of Pol II processivity on MYC raises the question if there is a specific genomic feature in the genes at which this loss of processivity becomes apparent. One such genomic location known to affect Pol II processivity is the presence of promoter proximal alternative poly-adenylation sites (aPAS) (Derti et al., 2012) which when utilized result in premature termination of transcription (Fong et al., 2015). To check if aPAS utilization is dependent on MYC levels, reads from 4sU-Seq in MYC ON and MYC OFF conditions were aligned to the first alternative polyadenylation site and the drop in nascent RNA synthesis seemed to start at precisely this location (Fig 4.22A). This points towards the hypothesis that MYC driven gene regulation could take place through increasing Pol II processivity and suppressing premature termination.

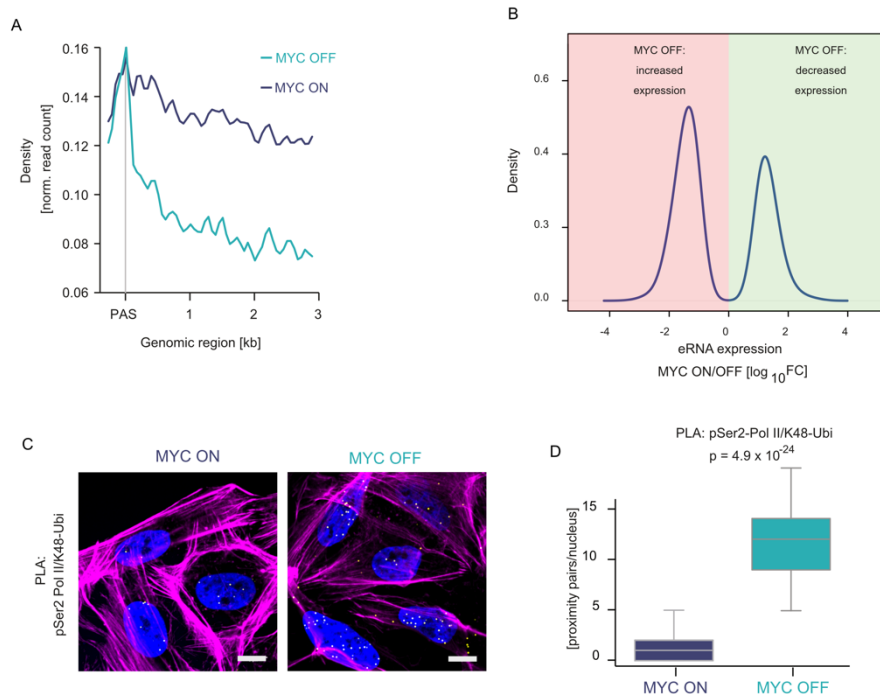


Figure 4.22 MYC enables productive transcription by RNA Pol II

(A) Average read density from 4sU-Seq data of U2OS^{MYC-Tet-On} cells in the absence (blue) and presence (purple) of MYC, representing MYC OFF and MYC ON conditions. Reads were plotted around the first alternative polyadenylation-site (PAS) downstream the promoter. (B) Kernel density plots of ratio between

reads from eRNA locations (The FANTOM Consortium et al., 2014) between MYC ON and MYC OFF conditions. (C) Immunofluorescence images taken on confocal microscope of PLAs between pSer2 Pol II and K48-Ub in U2OS^{MYC-Tet-On} cells at normal (MYC ON) and oncogenic (MYC HIGH) levels of MYC (under MG132 treatment for 4 hours), achieved by overexpressing MYC through Dox treatment. (D) Quantification of PLAs shown in (C). (yellow dots: intensity centres of proximity pairs; blue: Hoechst stained nuclei; magenta: Phalloidin staining; scale bar: 5 μ m). Parts of this figure appears in a similar form in (Balupuri et al., 2019).

Yet another aspect which needs clarification in this regard is the fate of such prematurely terminated Pol II. One possibility is that such Pol II complexes when disengaged from sense strand, get engaged at multiple other locations on the genome and engage in pervasive transcription. This indeed seems to hold true for upstream antisense promoters, which are transcribed at low MYC levels (Fig 4.20), but it remains open if there are other such genomic locations as well.

Another category of such pervasively transcribed but well-defined genomic features is enhancer RNA (eRNA) (The FANTOM Consortium et al., 2014). To test if this category is affected as well, eRNA regulation based on reads from 4sU-Seq was calculated, and the density of eRNA showing increased expression was plotted to compare with the eRNA which show decreased expression in MYC OFF condition. This analysis (Fig 4.22 B) showed that the number of eRNA loci which increase transcription upon MYC depletion is higher than that of those loci which decrease transcription. It can thus be concluded that Pol II which get disengaged from genes due to premature termination get utilized elsewhere to carry out pervasive transcription.

However, it is also possible that the reduction in Pol II processivity results in permanent stalling, and do not disengage from the chromatin. Such permanently stalled Pol II complexes are known to be ubiquitinated at lysine (K-48) moieties which causes it to be degraded by the proteasomal system (Noe Gonzalez et al., 2020).

In order to test if such un-processive Pol II are undergoing ubiquitination, PLA between K48-Ubi molecules and pSer2-Pol II were carried out in presence of MG132 in U2OS cells corresponding to the MYC ON and MYC OFF system mentioned earlier. Indeed, this experiment showed increased signal in MYC OFF condition (Fig 4.22 C, D), pointing out towards the fact that not all Pol II molecule disengaged from chromatin and recycled, and some of the non-processive Pol II get tagged for degradation by proteasomal machinery. However, in absence of verification of Pol II ubiquitination by precise and quantitative methods, these results can be considered to be indicative in nature.

That being said, it can be concluded from these analyses and experiments that MYC levels are instrumental in shifting the Pol II transcription from a pervasive and non-productive manner to a processive and productive one.

4.7 MYC regulates gene expression by altering speed, processivity and directionality of Pol II

4.7.1 MYC mediated defects in elongation rate, processivity and directionality correlate with each other

It has been shown from the work originating from Bentley lab that mutant Pol II which transcribes with slower elongation rate shows less directionality as well in transcription (Fong et al., 2017). Thus, it was likely that MYC mediated changes in processivity and directionality also correlate with each other. In order to test this, the changes in directionality and processivity between MYC ON and MYC OFF were calculated per gene, and the median value of the difference from groups of equal sized bins was plotted against each other.

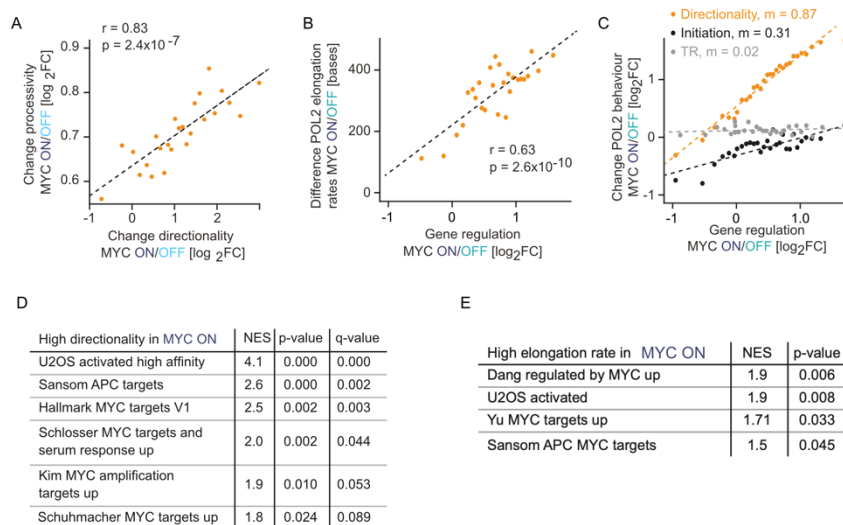


Figure 4.23 MYC mediated effects on Pol II behaviour correlate and affect gene regulation

(A) Scatterplot showing the correlation between MYC-driven changes in Pol II behaviour. Mean values of bins containing an equal number of genes are shown. The x axis indicates the changes in promoter directionality by calculating the differences in the directionality score (\log_2FC) between MYC ON and MYC OFF conditions based on 4sU-Seq data. The y axis indicates the change, calculated similarly for the processivity scores. (B) Scatterplot analysing the correlation between MYC-mediated changes in Pol II elongation rates and gene regulation. Mean values of bins containing an equal number of genes are shown. The x axis displays the change in gene expression by comparing the \log_2FC from 4sU-Seq data in the presence and absence of MYC (MYC ON, MYC OFF). The y axis depicts the difference in the elongation rates in 10 minutes release experiment of 4sUDRB-Seq. (r : correlation coefficient). (C) Correlation between MYC-driven changes in gene regulation and other parameters of Pol II behaviour. Mean values of bins containing an equal number of genes are shown. The x axis shows MYC-mediated changes in gene

regulation (\log_2FC) in the presence and absence of MYC from 4sU-Seq data. The y axis shows MYC-dependent changes in directionality (orange), ratio of 4sU read density in the promoter region (“initiation”, black), and traveling ratio (TR) based on ChIP-Seq of total Pol II (grey). m indicates slope of regression curve. This figure appears in a similar form in (Balupuri et al., 2019). (D) Gene sets most affected by MYC-directed changes in directionality of Pol II. Change in directionality score was used as a ranking parameter for gene set enrichment analysis between 4sU-Seq data from cells with and without MYC expression. A positive normalized enrichment score (NES) indicates gene sets with bi-directional transcription upon depletion of MYC. (E) Gene sets most affected by MYC-mediated effects on elongation when gene set enrichment analysis was performed on genes ranked for change in elongation rate between cells with and without MYC expression. A positive normalized enrichment score (NES) indicates gene sets with reduced elongation rate upon depletion of MYC.

This analysis (Fig 4.23 A) showed that the set of genes on which Pol II are least processive at low MYC levels, on the very same genes, Pol II also loses its uni-directional transcribing property and transcribed anti-sense transcripts as well.

The above-mentioned MYC dependent changes on Pol II behaviour are also likely to have a functional role in terms of regulation of gene expression. Given that MYC binds to all promoters and could hand over elongation factors to Pol II (see 1.2.2), it raised the possibility that by regulation of the processivity and elongation rates of Pol II, MYC could regulate the output of transcription and thus cause the genes to be activated. This hypothesis would be in line with “global amplifier” model of MYC function as proposed by the Levens and Young labs (Lin et al., 2012) (Nie et al., 2012), and would mean that all binding of MYC has a functional relevance.

In order to test this hypothesis, the gene regulation values from newly synthesized RNA (4sU-Seq) were correlated with MYC mediated changes in Pol II behaviour. First, the changes in Pol II elongation rates as measured by 4sUDRB-Seq were correlated with gene regulation, and it was found that the gene groups which showed the largest difference in elongation rates between MYC ON and MYC OFF conditions, were also the genes which were activated by MYC the most (Fig 4.23 B).

Second, since multiple steps in transcription could be regulated at the same time (Michel and Cramer, 2013), and in certain cases MYC indeed seems to be regulating pause-release of Pol II (Rahl et al., 2010), parameters indicative of change in each step of transcription cycle were generated (change in 4sU-Seq reads in promoter region for “initiation”, traveling ratio (Rahl et al., 2010) for “pause-release” and directionality score for “elongation”), and correlated again with gene-regulation. Importantly, MYC driven changes in directionality, and to some extent changes in initiation, were found to be stratified with gene regulation in a stronger fashion ($r=0.87$, directionality vs gene

regulation) as compared to traveling ratio, indicating that transcription elongation is the key rate limiting step in terms of driving MYC mediated gene regulation. (Fig 4.23 C)

Next, to confirm that MYC mediated changes in elongation rate impact known set of MYC target genes, the changes in elongation rates were used as a ranking parameter and this ranked list was then used to perform gene set enrichment analyses (GSEA). This analysis showed that the genes on which elongation rates changed the most, were found to be enriched in well-known sets of MYC- induced genes (Fig 4.23 E). Finally, the same analysis was performed using changes in directionality as a ranking parameter, and known sets of MYC target genes were yet again found to contain the genes on which directionality was most impacted by MYC (Fig 4.23 D).

Thus, from the above analysis and experiments, it can be concluded that MYC driven change in Pol II behaviour, specifically in the step of transcription elongation, is what stratifies global gene regulation by MYC.

4.7.2 Gene length stratifies for MYC's impact on Pol II behaviour

Since it is known that genomic features like gene length correlate with elongation rates (Veloso et al., 2014), it was possible that the above mentioned MYC mediated effects on elongation can also be stratified based on the length of genes. In order to test if such a correlation holds true for MYC mediated gene regulation and its effects of elongation, directionality scores and gene regulation values were compared on single-gene basis, and the behaviour of genes on the extreme ends of this correlation was analysed at browser track level.

Firstly, the genes which showed the maximum change in transcription directionality were also regulated by MYC the most (Fig 4.24 A). While this is in line with the analysis presented earlier (Fig 4.23 C), it was found upon inspection of such genes in Integrated Genome Browser (IGB) tracks, that these genes have length which are many-folds longer than mean value of gene length in the human genome (24 kb). One such example is the *CDK13* (Fig 4.24 B) gene (150 kb long), which showed drastic change in sense and antisense transcription when compared between MYC ON and OFF conditions, and at the same time is activated by MYC ($\log_2FC = 1.8$). On the other hand, genes which were not activated by MYC, showed no difference in antisense transcription between MYC ON and OFF conditions, and were found to be shorter compared to average gene length

(Fig 4.24 C). The correlation of gene regulation by MYC was found to hold true for all MYC regulated genes ($r=0.91$) (Fig 4.24 D), and could be confirmed on single gene levels as well (Fig 4.24 E, $r= 0.69$).

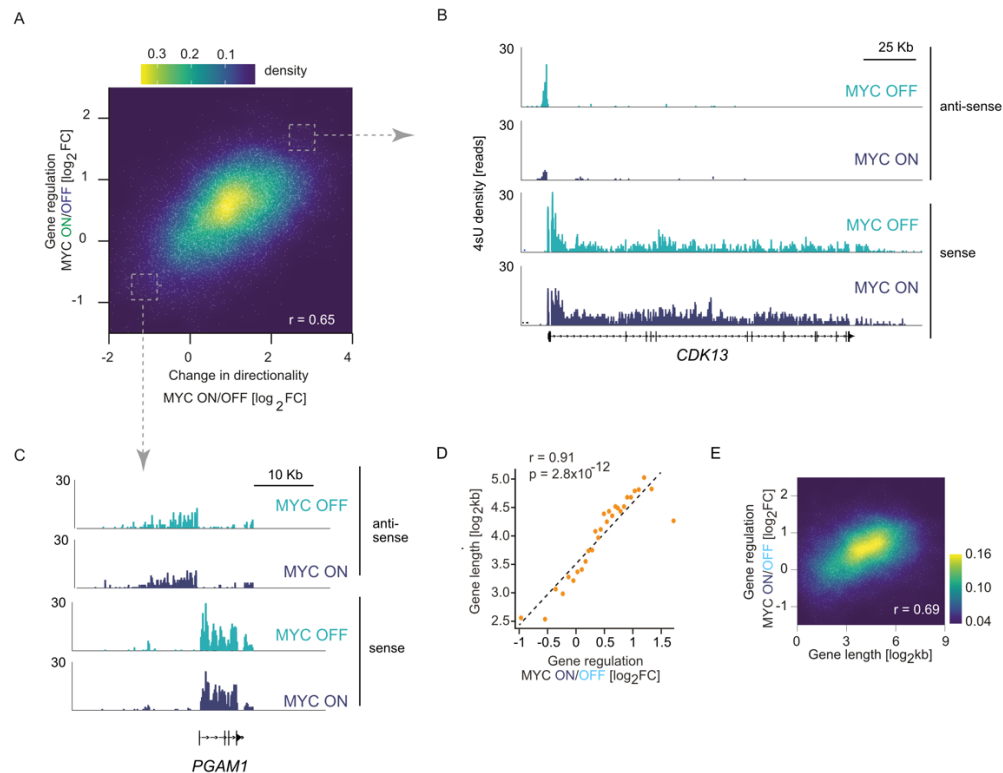


Figure 4.24 Gene length correlates with MYC-mediated gene regulation

(A) 2D-density plot showing the correlation between MYC-mediated gene regulation and change in directionality scores between MYC ON and MYC OFF for significantly expressed genes. Yellow colour indicates higher number of genes in the plot region as compared to purple colour which indicates a smaller number of genes in the plot region. Each white dot indicates a single gene. (B) Indicative browser track showing sense and antisense transcription for *CDK13* gene which is upregulated in MYC ON condition as compared to MYC OFF (C) Indicative browser track showing sense and antisense transcription for *PGAM1* gene which is not upregulated in MYC ON condition as compared to MYC OFF (D) Scatterplot showing the correlation between MYC-mediated gene regulation and length of genes. Mean values of bins containing an equal number of genes are shown. The x axis displays the change in gene expression by comparing the \log_2FC from 4sU-Seq data in the presence and absence of MYC (MYC ON, MYC OFF). The y axis indicates the gene length in terms of \log_2kb . One-way ANOVA was performed for p-value and r value calculation using linear regression model fitting in R version 3.4.4 (E) 2D-density plot showing the correlation between MYC-mediated gene regulation and length of genes for 7,343 significantly expressed genes. Parts of this figure appear in a similar form in (Balupuri et al., 2019).

These analyses show that since the regulation of genes happens at the step of elongation, longer genes are more susceptible to regulation by MYC as compared to shorter genes.

4.7.3 MYC drives the regulation of genes involved in UV response

However, long genes depend additionally on CDK12 for their regulation, and this is known to be most pronounced on DNA damage response genes (Krajewska et al., 2019). These results were primarily based on chemical inhibition of CDK12 by THZ531, and since CDK7 and CDK12 partially share similarity in structure, it is a reasonable assumption that such inhibitors have overlap in their target genes. Given CDK7 was involved in MYC dependant handover of SPT5 to Pol II, and long genes were most affected by low MYC levels, it was then tested if ranked GSEA of directionality and elongation rates show enrichment of genes involved in DNA damage response or not.

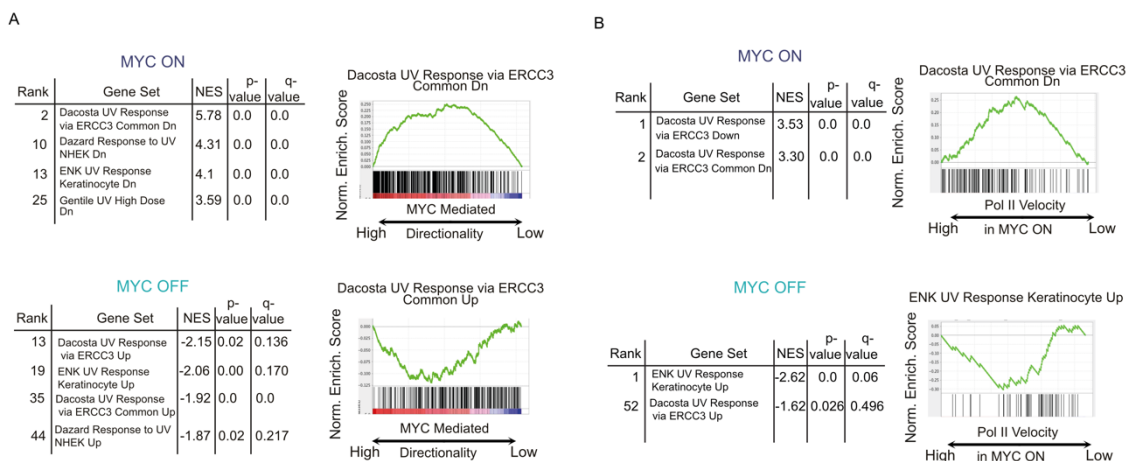


Figure 4.25 MYC drives the regulation of genes involved in UV response

Gene sets affected by MYC-mediated effects on Pol II behaviour. Gene set enrichment analysis was performed on genes ranked for change in directionality score (A) and elongation rate (B) between cells with and without MYC expression. A positive normalized enrichment score (NES) indicates gene sets with bi-directional transcription (A) or reduced elongation rate (B) upon depletion of MYC.

Correspondingly, genes downregulated upon UV treatment in cells were found to be enriched among genes showing least antisense transcription in MYC ON condition (Fig 4.25 A). Conversely, the genes upregulated in UV response were found among the gene sets enriched in MYC OFF condition, indicating that indeed MYC drives the regulation of genes involved in UV response. Gratifyingly, similar results were obtained when the

ranking parameter in GSEA was changed to Pol II velocity measured by 4sUDRB-Seq (Fig 4.25 B).

Taken together, these experiments and analyses point towards the conclusion that MYC is a global regulator of Pol II behaviour and ensures that Pol II remains processive, directional and maintains fast elongation rates to enforce the regulation of its target genes. These target genes are global in nature, though the effects of such regulation are more pronounced on long genes.

4.8 MYC Shapes its Tumour Specific Gene Expression Profile by Squelching Elongation Factors away from Pol II

4.8.1 High MYC levels reduce Pol II elongation rates and processivity

U2OS cells express less than 100,000 molecules of MYC per cell, comparable to many normal proliferating cells. However, in general, cancer cells express high levels of MYC, with multiple myeloma and colorectal cancer cell lines reaching up to a million MYC molecules per cell (Lin et al., 2012) (Lorenzin et al., 2016).

Hence, to study the effect of oncogenic MYC on transcriptional elongation, in contrast to the MYC ON condition, MYC expression was induced by addition of doxycycline without siRNA-mediated silencing of MYC in U2OS^{MYC-Tet-On} cells, resulting in ten-fold increase of MYC levels from 112,744 molecules in EtOH treated cell to 3,414,787 molecules in Dox treated cells (Lorenzin et al., 2016) (Fig. 4.26 A,B). This increase in MYC concentration is spread uniformly among the cells, resulting in MYC driven oncogenic gene expression signature (Fig. 4.26 C) (Walz et al., 2014).

In order to explore the regulation of transcription elongation by oncogenic MYC, elongation rates, processivity scores and directionality scores were required to be calculated in this cellular system additionally. Thus, in this cellular system, termed as MYC HIGH, 4sUDRB-Seq and 4sU-Seq were used to measure Pol II elongation rates and processivity.

Results

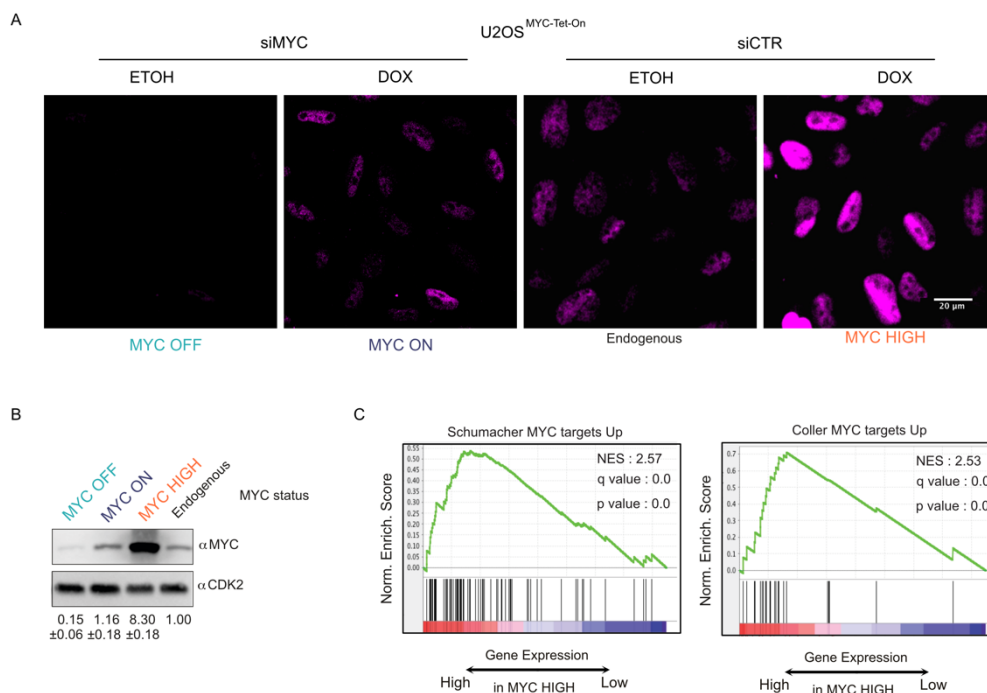


Figure 4.26 High MYC cellular system

(A) Indirect immunostaining images captured by confocal microscope and (B) immunoblot of U2OS^{MYC-Tet-On} cells depleted of MYC (MYC OFF: siMYC), in the “MYC ON” condition (siMYC, doxycycline) with oncogenic levels (MYC-HIGH: doxycycline) and untreated to show endogenous levels. (C) Gene set enrichment analysis indicating that MYC HIGH condition is representative of oncological levels of MYC since the gene regulation shows upregulation of gene sets defining MYC-amplified tumours. MYC is shown in magenta colour in (A). Parts of this figure appear in a similar form in (Baluapuri et al., 2019).

Contrary to expectations, high MYC levels didn’t further increase, but caused decrease in Pol II elongation rates compared to cells with endogenous MYC levels. Inspection of individual genes showed reduction in edge of transcriptional wave front from 54 kb in MYC ON to 41 kb in MYC HIGH on the *EFR3A* gene in 15 minutes DRB release experiment (Fig 4.27 A).

These results were not restricted to a single release time point but were also found in 10 min release experiments with a concomitant reduction in processivity on the same genes (Fig 4.27 B).

Results

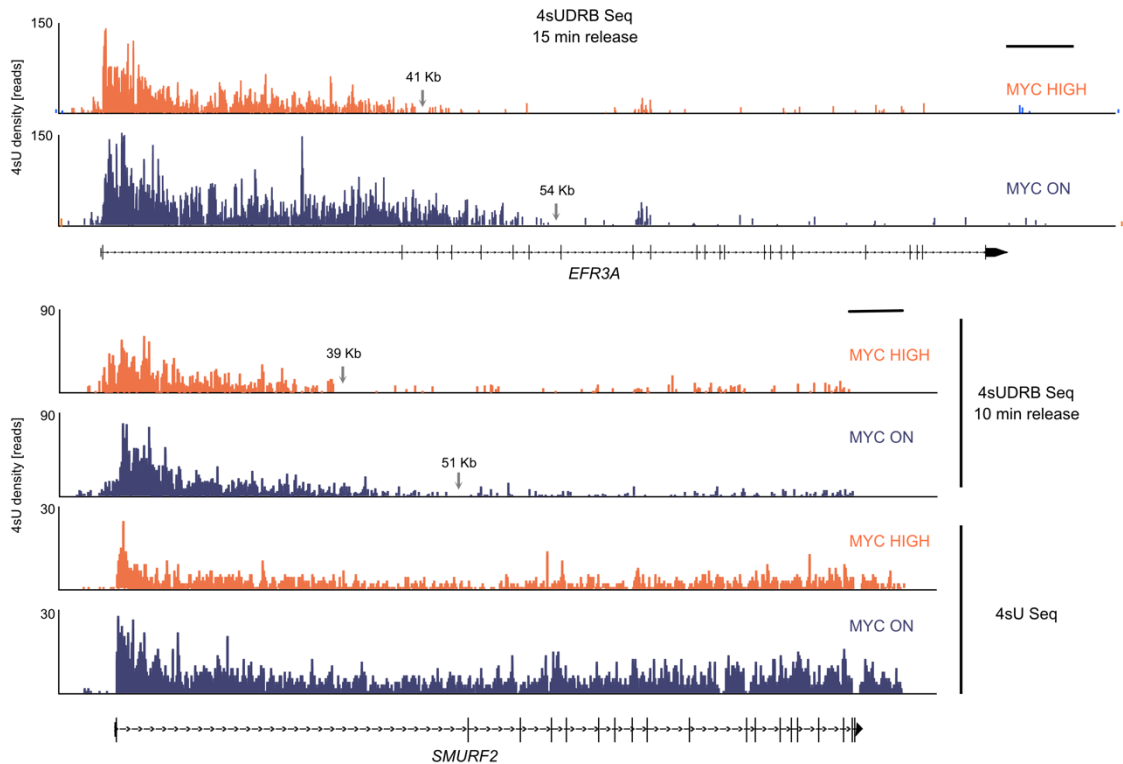


Figure 4.27 High MYC levels reduce Pol II elongation rates and processivity

(A) Genome browser pictures of the *EFR3A* gene in a 4sUDRB-Seq experiment. The wave front indicates the location of Pol II 15 min after release from DRB inhibition at normal (54 kb) and high (41 kb) levels of MYC. (B) Comparison of genome browser pictures of the *SMURF2* gene in a 4sUDRB-Seq and 4sU-Seq experiment. The wave front indicates the location of Pol II 10 min after release from DRB inhibition at normal (51 kb) and high (39 kb) levels of MYC, and shows decreased reads in distal regions as compared to proximal regions of the gene in MYC HIGH for 4sU-Seq. Scale bars represent 10 kb, and arrows with values indicate the approximate distance from TSS.

When read density of Pol II was compared at high MYC levels, the signal of nascent RNA in 10 min release experiment for 4sUDRB-Seq was found to fall to background level approx. 10 kb behind the signal from MYC ON cells (Fig 4.28 A). Further, fast transcribing genes were found to be again more susceptible to loss of elongation rates, as shown by heat maps sorted for Pol II elongation rates (Fig 4.28 B).

Precise quantification of elongation rates based revealed reduction in the median elongation rate in from 3270 bases/min to 2940 bases/min (Fig 4.28 C) in 10 minutes release experiment which correlates with the reduction in rates in 15 minutes release experiment as well (Fig 4.28 D).

Results

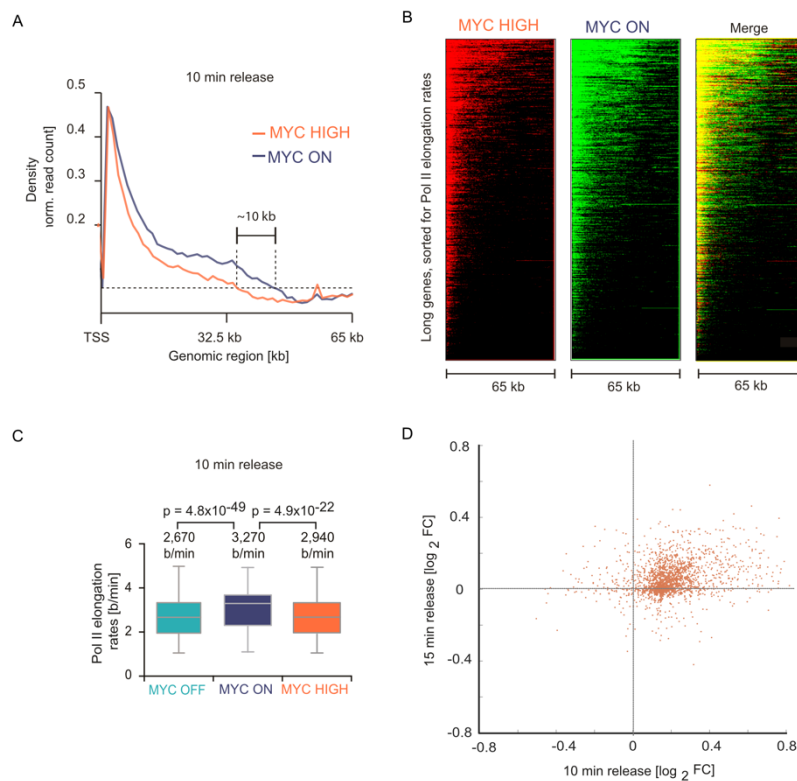


Figure 4.28 Gene specific changes in elongation rates at high MYC levels

(A) Averaged density profile for normalized reads from 4sUDRB-Seq for 3,732 genes (50–100 kb long) 10 min after DRB release, at normal (MYC ON) and oncological (MYC HIGH) MYC levels. The reads were aligned to each TSS and averaged (B) Heatmaps of normalized 4sU reads (10 min after DRB release) in the absence and presence of MYC sorted for elongation rates for 2,018 genes on which elongation rates were calculated (C) Pol II elongation rates as calculated by TERate (Zhang et al., 2016) in the absence (OFF) and presence of MYC (ON) and at oncogenic MYC levels (HIGH). Statistical significance was determined by a two tailed unpaired Wilcoxon test. (D) Correlation in the changes of elongation rates ($\log_2 FC$) between MYC ON and MYC HIGH conditions for 10- and 15-minutes release 4sUDRB-Seq experiments. Parts of this figure appears in a similar form in (Balupuri et al., 2019).

The experiments indicate that at high concentrations, MYC reduced elongation rates and processivity of Pol II, thus causing a repression of specific genes.

4.8.2 SPT5 and SPT6 are squelched away from Pol II

When MYC is overexpressed in the cells to the levels indicated in Fig 4.26, the amount of soluble, non-chromatin bound MYC increases in the nucleoplasm (Lorenzin et al., 2016). It has also been observed that when transcription factors are present in excess in nucleus, they could play a repressive role in transcription (Cahill et al., 1994) and this phenomenon is called squelching (Gill and Ptashne, 1988).

Since it was shown already earlier in this study that MYC interacts with elongation factors SPT5 and SPT6 and hands them over to Pol II, it raised the possibility that MYC could

squelch these elongation factors away as well. This would imply higher degree of proximity between MYC and these elongation factors if they indeed undergo squelching. To test if interaction with elongation factor like SPT5 increased upon MYC overexpression, PLA were carried out between MYC and SPT5 in MYC HIGH conditions. As compared to MYC ON, the PLA signal was found to be higher in MYC HIGH (Fig 4.29).

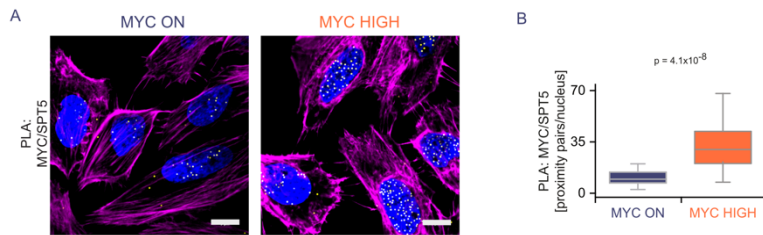


Figure 4.29 MYC and SPT5 share higher proximity in MYC HIGH condition

(A) Immunofluorescence images taken on confocal microscope of PLAs between MYC and SPT5 in U2OS^{MYC-Tet-On} cells at normal (MYC ON) and oncogenic (MYC HIGH) levels of MYC, achieved by overexpressing MYC through Dox treatment. (B) Quantification of PLAs shown in (A). Statistical significance was determined by a two tailed unpaired Wilcoxon test. (yellow dots: intensity centres of proximity pairs; blue: Hoechst stained nuclei; magenta: Phalloidin staining; scale bar: 5 μ m). This figure appears in a similar form in (Balupuri et al., 2019).

On the contrary, when PLA between pSer2-Pol II and SPT5 was carried out, the signal decreased in MYC HIGH condition (Fig 4.30) indicating less SPT5 was associated with Pol II and more was associated with MYC under these conditions.

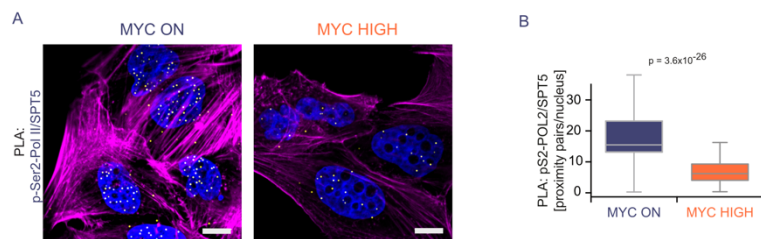


Figure 4.30 SPT5 is squelched away from Pol II at high MYC levels

(A) Immunofluorescence images taken on confocal microscope of PLAs between pSer2 Pol II and SPT5 in U2OS^{MYC-Tet-On} cells at normal and oncogenic levels of MYC, achieved by overexpressing MYC through Dox treatment. (B) Quantification of PLAs shown in (A). Statistical significance was determined by a two tailed unpaired Wilcoxon test. (yellow dots: intensity centres of proximity pairs; blue: Hoechst stained nuclei; magenta: Phalloidin staining; scale bar: 5 μ m). This figure appears in a similar form in (Balupuri et al., 2019).

Similar results were seen for the other elongation factor SPT6 which was found to interact with MYC (Fig 4.11A). Upon overexpression of MYC, PLA signal between pSer2-Pol II and SPT6 was found to be reduced, indicating that Pol II complex stays without critical elongation factors in cells where MYC is overexpressed to oncological levels (Fig 4.31).

These experiments offer a possible mechanism of MYC mediated gene repression via squelching of elongation factors from Pol II to reduce elongation rates and processivity of Pol II.

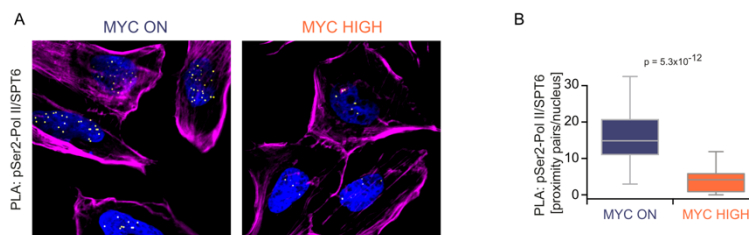


Figure 4.31 SPT6 is squelched away from Pol II at high MYC levels

(A) Immunofluorescence images taken on confocal microscope of PLAs between pSer2 Pol II and SPT6 in U2OS^{MYC-Tet-On} cells at normal and oncogenic levels of MYC, achieved by overexpressing MYC through Dox treatment. (B) Quantification of PLAs shown in (A). Statistical significance was determined by a two tailed unpaired Wilcoxon test. (yellow dots: intensity centres of proximity pairs; blue: Hoechst stained nuclei; magenta: Phalloidin staining; scale bar: 5 μ m). This figure appears in a similar form in (Baluaपुरi et al., 2019).

4.8.3 Squelching by MYC Establishes its Tumour Specific Gene Expression Profile

Since the elongation factors were found to be squelched away from Pol II by MYC, and loss of directionality and processivity was noted as a consequence of increased MYC levels, it was then tested which genes are most susceptible to loss of productive Pol II behaviour.

In order to predict these susceptible set of genes, processivity and directionality ratios were calculated from 4sU-Seq experiments performed in HIGH MYC conditions in duplicate in U2OS cells. These ratios were then used to stratify the resulting gene lists and compared with the ratios calculated between MYC ON and MYC OFF conditions earlier. The most repressed genes between MYC OFF/ON conditions (Fig 4.32 A), showed a drastic reduction in processivity at oncogenic levels (HIGH/ON) when

Results

compared to physiological levels. Similar results were noted when directionality between the two comparison was plotted next to each other (Fig 4.32 B).

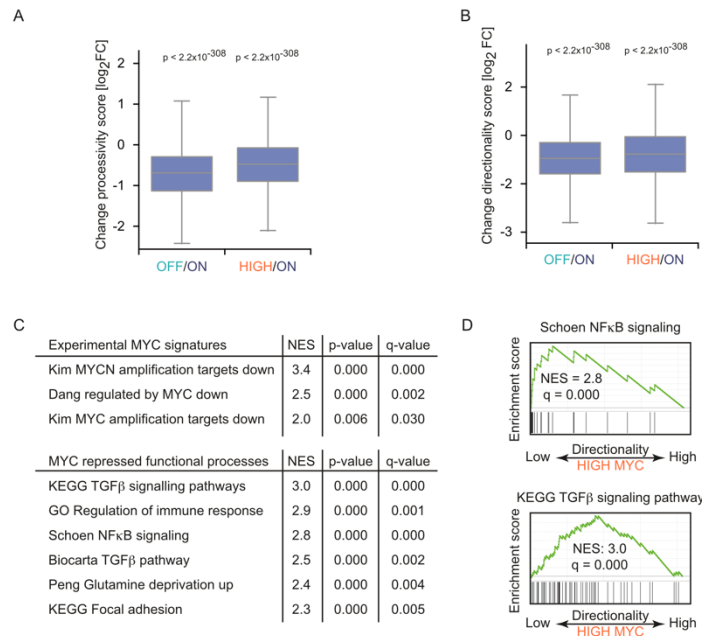


Figure 4.32 MYC's impact on Pol II function drives gene regulation in cancer

(A) Effects on MYC levels on processivity, calculated for U2OS cells with low (OFF), physiological (ON) or oncogenic (HIGH) levels of MYC, as shown by scores calculated based on data from 4sU-Seq. The dependency on MYC was shown by calculating the MYC-driven change in processivity (\log_2FC) per gene. Negative values indicate processive Pol II at physiological levels of MYC (OFF/ON), whereas for oncogenic levels (HIGH/ON), they denote decrease of Pol II processivity. Statistical significance was determined by a two tailed unpaired Wilcoxon test. (B) Effects on MYC levels on directionality, calculated for U2OS cells with low (OFF), physiological (ON) or oncogenic (HIGH) levels of MYC, as shown by scores calculated based on data from 4sU-Seq. The dependency on MYC was shown by calculating the MYC-driven change in directionality (\log_2FC) per gene. Negative values indicate uni-directional Pol II at physiological levels of MYC (OFF/ON), whereas for oncogenic levels (HIGH/ON), they denote bi-directionally transcribing Pol II (C, D) Gene set enrichment analysis where the ranking parameter used was change in directionality score between normal and high (oncogenic) levels of MYC. Statistical significance was determined by a two tailed unpaired Wilcoxon test. (C) Gene sets with the highest normalized enrichment scores (NES). (D) Enrichment plots for the “Schoen NF κ B” gene set and “KEGG TGF β signalling pathway” gene set. Vertical black bars indicate the position of genes in the ranked gene list; the enrichment score is shown as a green line. This figure appears in a similar form in (Balupuri et al., 2019).

Gene set enrichment analysis with genes ranked by change in directionality score between normal and high (oncogenic) levels of MYC showed that the decrease in directionality was strongest for sets of genes that are well established as MYC- repressed genes (Fig 4.32 C), including immune response genes and genes encoding components of the TGF- β pathway (Fig 4.32 D).

Hence, these experiments show that the high levels of MYC that are expressed in human tumours do not further enhance transcriptional elongation but rather bring about squelching of SPT5 and SPT6, reducing the processivity of Pol II transcription. This is seen specifically and strongly on genes that are known targets of MYC-dependent repression and regulators of the interactions of tumour cells with the immune system, resulting in suppression of tumour-suppressive genes.

4.9 MYC shows capacity to form condensate *in vivo*

Since Pol II itself can phase separate into membrane-less compartments which impact the output of transcription (Sabari et al., 2018), it raises the question that do all above mentioned mechanisms take place in these condensates. Also, in addition to SPT5 and SPT6, several other elongation factors are now known to be transferred to Pol II by MYC (Gerlach et al., 2017) (Kalkat et al., 2018) pointing towards its multi-valent property essential for formation of such phase-separated condensates. Additionally, since MYC has been shown to phase-separate with transcription factors like the Mediator complex *in vitro* (Boija et al., 2018), it remains possible that MYC is a part of such condensates *in vivo* as well.

However, to confirm the presence of such condensates with conventional fluorescence microscopy, it requires ability to discern single molecules from clusters of molecules, and the dependency of diffraction limit on the size of molecules being resolved, presents a barrier to this ability (Abbe, 1873).

One of the microscopic techniques which can achieve a resolution higher than what is restricted by the diffraction limit (resolution limit approx. 40 nm, as compared to 250 nm of traditional confocal microscopy) (van de Linde et al., 2011) and allows single molecule imaging is *direct* Stochastic Optical Reconstruction Microscopy (*d*STORM), which can also be utilized to study single molecule localization and clustering of MYC *in vivo*.

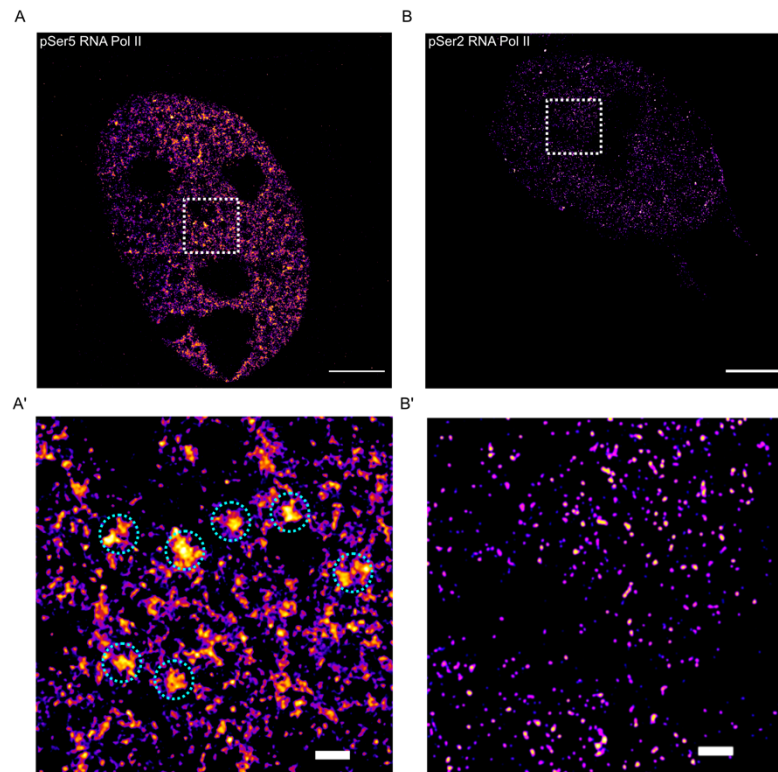


Figure 4.33 pSer5 Pol II condensates can be visualized by *d*STORM

Reconstructed image from rapidSTORM (Wolter et al., 2012) showing single molecule localizations of (A) pSer5 and (B) pSer2 Pol II based on photo-blinking events recorded during *d*STORM imaging. Part of the nucleus (indicated by dotted square) is shown for zoomed in image (A'), (B'). Scale bars denote 5 μm in (A), (B) and 500 nm in (A'), (B'). *d*STORM performed by Patrick Eiring and Marvin Jungblut (Department of Biotechnology, University of Würzburg) on staining performed by Apoorva Baluapuri. Post reconstruction image processing was performed by Apoorva Baluapuri.

Towards this end, it was first required to test if *d*STORM method can be utilized for studying molecular clusters. Hence, initially, those molecules which are known to undergo clustering were imaged. pSer5 and pSer2 forms of Pol II, which can be stained using phospho-specific fluorophore labelled antibodies were tested in unperturbed U2OS cells following paraformaldehyde fixation since it is known that pSer5 Pol II undergoes condensate formation while the elongating form of Pol II (pSer2) gets released from these condensates (Guo et al., 2019) (Boehning et al., 2018). The resulting images from reconstruction of localization events showed pSer5 Pol II undergoing marked clustering behaviour (Fig 4.33 A). Correspondingly, the same staining with pSer2 specific Pol II antibody did not show any signs of cluster formation (Fig 4.33 B).

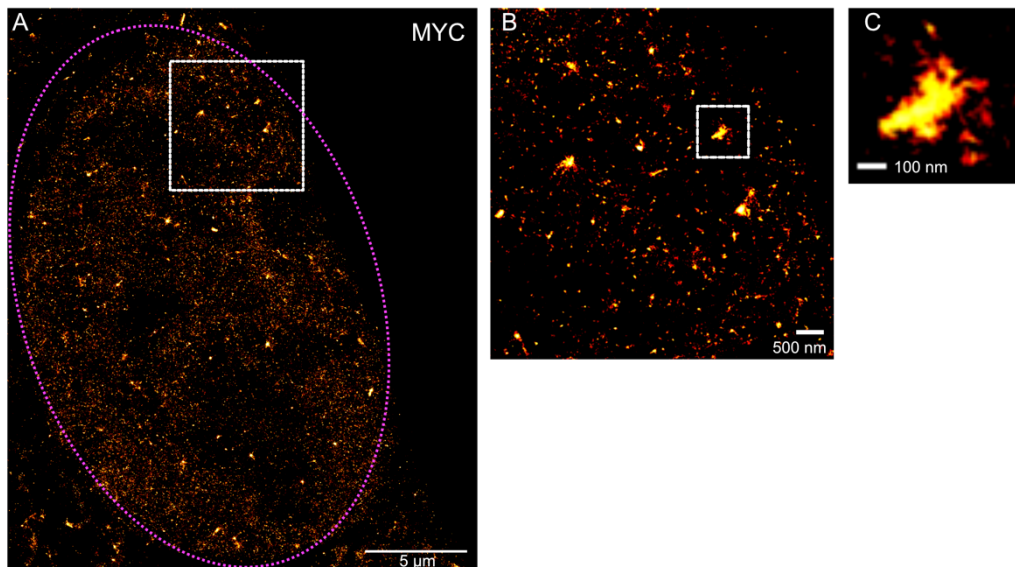


Figure 4.34 Single molecule localization of MYC

(A) Reconstructed image from rapidSTORM (Wolter et al., 2012) showing single molecule localizations of MYC molecules in nucleus (indicated by magenta dotted line) based on photoblinking events recorded during *d*STORM imaging. (B) Part of the nucleus (indicated by dotted square) is shown for zoomed in image next to full scale image, and a selected example of supposed condensate formation with additional zoom is shown in (C). Scale bars denote 5 μm in (A), 500 nm in (B) and 100 nm in (C). *d*STORM performed by Patrick Eiring and Marvin Jungblut (Department of Biotechnology, University of Würzburg) on staining performed by Apoorva Baluapuri. Post reconstruction image processing was performed by Apoorva Baluapuri.

Since it was evident that *d*STORM can detect both clusters and monomers of molecules, the next step was to test if MYC clusters can be identified and differentiated from single molecules in cells or not. Towards this end, single molecule localization for MYC was carried out using monoclonal antibody against MYC which recognized exactly one site per MYC molecule.

The localization events thus recorded were used to quantify the concentration of MYC molecules in cells, which resulted in approx. 250,000 molecules per nuclei. This was in close proximity to the number from earlier studies (Lorenzin et al., 2016) which indicated that MYC is always present in excess in the nucleus to bind to the chromatin.

Finally, *d*STORM was then deployed to test the evidence of MYC cluster formation (Fig 4.34). While there were obvious signs of condensates which were more than 300 nm wide, sophisticated image analysis is required to test if such the occurrence of such clusters is statistically significant. Also, in order to rule out that these are artefacts from imaging, further work like imaging with tagged fluorophores is required.

Since MYC shows signs of phase separation *in vitro* (Boija et al., 2018) and undergoes condensate formation *in vivo* as shown above in unperturbed cells, it opens the possibility that MYC can undergo phase separation in response to chemical perturbation and stress effectors as well.

4.10 Proteasomal inhibition drives phase separation capacity of MYC

It has been known that proteasomal inhibitor MG132 re-organizes MYC distribution to nucleoli (Arabi, 2003) and PML bodies (Smith et al., 2004). However, not only the functional relevance of this phenomenon is unknown, it is also not clear to what extent does this phenomenon occur, and if these are the only organelles to which MYC is evicted upon MG132 treatment. It is also possible that other such compartments are visible in detail only if the structural information within the cell at microscopic level is kept intact. Additionally, co-staining is required to confirm the compartment identity, and multi-colour imaging has been so far challenging with traditional high-resolution imaging. Thus, it was required to carry out high-resolution imaging oriented characterisation of MYC's localisation within the nucleus upon proteasomal inhibition.

4.10.1 MG132 treatment squelches MYC away from chromatin

Thus, to check for such compartments, another super resolution imaging technique, called *Structured Illumination Microscopy* (SIM) was deployed which allows multi-colour imaging and at the same time, allows observation of several morphological aspects of cells which are not possible by conventional microscopy (Schermelleh et al., 2008). U2OS^{MYC-tet-on} cells with MYC at low and high levels were additionally treated with MG132 for 4 hours before fixation with paraformaldehyde, after which the cells were permeabilized and stained using antibody against MYC along with a nuclear and actin marker co-staining. SIM images from these cells were then compared to the images from the cells which were not treated with MG132.

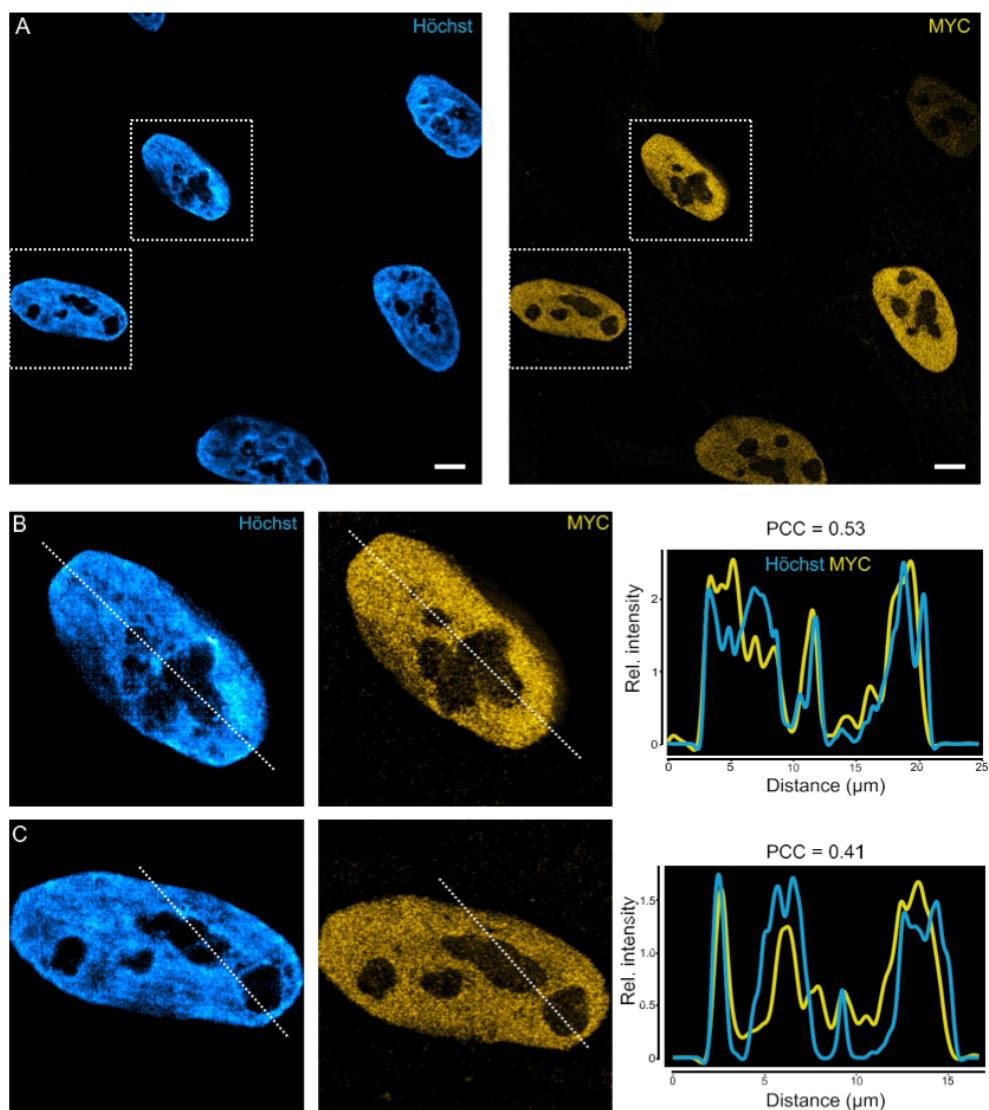


Figure 4.35 SIM reveals MYC is restricted to DNA containing compartment in the nucleus.

(A) Overview of super-resolved image from SIM showing localizations of Höchst and MYC. Two nuclei in the image (indicated by dotted square) are shown for zoomed in image (B), (C). Pearson's correlation coefficient between Höchst and MYC is indicated for the two nuclei along with the profile plot of the intensity over a single pixel wide (dotted line) line across the indicated region. Scale bars denote 5 μm in (A).

SIM imaging showed that under unperturbed conditions, MYC is distributed unevenly in the nucleus (Fig 4.35). To a minor extent, MYC is found in Höchst negative regions of the nucleus, but the majority of the MYC molecules were found to be colocalized with Höchst staining (Pearson's correlation coefficient (PCC) = 0.53 & 0.41). This observation

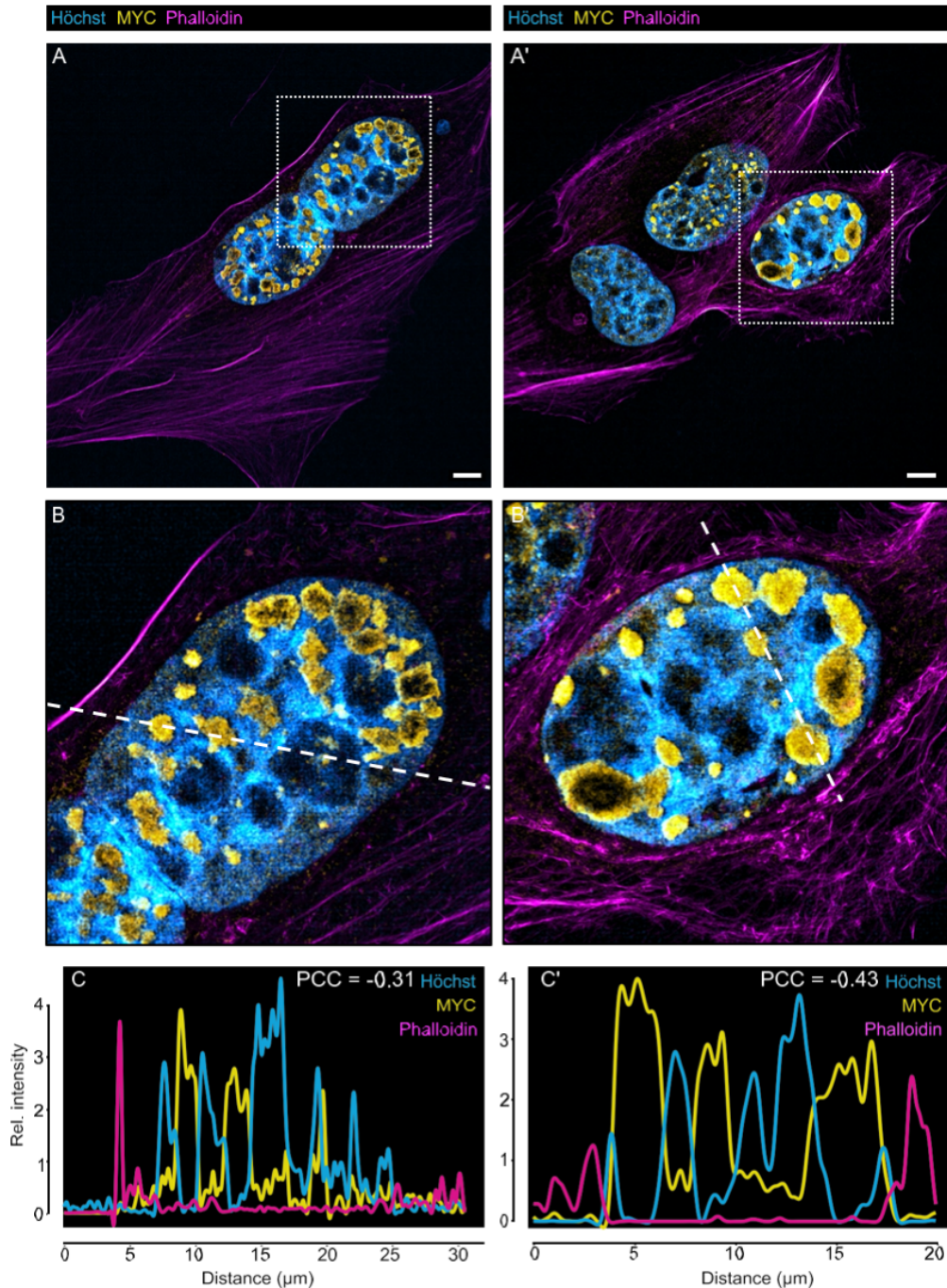


Figure 4.36 MYC is squelched away into DNA free compartment upon MG132 treatment

(A), (A') Two independent overview of super-resolved images from SIM showing localizations of Höchst, Phalloidin and MYC. (B), (B') One nucleus in each image (indicated by dotted square) is shown as zoomed in image. (C), (C') Pearson's correlation coefficient between Höchst and MYC is indicated for the two nuclei along with the profile plot of the intensity over a single pixel wide (dashed line) line across the indicated region. Scale bars denote 5 μm in (A).

is not surprising and has been recorded multiple times, and is in line with DNA binding capacity of MYC. However, when the cells treated with MG132 were imaged under SIM and analysed (Fig 4.36 A, B), MYC localization showed drastic anti-correlation (PCC = -0.31 and -0.43, Fig 4.36 C, C') with Höchst positive regions of the nuclei. This indicates

that MYC is evicted away from chromatin and squelched away into Höchst negative compartments of the cell.

These Höchst negative regions gave the impression of nucleolar compartments, and since the nucleoli are themselves phase separated compartments involved in protein quality control (Lafontaine et al., 2020), it was necessary to test the extent of colocalization of nucleolar compartment with squelched MYC. Since nucleoli are sites of rRNA transcription and processing, staining of total RNA in the nucleus gives a proximal location indicator of nucleolar regions, since rRNA is present in gross excess compared to mRNA and tRNA.

When total RNA was stained with SYTO-RNA dye and probed for colocalization with MYC in unperturbed U2OS^{MYC-tet-on} cells under confocal microscope, it was found to anti-correlate with MYC and DNA staining. However, when the same was checked for MG132 treated cells, MYC co-localized to a high degree with RNA (Fig 4.37).

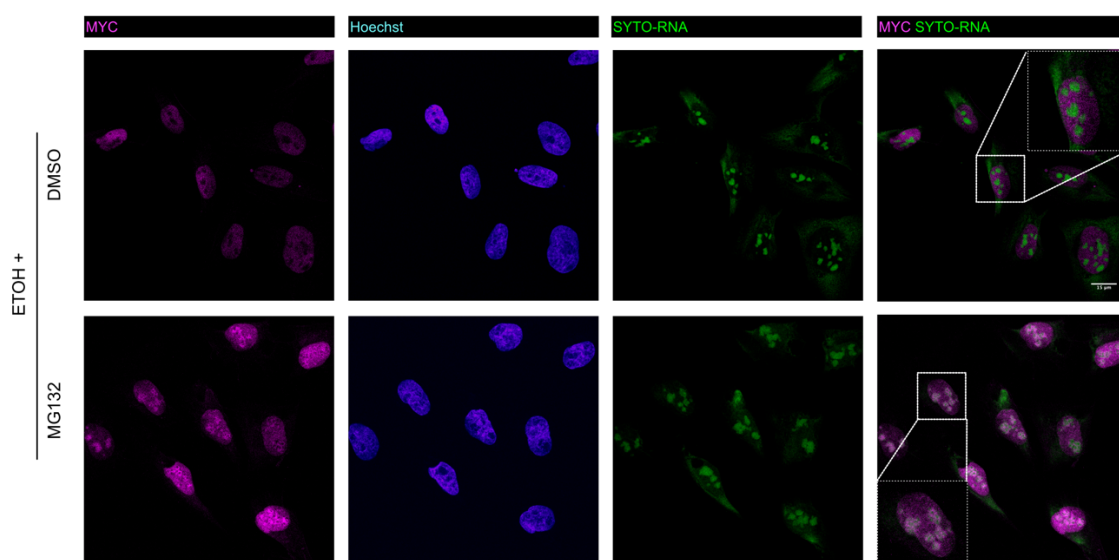


Figure 4.37 MYC partially relocates to nucleoli upon MG132 treatment

Immunofluorescence images taken on confocal microscope showing relative localization of MYC and SYTO-RNA stained RNA (as a proxy marker for nucleoli) within the nucleus in response to treatment of cells with MG132 (bottom) and compared with cells which were not treated with MG132 (top) (MYC is shown in magenta, Hoechst in blue and SYTO-RNA in green).

Further, when U2OS^{MYC-tet-on} cells were treated concomitantly with Dox and MG132, and probed for co-localization with RNA, similar results were noted (Fig 4.38). These experiments resulted in the conclusion that indeed MYC re-localizes to nucleolar compartment upon proteasomal inhibition.

However, in all the above immunostainings, it was evident that not all compartments to which MYC is sequestered away, are definitively nucleolar compartments. Some of the MYC positive compartments in MG132 treated cells remained distinctly negative for RNA staining (Fig 4.38) indicating that there could be other phase separated compartments to which MYC can re-locate to upon proteasomal inhibition.

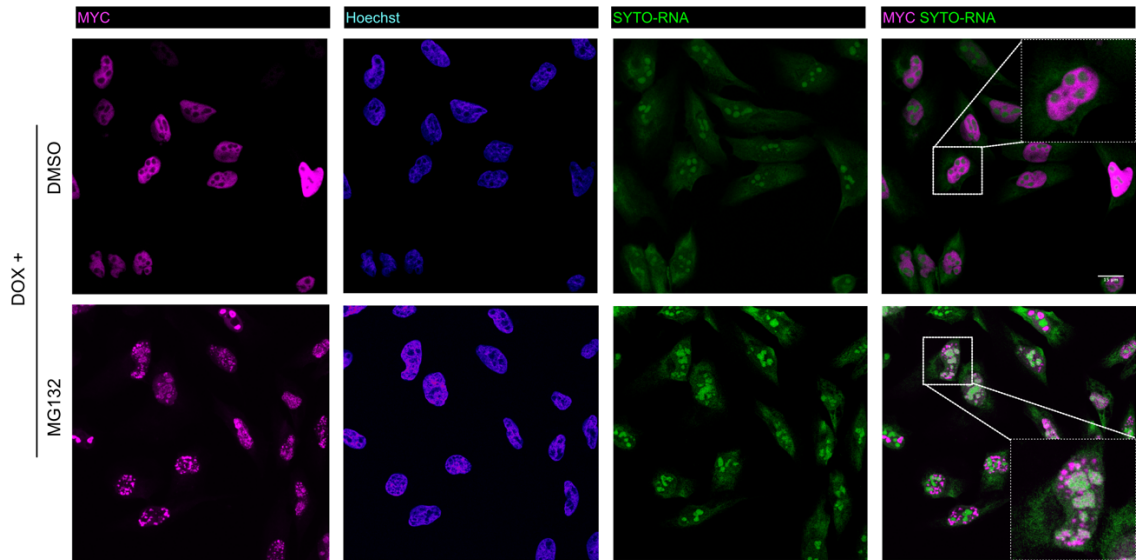


Figure 4.38 MYC behaviour is independent of levels in response to MG132

Immunofluorescence images taken on confocal microscope showing relative localization of MYC and SYTO-RNA stained RNA (as a proxy marker for nucleoli) within the nucleus in response to treatment of cells with MG132 while overexpressing MYC by addition of Dox (bottom) and compared with cells which were not treated with MG132 (top) (MYC is shown in magenta, Hoechst in blue and SYTO-RNA in green).

To test which other compartments can MYC re-localize to, MYC was co-stained with markers for PML bodies as well, which are hubs for sequestration, modification or degradation of partner proteins (Lallemand-Breitenbach and de The, 2010) (Fig 4.39 A). Close inspection of the PML bodies revealed high degree of MYC co-localization with PML protein (Fig 4.39 B).

These results indicate that MYC molecules translocate to phase separated compartments of the nucleus upon proteasomal inhibition, thus pointing towards the argument that MYC itself could be undergoing phase-separation in these compartments.

In order to test this hypothesis, MG132 treated cells were further treated with aliphatic hydrocarbon compound (1,6 Hexane-diol (1,6 HD)) to disrupt the weak interactions that

drive the phase separation phenomenon, and then stained for MYC to check if 1,6 HD treatment reverts the distribution of MYC to H \ddot{o} chst positive regions.

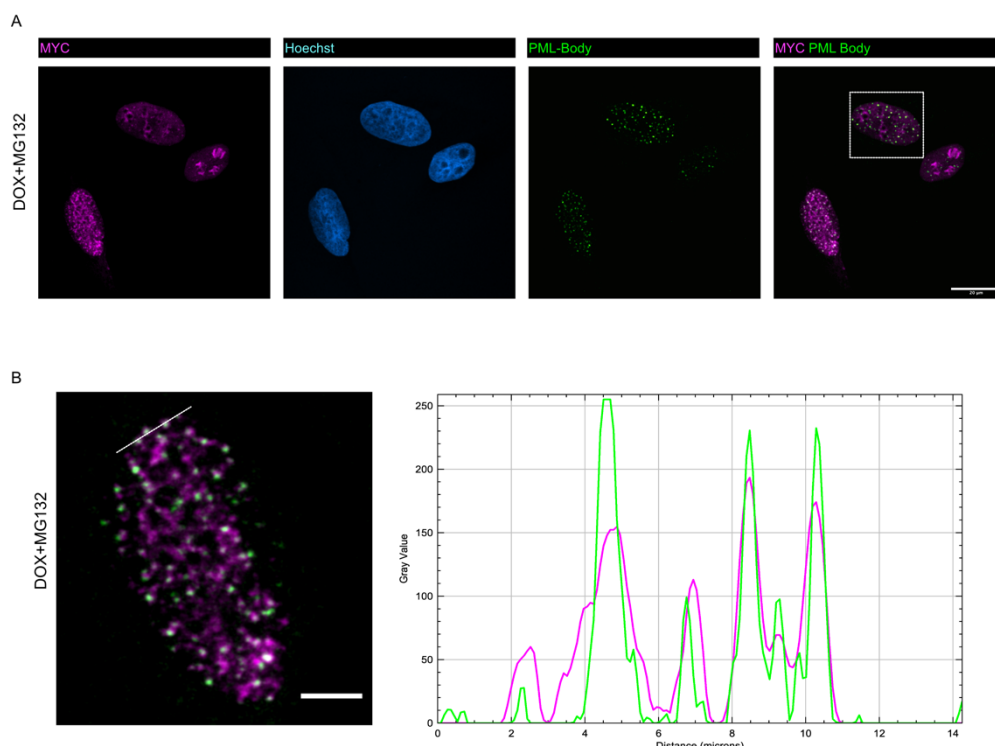


Figure 4.39 MYC partially relocates to PML bodies upon MG132 treatment

(A) Immunofluorescence images taken on confocal microscope showing colocalization of MYC and PML protein in cells overexpressing MYC and treated with MG132. (B) Zoomed in image of cell marked with dotted square in (A), with single pixel wide dotted line indicating the location used to generate the respective line profiles shown (MYC is shown in magenta and PML protein in green; overlap is indicated in white).

Upon observing the MYC staining in these cells, it was found that these condensates appeared to be sensitive to 1,6 HD treatment. MYC staining co-localized with DNA in cells treated with MG132 and 1,6 HD, as compared to cells which were treated only with MG132, where MYC localization anti-correlated with DNA staining (Fig 4.40). These observations result in the conclusion that MYC condensates formed in response to MG132 treatment are indeed liquid-liquid phase separated entities.

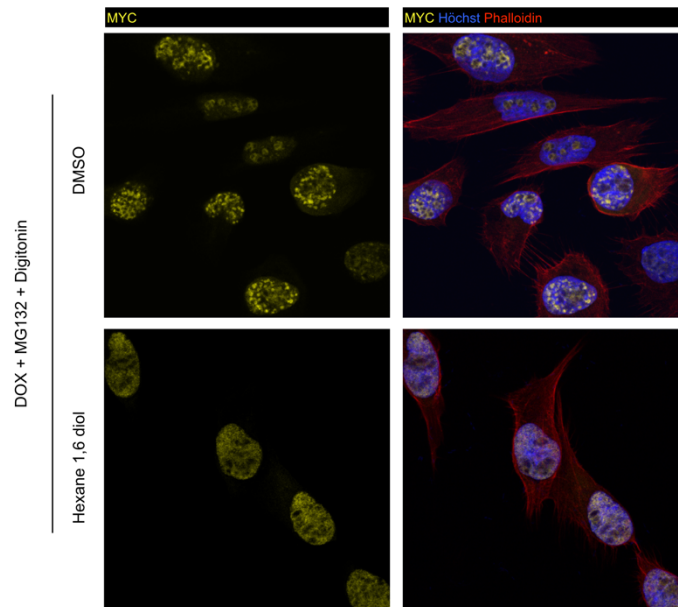


Figure 4.40 MG132 dependant MYC condensates are sensitive to hexane 1,6 diol

Immunofluorescence images taken on confocal microscope showing dissolving of MG132 induced MYC condensates upon treatment of cells with hexane 1,6 diol for 15 minutes (MYC is shown in yellow, Hoechst in blue and Phalloidin in red).

4.10.2 MG132 treatment super-squelches SPT6 and SPT5 with MYC

Since it was noted earlier in this study that at supraphysiological levels in the nucleus, MYC squelches away elongation factors like SPT5 and SPT6 from Pol II (Fig 4.30, 4.31), it raised the possibility that in presence of MG132, MYC can undergo “super-squelching” behaviour and phase-separate along with such elongation factors.

Hence, in order to further dissect the composition of these MYC positive phase-separated compartments, co-staining of MYC was carried out with multiple known MYC interactors under MG132 treated conditions. Surprisingly, only SPT6 (Fig 4.41 B), and to certain extent, SPT5 (Fig 4.41 A) were found to be colocalized with MYC in these condensates formed upon proteasomal inhibition. These results were highly suggestive of the fact that MYC squelches away elongation factors from chromatin bound Pol II to condensates which could affect gene regulation, and offer a new mechanism of gene regulation which is specific to proteasomal inhibition stress.

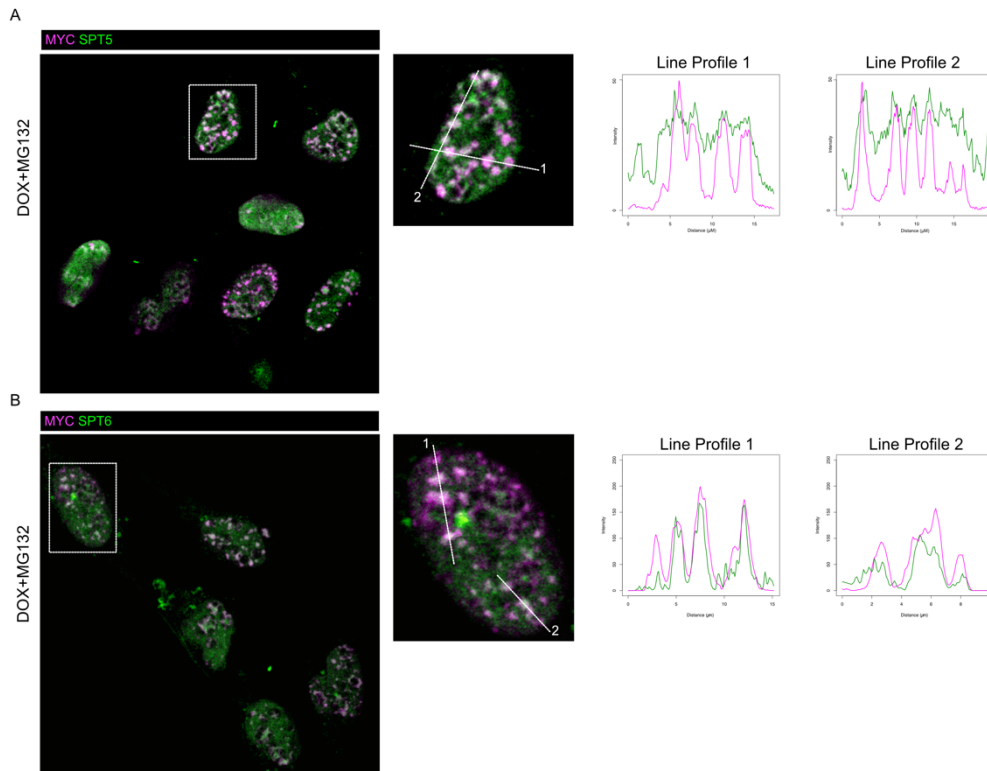


Figure 4.41 SPT5 and SPT6 are squelched into phase-separated compartments

(A) Immunofluorescence images taken on confocal microscope showing colocalization of MYC and SPT5 in cells overexpressing MYC and treated with MG132. Zoomed in image of cell marked with dotted square is shown in inset, with single pixel wide dotted line indicating the location used to generate the respective line profiles shown (MYC is shown in magenta and SPT5 in green; overlap is indicated in white). (B) Immunofluorescence images taken on confocal microscope showing colocalization of MYC and SPT6 in cells overexpressing MYC and treated with MG132. Zoomed in image of cell marked with dotted square is shown in inset, with single pixel wide dotted line indicating the location used to generate the respective line profiles shown (MYC is shown in magenta and SPT6 in green; overlap is indicated in white).

4.11 MG132 treatment changes MYC gene regulation profile

One of the direct effects of squelching elongation factors away from Pol II by MYC would be changes in gene expression. To confirm this effect, fold changes in nascent RNA synthesis upon treatment with MG132 for 5 hours via 4sU-Seq in biological triplicates were measured in U2OS^{MYC-tet-on} cellular system with high and low levels of MYC (Fig 4.42).

Results

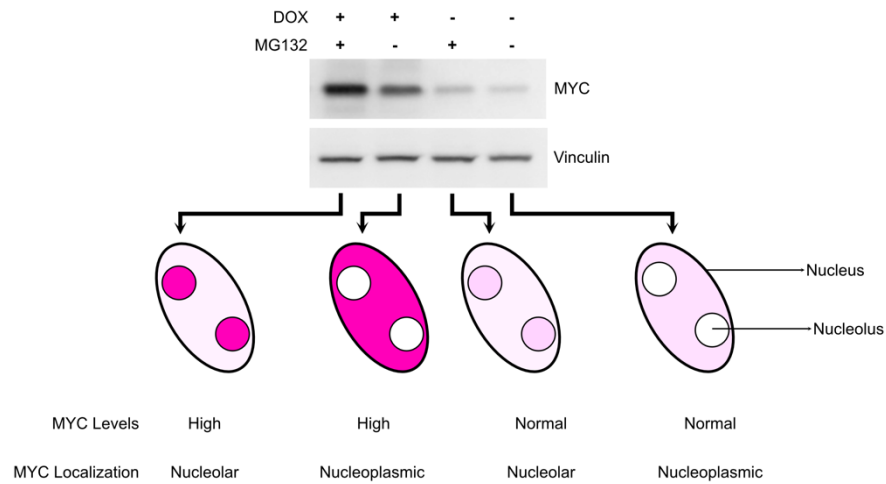


Figure 4.42 MYC levels upon MG132 treatment and MYC overexpression

Immunoblot showing MYC levels and localization inside the nucleus after 8 hours of DOX and/or 5 hours of MG132 treatment on U2OS^{MYC-Tet-On} cells.

In order to compare the effect of MG132 with effect of MYC on gene expression, \log_2FC values based on read counts in the intronic regions of the genes were calculated between the respective conditions, and plotted against the significance ($-\log_{10}(p\text{-value})$). Specifically, certain known MYC regulated genes were highlighted to see the effect of MG132 on MYC regulated genes in such volcano plots (Fig 4.43). This analysis showed that while MG132 affected some of the MYC repressed genes mildly, MYC activated genes (*NPM1*, *NCL*, *RPL8*, *RGS16*) were heavily repressed by MG132 treatment. Additionally, MYC transcription itself seems to be repressed by MG132, however, this does not reflect at MYC protein level (Fig 4.42) due to proteasomal inhibition. Thus, the repression of the aforementioned genes could not be ascribed to change in MYC protein levels. Also, genes which are involved in immune surveillance of tumour cells, which are particularly prone to squelching by MYC (*ITGB1* and *NFKB2*) were also found to be repressed by MG132 treatment.

Results

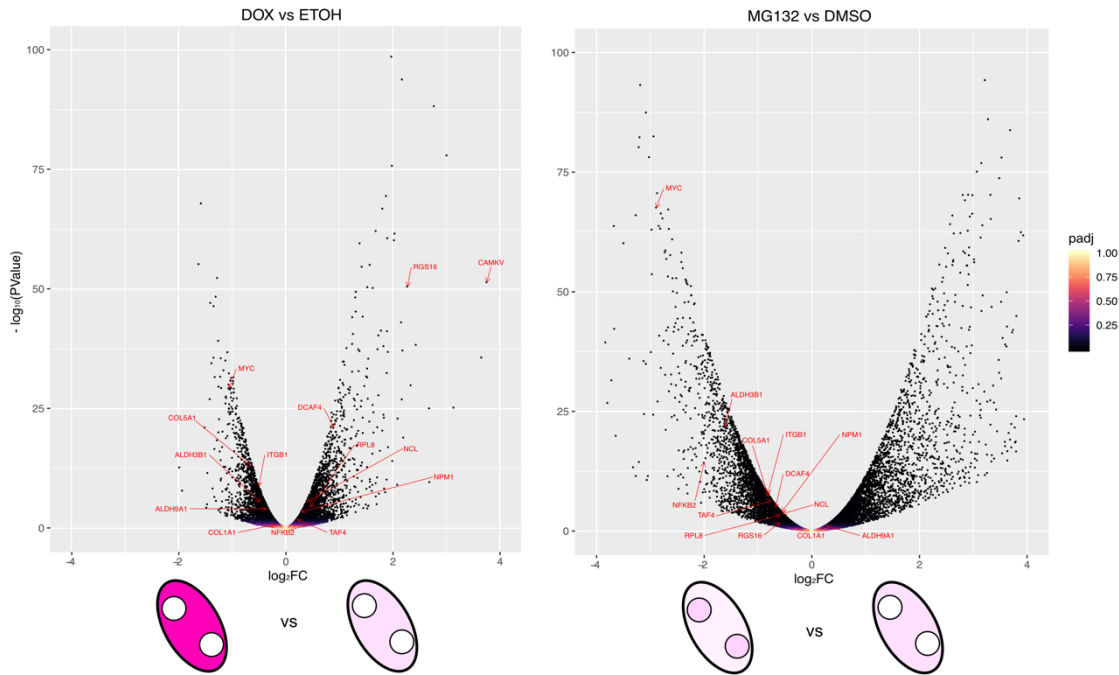


Figure 4.43 MG132 treatment changes MYC gene regulation profile

Comparison of regulation of selected MYC target genes (marked in red) in 4sU-Seq from cells overexpressing MYC (left) or treated with MG132 (right) for 13,182 and 13,578 expressed genes respectively. 4sU-Seq was performed by Pranjali Bhandare (Department of Molecular Biology and Biochemistry, University of Würzburg) under supervision of Apoorva Baluapuri. All subsequent data analysis was performed by Apoorva Baluapuri.

In addition to selected genes, regulation of gene-sets indicating expression signatures that demonstrate patterns of pathway deregulation by MYC (Bild et al., 2006) (Zeller et al., 2003) were also compared between MYC overexpression and MG132 treatment. These gene-sets, when highlighted in volcano plots (Fig 4.44 A) showed that MYC regulated genes, irrespective of them being repressed or activated, were both affected by MG132 treatment. In order to carry out a more global comparison of the gene regulation between MYC overexpression and MG132 treatment, \log_2FC of all significantly regulated genes was plotted as a 2D density plot where the colour coded information on the density of genes in a location of plot was overlaid on the actual genes as well. This comparison showed (Fig 4.44 B) that MG132 massively and globally affects the genes normally regulated by MYC. Not only the genes normally repressed by MYC in a mild manner are strongly repressed by MG132 treatment, the genes which are activated by MYC get repressed as well.

These experiments indicate that MG132 treatment could result in a major change in Pol II complex, particularly by squelching of elongation factors by MYC, which results in repression of genes at a global level, including MYC target genes.

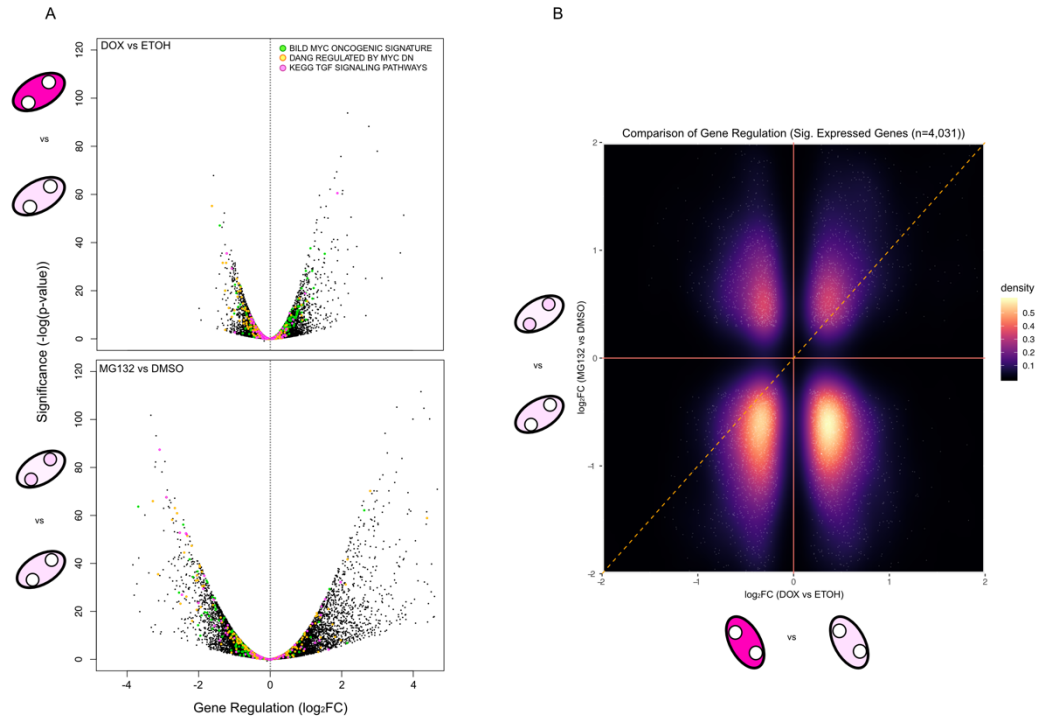


Figure 4.44 MYC activated and repressed genes are affected by MG132

(A) Comparison of regulation of MYC target gene-sets in 4sU-Seq from cells treated with Dox (top) or MG132 (bottom) for 13,182 and 13,578 expressed genes respectively. (B) 2D Density plot showing correlation of gene regulation (\log_2FC) by MYC (DOX vs ETOH) and MG132 (MG132 vs DMSO) for 4,031 significantly expressed genes in both comparisons. Yellow colour indicates higher number of genes in the plot region as compared to black colour which indicates a smaller number of genes in the plot region. Each white dot indicates a single gene

5 DISCUSSION

5.1 MYC as a regulator of Transcription Elongation

Pol II recruited to promoters undergoes a series of transitions during which it progressively associates with transcription elongation and RNA processing factors to form a complete Pol II elongation complex (Chen et al., 2018). During this assembly process, Pol II acquires the ability to transcribe in a processive fashion to generate full-length coding transcripts. This raises the important question which factors control promoter directionality and processive transcription by Pol II. The line of evidence presented in this thesis indicates that MYC facilitates assembly between Pol II and SPT5 early after initiation, since the transfer of SPT5 from MYC to Pol II depends on the activity of CDK7. A global redistribution of Pol II on chromatin from the promoter to the gene body upon MYC activation has been noted earlier, but the molecular mechanisms underlying this change had remained unclear.

The findings presented here argue that MYC facilitates the assembly of a paused Pol II complex and promotes transcription elongation by promoting the accumulation of a substrate - SPT5/Pol II complexes - for subsequent pause-release by CDK9. MYC-mediated transfer of SPT5 induces a change in Pol II behaviour from a slow and poorly processive to a fast and highly processive mode, thus promoting unidirectional transcription. However, this function of MYC posits the key question, why elongation would be a regulated step in the transcription cycle.

Firstly, among other things, the easiest explanation could be the same line of reason, as to why Pol II undergoes pausing: to keep promoters open (Adelman and Lis, 2012) and Pol II ready to transcribe within short time interval from acceptance of growth signals. This function could be especially critical in the light of reasoning that the packaging of chromatin at promoters can function as a barrier to transcription by preventing access to DNA recognition motif. Abortive transcription, while not contributing to gene expression, would allow access to such parts of DNA. Since this has been long known for heat shock response genes (Costlow and Lis, 1984), it is certainly possible for acute and direct regulation of MYC target genes as well. This way, abortive transcription could

contribute to an additional regulatory step in the transcription cycle, in combination to Pol II recruitment which would allow MYC to work together with factors that stimulate recruitment and other steps to establish a finely tuned regulation of transcription.

Secondly, a more central answer would be in the line of reasoning that, it is one of the steps that actually can be regulated in a way that impacts the expression of genes. Other processes related to transcription like nucleotide synthesis are much harder to regulate by molecular pathways. Additionally, other biological processes, like splicing (Saldi et al., 2016) (Fong et al., 2014) and translation (Slobodin et al., 2017) are directly linked to elongation rates of Pol II. Thus, regulating elongation could serve as a central hub for regulating the output from other processes as well.

This would be in line with the recently demonstrated role of MYC in splicing (Cossa et al., 2020) where deregulated MYC overrides a quality check on transcript elongation, whereby splicing defects that would otherwise trigger premature termination are ignored, leading to locking up of Pol II in unproductive elongation complexes (Fisher, 2020). Further, Protein Phosphatase 1 (PP1) is a part of complex that associates with the splicing machinery and provides a feedback control of transcription that can be overridden by deregulated MYC. Inhibitory phosphorylation of PP1 is carried out by CDK9 (Parua et al., 2018), whereas dephosphorylation of SPT5 itself is carried out by PP1 (Parua et al., 2020), offering multiple points of elongation control which is coupled to transcription termination and splicing.

Thirdly, recent evidence suggests that the previously assumed non functionality of anti-sense and other pervasively transcribed species might not be completely true. For example, based on mass spectrometry and ribosome profiling, it has now been shown that protein translation outside of annotated exonic sequences in mRNAs is of pervasive nature as well (Chen et al., 2020). Further, since some HLA-I peptides for certain HLA alleles are now known to be translated from cryptic open reading frames in supposedly non-coding region, including anti-sense promoters (Erhard et al., 2020), it could be hypothesized that MYC could play a role in suppressing similar cryptic peptides, which could explain the biological relevance of antisense transcription to exist in the first place.

5.2 Role of other MYC interactors in gene regulation

The data shown in this thesis argue that MYC driven regulation is primarily carried out during elongation step of transcription, via its interaction with SPT5 and SPT6. However, it is possible the transcriptional cycle is regulated at multiple steps by MYC, and other body of works (Kalkat et al., 2018) (Rahl et al., 2010) would indicate these steps to be driven by MYC's interaction with CDK9.

This raises an important question. Why do multiple mechanisms of regulation even exist, and if so, does MYC impact the other steps as well? In part, the answer to such a question is a thematic one – to be an efficient TF, it needs to modulate multiple steps of transcription. Thus, it is plausible that CDK9 dependent pause release might be a regulated step of MYC driven gene regulation (Rahl et al., 2010), but Pol II molecules which are released from the pause site still need to undergo processive elongation to bring about a productive gene regulation. Moreover, it could very well be that travelling ratio-based claims of MYC regulating pause release were simply misinterpretation of changes on processivity (Ehrensberger et al., 2013). Nevertheless, MYC needs to catalyse the key transitions of all steps of transcription in order to achieve high levels of transcription (Kouzine et al., 2013).

5.3 Molecular mechanisms of MYC function at core promoters

In context of transcription regulation, CDKs can regulate the behaviour of Pol II by phosphorylating not only the CTD of Pol II, but also multiple other regulatory TFs involved.

During the initiation phase of transcription, CDK7 exerts control on transcription by allowing a stable Pol II pre-initiation complex to be formed. This action is performed in multiple steps. CDK7 phosphorylates the Ser5 of Pol II CTD repeats, and additionally, phosphorylate HCE (Human Capping Enzyme) to allow mRNA-capping. Further, TFIIE also serves as a substrate for CDK7 (Larochelle et al., 2012), which results in its dissociation from Pol II. The very same site, which was hitherto occupied by TFIIE, can now be bound by SPT5 (Nilson et al., 2015), thus resulting in a Pol II complex which can now transcribe (Fig 5.1, steps 1-3). Further SPT5 itself is a substrate of CDK7 and helps in capping of nascent RNA as it is transcribed from Pol II.

It has been shown that inhibition of CDK7 increases retention of TFIIE with Pol II thus preventing SPT5 recruitment and attenuates pausing in human cells (Larochelle et al., 2012). Thus, a prerequisite for establishment of Pol II pausing is the exchange of TFIIE for SPT5 which compete for the same docking sites on the surface of Pol II (Grohmann et al., 2011). The resulting model would indicate that CDK7 activity is required for MYC driven handover of SPT5 to Pol II, since TFIIE eviction is dependent on CDK7 activity.

Building on this model, it is shown, here in this thesis, that depletion of TFIIE facilitates CDK7-mediated SPT5 recruitment, thus bolstering role of CDK7 in MYC's handover of SPT5 to Pol II (Fig 5.1, steps 3-5).

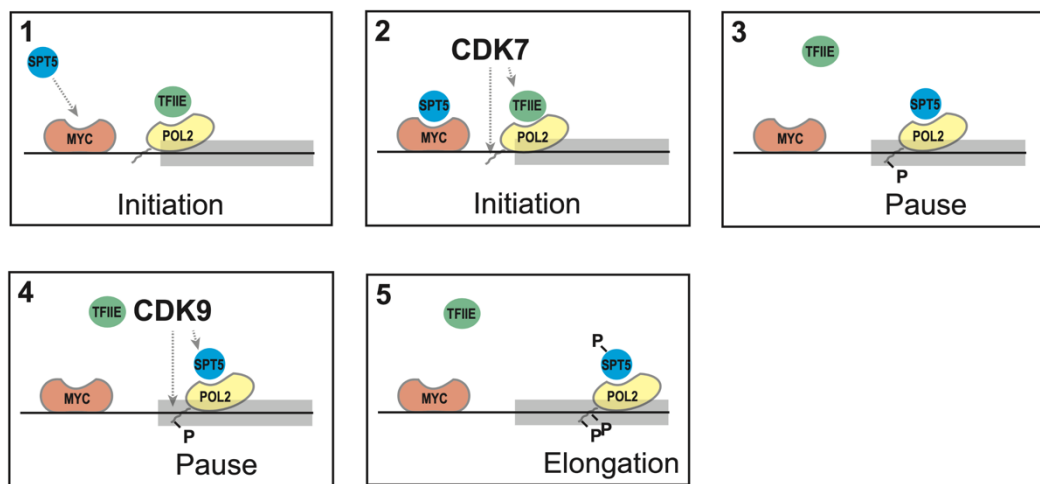


Figure 5.1 Model of MYC's effect on Pol II transcription

MYC binds to chromatin and recruits SPT5 to promoters. CDK7 converts Pol II into a receptive state for SPT5 by promoting phosphorylation and eviction of TFIIE, in addition to initiating transcription by CTD phosphorylation at Ser5 residues. Eventually, CDK9 drives pause release of Pol II by phosphorylating SPT5 and the Ser2 residues of the Pol II CTD.

Although TFIIE and SPT5 both contain sites phosphorylated by CDK7 *in vitro* (Larochelle et al., 2012), which factors and phospho-sites CDK7 utilizes to promote this factor exchange is not yet known. Nevertheless, data presented in this thesis now gives rise to a model where eviction of TFIIE by CDK7 is critical for CDK9's substrate SPT5 to bind to Pol II, without which productive elongation resulting from SPT5 phosphorylation cannot take place.

Since CDK7 hasn't been shown to interact with MYC, it could very well be that MYC mediates the handover of the factors dependent on CDK7 activity in order to bring about gene regulation via elongation. The data from chemical inhibitor screen presented in this

thesis show that CDK7 is critical for handover of SPT5 to Pol II by MYC. Given SPT5 is a direct substrate of CDK9, these data present now a cohesive model how MYC could regulate the assembly of productive Pol II elongation complex.

Not only this model enhances our understanding of MYC's impact on transcription, it also highlights the relevance of CDK7 and CDK9 as attractive targets for targeting transition in the transcription cycle for cancer therapy (Parua and Fisher, 2020) for MYC driven tumours.

5.4 Squelching as a mechanism of gene repression

Regulation of transcription elongation provides an attractive mechanism in terms of gene activation. However, in addition to activating genes, MYC can also repress certain genes, and this repression is important in driving its oncogenic role (Gartel et al., 2001) (Seoane et al., 2001) (Herkert and Eilers, 2010) (Walz et al., 2014) (Sabò et al., 2014). This MYC driven repression can be explained by the MYC/MIZ1 ratio at promoters, i.e., an equal binding of MYC and MIZ1 results in direct repression of genes, whereas higher MYC binding relative to MIZ1 results in gene activation (Peukert et al., 1997) (Walz et al., 2014).

In addition to the above-mentioned direct mechanisms of MYC mediated gene repression, an indirect mechanism of repression by MYC has also been proposed. In this mechanism, reduction of MYC binding relative to the total chromatin bound MYC upon increase in its cellular concentration drives repression of the MYC target genes (Tesi et al., 2019). This “relative” amount of MYC at each promoter was defined as “MYC share” (de Pretis et al., 2017). Strikingly, the genes with lowest shares are the most strongly repressed genes. However, no direct mechanism or evidence for such a model has been presented till now.

The data analysed and compared in this thesis between MYC ON and MYC HIGH show gene repression of specific genes. This gene repression could stem from increase in local concentration of soluble MYC which is actually not bound to chromatin, and exhibits squelching of SPT5 away from Pol II complex. This could be considered as a confirmatory piece of evidence in favour of decreased “MYC share” driving gene repression. The genes on which squelching is most prominent, are the well-known MYC-repressed genes like the ones encoding proteins of the TGF beta pathway or regulators of

the interactions of tumour cells with the immune system. This concept brings into focus the hypothesis that while at normal levels MYC acts as a global activator of gene regulation by generating productive Pol II elongation complexes, at high oncogenic levels, MYC additionally acts as a specific repressor of genes. Among other factors, this could explain the selection of cells with high levels of MYC during tumour progression.

However, MYC driven super-squelching of elongation factors is also noted in this thesis in response to MG132 treatment. Since inhibition of MYC degradation results in locking of PAF proteins with MYC, this results in dysregulation of MYC target genes (Jaenicke et al., 2016). The MG132 driven gene regulation and MYC squelching data presented here are in agreement with these observations since MYC and the associated elongation factors, SPT5 and SPT6, are actually squelched away from nucleoplasm where MYC target genes' DNA resides, and is shuttled into nucleolar, PML compartments and possibly other storage compartments. This indeed needs to be demonstrated conclusively by measuring the chromatin association of MYC, SPT5 and SPT6 after MG132 treatment using ChIP-Seq, and indicates scope for future work as an extension of this study. Further, other interactors could also be a part of these storage compartments, and studying the interactome of MYC in presence of MG132 could clarify the relevance of these compartments.

However, future work will determine if proteasomal inhibition is the only stress which results in MYC squelching and if this has any impact on ribosomal maturation, since MYC actively evades the site of rRNA transcription while occupying nucleolar space under MG132 treatment.

5.5 Conclusions

In order to find the exact molecular mechanism by which MYC regulates transcription, a multi-omics approach was utilized, which included analysis of nascent transcription (4sU-Seq) combined with measurement of precise elongation rates in kilobases per minute (4sUDRB-Seq). The data from these techniques was then combined with sophisticated bioinformatic workflows to establish a model of rate-limiting step in transcription cycle affected by MYC. To find the mechanisms how MYC affects this step, a combination of mass spectrometry-based proteomics, biochemical protein-protein proximity assay and small molecule inhibitor screen was used.

In absence of MYC, elongation rates of Pol II, as measured by 4sUDRB-Seq, decrease to 2700 bases/minute as compared 3200 bases/minute in normal cells. This was accompanied by a loss of processive Pol II in cells without MYC, as noted by 4sU-Seq. This regulation of elongation rate and processivity was found to be in strong correlation with MYC driven gene regulation. Thus, these results establish transcription elongation as the key rate limiting step for MYC's impact on changing mRNA levels of all genes in the cell. Surprisingly, this loss of elongation rate is not only accompanied by reduction in the number of Pol II molecules which complete transcription to make meaningful transcripts, but also by an increased number of Pol II molecules transcribing in anti-sense direction. These results showcase MYC's ability to switch the transcription machinery from pervasive mode to productive mode, resulting in generation of stable full-length mRNAs and thus driving gene regulation through an unexpected mechanism.

Mass spectrometric analyses combined with biochemical approaches revealed that MYC directly binds to a critical elongation factor SPT5 and hands it over to Pol II in a CDK 7 dependent manner. Importantly, chemical inhibition of only CDK7, and not CDK9 resulted in loss of MYC's ability to hand over SPT5 to Pol II at the promoters. These results showcase MYC as central regulatory hub of biology in the cell due to its ability to integrate signals from growth factors. This brings cells from resting to proliferating stage by rapidly switching transcription to a productive mode by handing over SPT5 to Pol II. As a result of this handover, generation of stable mRNAs and building up of biomass takes place due to growth-stimuli-induced MYC activation.

In order to find the exact mechanism of gene regulation, when the levels of MYC are increased to high levels in the cells, a similar multi-omics approach as above was applied to study Pol II behaviour. Surprisingly, elongation rate and processivity of Pol II did not further increase with rise in MYC levels, but actually decreased. This caused the genes to get repressed by premature termination of Pol II, especially on genes which are involved in immune surveillance of tumour cells. Further biochemical analyses revealed that this loss of elongation was due to MYC's ability to squelch away elongation factors like SPT5 from the Pol II transcription complex, thus reducing the ability of Pol II to make full length transcripts. The squelching phenomenon was not only noted at high concentrations of MYC, but was also found in cells where proteasomal machinery was inhibited by adding MG132.

The work presented here shows that specificity of gene regulation by MYC is largely generated by the mechanism by which MYC regulates Pol II function. Specifically, the high levels of MYC that are expressed in human tumours do not further enhance transcriptional elongation but rather bring about squelching of SPT5, reducing the processivity of Pol II transcription, especially on genes that are known targets of MYC-dependent repression, like genes encoding proteins of the TGF-beta pathway and regulators of the interactions of tumour cells with the immune system. This body of work has thus unravelled that tumours exploit MYC-dependent sequestration of SPT5 to repress tumour-suppressive genes, providing an explanation as to how the high levels of MYC cause tumour progression.

6 REFERENCES

Abbe, E. (1873). Ueber einen neuen Beleuchtungsapparat am Mikroskop. *Arch. Für Mikrosk. Anat.* *9*, 469–480.

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

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

Annibali, D., Whitfield, J.R., Favuzzi, E., Jauset, T., Serrano, E., Cuartas, I., Redondo-Campos, S., Folch, G., González-Juncà, A., Sodik, N.M., et al. (2014). Myc inhibition is effective against glioma and reveals a role for Myc in proficient mitosis. *Nat. Commun.* *5*, 4632.

Arabi, A. (2003). Accumulation of c-Myc and proteasomes at the nucleoli of cells containing elevated c-Myc protein levels. *J. Cell Sci.* *116*, 1707–1717.

Baluapuri, A., Hofstetter, J., Dudvarski Stankovic, N., Endres, T., Bhandare, P., Vos, S.M., Adhikari, B., Schwarz, J.D., Narain, A., Vogt, M., et al. (2019). MYC Recruits SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation. *Mol. Cell* *74*, 674-687.e11.

Baluapuri, A., Wolf, E., and Eilers, M. (2020). Target gene-independent functions of MYC oncoproteins. *Nat. Rev. Mol. Cell Biol.*

Beaulieu, M.-E., Jauset, T., Massó-Vallés, D., Martínez-Martín, S., Rahl, P., Maltais, L., Zacarias-Fluck, M.F., Casacuberta-Serra, S., Serrano del Pozo, E., Fiore, C., et al. (2019). Intrinsic cell-penetrating activity propels Omomyc from proof of concept to viable anti-MYC therapy. *Sci. Transl. Med.* *11*, eaar5012.

Bensaude, O. (2011). Inhibiting eukaryotic transcription. Which compound to choose? How to evaluate its activity?: Which compound to choose? How to evaluate its activity? *Transcription* *2*, 103–108.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* *15*, 163–175.

Bernecky, C., and Taatjes, D.J. (2012). Activator–Mediator Binding Stabilizes RNA Polymerase II Orientation within the Human Mediator–RNA Polymerase II–TFIIF Assembly. *J. Mol. Biol.* *417*, 387–394.

Bernecky, C., Plitzko, J.M., and Cramer, P. (2017). Structure of a transcribing RNA polymerase II–DSIF complex reveals a multidentate DNA–RNA clamp. *Nat. Struct. Mol. Biol.* *24*, 809–815.

- Bild, A.H., Yao, G., Chang, J.T., Wang, Q., Potti, A., Chasse, D., Joshi, M.-B., Harpole, D., Lancaster, J.M., Berchuck, A., et al. (2006). Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439, 353–357.
- Blackwell, T.K., Huang, J., Ma, A., Kretzner, L., Alt, F.W., Eisenman, R.N., and Weintraub, H. (1993). Binding of myc proteins to canonical and noncanonical DNA sequences. *Mol. Cell. Biol.* 13, 5216–5224.
- Blackwood, E.M., and Eisenman, R.N. (1991). Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science* 251, 1211–1217.
- Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A.S., Yu, T., Marie-Nelly, H., McSwiggen, D.T., Kokic, G., Dailey, G.M., Cramer, P., et al. (2018). RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nat. Struct. Mol. Biol.* 25, 833–840.
- Boija, A., Klein, I.A., Sabari, B.R., Dall’Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* 175, 1842-1855.e16.
- Bortvin, A., and Winston, F. (1996). Evidence That Spt6p Controls Chromatin Structure by a Direct Interaction with Histones. *Science* 272, 1473–1476.
- Bouchard, C., Marquardt, J., Brás, A., Medema, R.H., and Eilers, M. (2004). Myc-induced proliferation and transformation require Akt-mediated phosphorylation of FoxO proteins. *EMBO J.* 23, 2830–2840.
- Brägelmann, J., Dammert, M.A., Dietlein, F., Heuckmann, J.M., Choidas, A., Böhm, S., Richters, A., Basu, D., Tischler, V., Lorenz, C., et al. (2017). Systematic Kinase Inhibitor Profiling Identifies CDK9 as a Synthetic Lethal Target in NUT Midline Carcinoma. *Cell Rep.* 20, 2833–2845.
- Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Julicher, F., and Hyman, A.A. (2009). Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science* 324, 1729–1732.
- Büchel, G., Carstensen, A., Mak, K.-Y., Roeschert, I., Leen, E., Sumara, O., Hofstetter, J., Herold, S., Kalb, J., Baluapuri, A., et al. (2017). Association with Aurora-A Controls N-MYC-Dependent Promoter Escape and Pause Release of RNA Polymerase II during the Cell Cycle. *Cell Rep.* 21, 3483–3497.
- Cahill, M.A., Ernst, W.H., Janknecht, R., and Nordheim, A. (1994). Regulatory squelching. *FEBS Lett.* 344, 105–108.
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discov.* 2, 401–404.

- Chao, S.-H., and Price, D.H. (2001). Flavopiridol Inactivates P-TEFb and Blocks Most RNA Polymerase II Transcription *in Vivo*. *J. Biol. Chem.* *276*, 31793–31799.
- Chen, F., Gao, X., and Shilatifard, A. (2015). Stably paused genes revealed through inhibition of transcription initiation by the TFIIH inhibitor triptolide. *Genes Dev.* *29*, 39–47.
- Chen, F.X., Smith, E.R., and Shilatifard, A. (2018). Born to run: control of transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* *19*, 464–478.
- Chen, J., Brunner, A.-D., Cogan, J.Z., Nuñez, J.K., Fields, A.P., Adamson, B., Itzhak, D.N., Li, J.Y., Mann, M., Leonetti, M.D., et al. (2020). Pervasive functional translation of noncanonical human open reading frames. *Science* *367*, 1140–1146.
- Cheng, S.-W.G., Davies, K.P., Yung, E., Beltran, R.J., Yu, J., and Kalpana, G.V. (1999). c-MYC interacts with INI1/hSNF5 and requires the SWI/SNF complex for transactivation function. *Nat. Genet.* *22*, 102–105.
- Chipumuro, E., Marco, E., Christensen, C.L., Kwiatkowski, N., Zhang, T., Hatheway, C.M., Abraham, B.J., Sharma, B., Yeung, C., Altabef, A., et al. (2014). CDK7 Inhibition Suppresses Super-Enhancer-Linked Oncogenic Transcription in MYCN-Driven Cancer. *Cell* *159*, 1126–1139.
- Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., et al. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* *361*, eaar2555.
- Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* *469*, 368–373.
- Conacci-Sorrell, M., McFerrin, L., and Eisenman, R.N. (2014). An Overview of MYC and Its Interactome. *Cold Spring Harb. Perspect. Med.* *4*, a014357–a014357.
- Conaway, R.C., and Conaway, J.W. (2019). The hunt for RNA polymerase II elongation factors: a historical perspective. *Nat. Struct. Mol. Biol.* *26*, 771–776.
- Conaway, R.C., Garrett, K.P., Hanley, J.P., and Conaway, J.W. (1991). Mechanism of promoter selection by RNA polymerase II: mammalian transcription factors alpha and beta gamma promote entry of polymerase into the preinitiation complex. *Proc. Natl. Acad. Sci.* *88*, 6205–6209.
- Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science* *322*, 1845–1848.
- Cortazar, M.A., Sheridan, R.M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K., and Bentley, D.L. (2019). Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a “Sitting Duck Torpedo” Mechanism. *Mol. Cell* *76*, 896-908.e4.

- Cossa, G., Roeschert, I., Prinz, F., Baluapuri, A., Silveira Vidal, R., Schüle-Völk, C., Chang, Y.-C., Ade, C.P., Mastrobuoni, G., Girard, C., et al. (2020). Localized Inhibition of Protein Phosphatase 1 by NUA1 Promotes Spliceosome Activity and Reveals a MYC-Sensitive Feedback Control of Transcription. *Mol. Cell* 77, 1322-1339.e11.
- Costlow, N., and Lis, J.T. (1984). High-resolution mapping of DNase I-hypersensitive sites of *Drosophila* heat shock genes in *Drosophila melanogaster* and *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 4, 1853–1863.
- Cowling, V.H., and Cole, M.D. (2007). The Myc Transactivation Domain Promotes Global Phosphorylation of the RNA Polymerase II Carboxy-Terminal Domain Independently of Direct DNA Binding. *Mol. Cell. Biol.* 27, 2059–2073.
- Cramer, P. (2019a). Eukaryotic Transcription Turns 50. *Cell* 179, 808–812.
- Cramer, P. (2019b). Organization and regulation of gene transcription. *Nature* 573, 45–54.
- Dang, C.V. (2012). MYC on the path to cancer. *Cell* 149, 22–35.
- Dauch, D., Rudalska, R., Cossa, G., Nault, J.-C., Kang, T.-W., Wuestefeld, T., Hohmeyer, A., Imbeaud, S., Yevsa, T., Hoenicke, L., et al. (2016). A MYC–aurora kinase A protein complex represents an actionable drug target in p53-altered liver cancer. *Nat. Med.* 22, 744–753.
- Derti, A., Garrett-Engele, P., Macisaac, K.D., Stevens, R.C., Sriram, S., Chen, R., Rohl, C.A., Johnson, J.M., and Babak, T. (2012). A quantitative atlas of polyadenylation in five mammals. *Genome Res.* 22, 1173–1183.
- Dubois, N.C., Adolphe, C., Ehninger, A., Wang, R.A., Robertson, E.J., and Trumpp, A. (2008). Placental rescue reveals a sole requirement for c-Myc in embryonic erythroblast survival and hematopoietic stem cell function. *Development* 135, 2455–2465.
- Dynan, W.S., and Tjian, R. (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* 35, 79–87.
- Eberhardy, S.R., and Farnham, P.J. (2002). Myc Recruits P-TEFb to Mediate the Final Step in the Transcriptional Activation of the *cad* Promoter. *J. Biol. Chem.* 277, 40156–40162.
- Ehara, H., Yokoyama, T., Shigematsu, H., Yokoyama, S., Shirouzu, M., and Sekine, S. (2017). Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* 357, 921–924.
- Ehrensberger, A.H., Kelly, G.P., and Svejstrup, J.Q. (2013). Mechanistic Interpretation of Promoter-Proximal Peaks and RNAPII Density Maps. *Cell* 154, 713–715.
- Eilers, M., and Eisenman, R.N. (2008). Myc’s broad reach. *Genes Dev.* 22, 2755–2766.

References

- Eilers, M., Picard, D., Yamamoto, K.R., and Bishop, J.M. (1989). Chimaeras of Myc oncoprotein and steroid receptors cause hormone-dependent transformation of cells. *Nature* *340*, 66–68.
- Erhard, F., Dölken, L., Schilling, B., and Schlosser, A. (2020). Identification of the cryptic HLA-I immunopeptidome. *Cancer Immunol. Res.* *canimm.0886.2019*.
- Faiola, F., Liu, X., Lo, S., Pan, S., Zhang, K., Lymar, E., Farina, A., and Martinez, E. (2005). Dual Regulation of c-Myc by p300 via Acetylation-Dependent Control of Myc Protein Turnover and Coactivation of Myc-Induced Transcription. *Mol. Cell Biol.* *25*, 10220–10234.
- Farrell, A.S., and Sears, R.C. (2014). MYC Degradation. *Cold Spring Harb. Perspect. Med.* *4*, a014365–a014365.
- Felsher, D.W., and Bishop, J.M. (1999). Reversible tumorigenesis by MYC in hematopoietic lineages. *Mol. Cell* *4*, 199–207.
- Fernandez, P.C. (2003). Genomic targets of the human c-Myc protein. *Genes Dev.* *17*, 1115–1129.
- Fisher, R.P. (2020). Splice or Die: When MYC Is Driving, Transcription Needs NUA1 to Avoid Fatal Pileups. *Mol. Cell* *77*, 1157–1158.
- Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., Diener, K., Jones, K., Fu, X.-D., and Bentley, D.L. (2014). Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev.* *28*, 2663–2676.
- Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M.A., Sheridan, R.M., Nguyen, T., Karp, S., and Bentley, D.L. (2015). Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Mol. Cell* *60*, 256–267.
- Fong, N., Saldi, T., Sheridan, R.M., Cortazar, M.A., and Bentley, D.L. (2017). RNA Pol II Dynamics Modulate Co-transcriptional Chromatin Modification, CTD Phosphorylation, and Transcriptional Direction. *Mol. Cell* *66*, 546-557.e3.
- Frank, S.R., Parisi, T., Taubert, S., Fernandez, P., Fuchs, M., Chan, H., Livingston, D.M., and Amati, B. (2003). MYC recruits the TIP60 histone acetyltransferase complex to chromatin. *EMBO Rep.* *4*, 575–580.
- Fuchs, G., Voichek, Y., Rabani, M., Benjamin, S., Gilad, S., Amit, I., and Oren, M. (2015). Simultaneous measurement of genome-wide transcription elongation speeds and rates of RNA polymerase II transition into active elongation with 4sUDRB-seq. *Nat. Protoc.* *10*, 605–618.
- Fuchs, M., Gerber, J., Drapkin, R., Sif, S., Ikura, T., Ogryzko, V., Lane, W.S., Nakatani, Y., and Livingston, D.M. (2001). The p400 Complex Is an Essential E1A Transformation Target. *Cell* *106*, 297–307.

- Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* *461*, 186–192.
- Gabay, M., Li, Y., and Felsher, D.W. (2014). MYC Activation Is a Hallmark of Cancer Initiation and Maintenance. *Cold Spring Harb. Perspect. Med.* *4*, a014241–a014241.
- Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* *6*, p11–p11.
- Gartel, A.L., Ye, X., Goufman, E., Shianov, P., Hay, N., Najmabadi, F., and Tyner, A.L. (2001). Myc represses the p21(WAF1/CIP1) promoter and interacts with Sp1/Sp3. *Proc. Natl. Acad. Sci.* *98*, 4510–4515.
- Gerlach, J.M., Furrer, M., Gallant, M., Birkel, D., Baluapuri, A., Wolf, E., and Gallant, P. (2017). PAF1 complex component Leo1 helps recruit *Drosophila* Myc to promoters. *Proc. Natl. Acad. Sci.* *114*, E9224–E9232.
- Gilchrist, D.A., Nechaev, S., Lee, C., Ghosh, S.K.B., Collins, J.B., Li, L., Gilmour, D.S., and Adelman, K. (2008). NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes Dev.* *22*, 1921–1933.
- Gill, G., and Ptashne, M. (1988). Negative effect of the transcriptional activator GAL4. *Nature* *334*, 721–724.
- Goodrich, J.A., and Tjian, R. (1994). TBP-TAF complexes: selectivity factors for eukaryotic transcription. *Curr. Opin. Cell Biol.* *6*, 403–409.
- Gossen, M., and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci.* *89*, 5547–5551.
- Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., and Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. *Science* *268*, 1766–1769.
- Gregersen, L.H., Mitter, R., Ugalde, A.P., Nojima, T., Proudfoot, N.J., Agami, R., Stewart, A., and Svejstrup, J.Q. (2019). SCAF4 and SCAF8, mRNA Anti-Terminator Proteins. *Cell* *177*, 1797-1813.e18.
- Gregersen, L.H., Mitter, R., and Svejstrup, J.Q. (2020). Using TTchem-seq for profiling nascent transcription and measuring transcript elongation. *Nat. Protoc.* *15*, 604–627.
- Grohmann, D., Nagy, J., Chakraborty, A., Klose, D., Fielden, D., Ebright, R.H., Michaelis, J., and Werner, F. (2011). The Initiation Factor TFE and the Elongation Factor Spt4/5 Compete for the RNAP Clamp during Transcription Initiation and Elongation. *Mol. Cell* *43*, 263–274.
- Guccione, E., Martinato, F., Finocchiaro, G., Luzi, L., Tizzoni, L., Dall’Olio, V., Zardo, G., Nervi, C., Bernard, L., and Amati, B. (2006). Myc-binding-site recognition in the human genome is determined by chromatin context. *Nat. Cell Biol.* *8*, 764–770.

- Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. *Cell* 130, 77–88.
- Guo, J., Li, T., Schipper, J., Nilson, K.A., Fordjour, F.K., Cooper, J.J., Gordân, R., and Price, D.H. (2014). Sequence specificity incompletely defines the genome-wide occupancy of Myc. *Genome Biol.* 15, 482.
- Guo, Y.E., Manteiga, J.C., Henninger, J.E., Sabari, B.R., Dall’Agnese, A., Hannett, N.M., Spille, J.-H., Afeyan, L.K., Zamudio, A.V., Shrinivas, K., et al. (2019). Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* 572, 543–548.
- Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of Cancer: The Next Generation. *Cell* 144, 646–674.
- Hayday, A.C., Gillies, S.D., Saito, H., Wood, C., Wiman, K., Hayward, W.S., and Tonegawa, S. (1984). Activation of a translocated human c-myc gene by an enhancer in the immunoglobulin heavy-chain locus. *Nature* 307, 334–340.
- He, T. (1998). Identification of c-MYC as a Target of the APC Pathway. *Science* 281, 1509–1512.
- Herkert, B., and Eilers, M. (2010). Transcriptional repression: the dark side of myc. *Genes Cancer* 1, 580–586.
- Herold, S., Kalb, J., Büchel, G., Ade, C.P., Baluapuri, A., Xu, J., Koster, J., Solvie, D., Carstensen, A., Klotz, C., et al. (2019). Recruitment of BRCA1 limits MYCN-driven accumulation of stalled RNA polymerase. *Nature* 567, 545–549.
- Hofmann, J.W., Zhao, X., De Cecco, M., Peterson, A.L., Pagliaroli, L., Manivannan, J., Hubbard, G.B., Ikeno, Y., Zhang, Y., Feng, B., et al. (2015). Reduced Expression of MYC Increases Longevity and Enhances Healthspan. *Cell* 160, 477–488.
- Huang, C.-H., Lujambio, A., Zuber, J., Tschaharganeh, D.F., Doran, M.G., Evans, M.J., Kitzing, T., Zhu, N., de Stanchina, E., Sawyers, C.L., et al. (2014). CDK9-mediated transcription elongation is required for MYC addiction in hepatocellular carcinoma. *Genes Dev.* 28, 1800–1814.
- Jaenicke, L.A., von Eyss, B., Carstensen, A., Wolf, E., Xu, W., Greifenberg, A.K., Geyer, M., Eilers, M., and Popov, N. (2016). Ubiquitin-Dependent Turnover of MYC Antagonizes MYC/PAF1C Complex Accumulation to Drive Transcriptional Elongation. *Mol. Cell* 61, 54–67.
- Jeronimo, C., and Robert, F. (2014). Kin28 regulates the transient association of Mediator with core promoters. *Nat. Struct. Mol. Biol.* 21, 449–455.
- Kadonaga, J.T. (2019). The transformation of the DNA template in RNA polymerase II transcription: a historical perspective. *Nat. Struct. Mol. Biol.* 26, 766–770.

- Kalkat, M., Resetca, D., Lourenco, C., Chan, P.-K., Wei, Y., Shiah, Y.-J., Vitkin, N., Tong, Y., Sunnerhagen, M., Done, S.J., et al. (2018). MYC Protein Interactome Profiling Reveals Functionally Distinct Regions that Cooperate to Drive Tumorigenesis. *Mol. Cell* *72*, 836-848.e7.
- Kanazawa, S., Soucek, L., Evan, G., Okamoto, T., and Peterlin, B.M. (2003). c-Myc recruits P-TEFb for transcription, cellular proliferation and apoptosis. *Oncogene* *22*, 5707–5711.
- Kawauchi, D., Robinson, G., Uziel, T., Gibson, P., Rehg, J., Gao, C., Finkelstein, D., Qu, C., Pounds, S., Ellison, D.W., et al. (2012). A mouse model of the most aggressive subgroup of human medulloblastoma. *Cancer Cell* *21*, 168–180.
- Kelso, T.W.R., Baumgart, K., Eickhoff, J., Albert, T., Antrecht, C., Lemcke, S., Klebl, B., and Meisterernst, M. (2014). Cyclin-Dependent Kinase 7 Controls mRNA Synthesis by Affecting Stability of Preinitiation Complexes, Leading to Altered Gene Expression, Cell Cycle Progression, and Survival of Tumor Cells. *Mol. Cell. Biol.* *34*, 3675–3688.
- Kim, J.B., and Sharp, P.A. (2001). Positive Transcription Elongation Factor b Phosphorylates hSPT5 and RNA Polymerase II Carboxyl-terminal Domain Independently of Cyclin-dependent Kinase-activating Kinase. *J. Biol. Chem.* *276*, 12317–12323.
- Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells. *Cell* *132*, 1049–1061.
- Kortlever, R.M., Sodik, N.M., Wilson, C.H., Burkhart, D.L., Pellegrinet, L., Brown Swigart, L., Littlewood, T.D., and Evan, G.I. (2017). Myc Cooperates with Ras by Programming Inflammation and Immune Suppression. *Cell* *171*, 1301-1315.e14.
- Kouzine, F., Wojtowicz, D., Yamane, A., Resch, W., Kieffer-Kwon, K.-R., Bandle, R., Nelson, S., Nakahashi, H., Awasthi, P., Feigenbaum, L., et al. (2013). Global Regulation of Promoter Melting in Naive Lymphocytes. *Cell* *153*, 988–999.
- Krajewska, M., Dries, R., Grassetti, A.V., Dust, S., Gao, Y., Huang, H., Sharma, B., Day, D.S., Kwiatkowski, N., Pomaville, M., et al. (2019). CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation. *Nat. Commun.* *10*, 1757.
- Kress, T.R., Sabò, A., and Amati, B. (2015). MYC: connecting selective transcriptional control to global RNA production. *Nat. Rev. Cancer* *15*, 593–607.
- Kwiatkowski, N., Zhang, T., Rahl, P.B., Abraham, B.J., Reddy, J., Ficarro, S.B., Dastur, A., Amzallag, A., Ramaswamy, S., Tesar, B., et al. (2014). Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* *511*, 616–620.
- Lafontaine, D.L.J., Riback, J.A., Bascetin, R., and Brangwynne, C.P. (2020). The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell Biol.*
- Lallemand-Breitenbach, V., and de The, H. (2010). PML Nuclear Bodies. *Cold Spring Harb. Perspect. Biol.* *2*, a000661–a000661.

- Larochelle, S., Batliner, J., Gamble, M.J., Barboza, N.M., Kraybill, B.C., Blethrow, J.D., Shokat, K.M., and Fisher, R.P. (2006). Dichotomous but stringent substrate selection by the dual-function Cdk7 complex revealed by chemical genetics. *Nat. Struct. Mol. Biol.* *13*, 55–62.
- Larochelle, S., Amat, R., Glover-Cutter, K., Sansó, M., Zhang, C., Allen, J.J., Shokat, K.M., Bentley, D.L., and Fisher, R.P. (2012). Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. *Nat. Struct. Mol. Biol.* *19*, 1108–1115.
- Lawlor, E.R., Soucek, L., Brown-Swigart, L., Shchors, K., Bialucha, C.U., and Evan, G.I. (2006). Reversible Kinetic Analysis of Myc Targets *In vivo* Provides Novel Insights into Myc-Mediated Tumorigenesis. *Cancer Res.* *66*, 4591–4601.
- Levens, D. (2013). Cellular MYC Economics: Balancing MYC Function with MYC Expression. *Cold Spring Harb. Perspect. Med.* *3*, a014233–a014233.
- Lewis, L.M., Edwards, M.C., Meyers, Z.R., Talbot, C.C., Hao, H., Blum, D., Reproducibility Project: Cancer Biology, Iorns, E., Tsui, R., Denis, A., et al. (2018). Replication Study: Transcriptional amplification in tumor cells with elevated c-Myc. *ELife* *7*, e30274.
- Liang, K., Smith, E.R., Aoi, Y., Stoltz, K.L., Katagi, H., Woodfin, A.R., Rendleman, E.J., Marshall, S.A., Murray, D.C., Wang, L., et al. (2018). Targeting Processive Transcription Elongation via SEC Disruption for MYC-Induced Cancer Therapy. *Cell* *175*, 766–779.e17.
- Lin, C.Y., Lovén, J., Rahl, P.B., Paranal, R.M., Burge, C.B., Bradner, J.E., Lee, T.I., and Young, R.A. (2012). Transcriptional Amplification in Tumor Cells with Elevated c-Myc. *Cell* *151*, 56–67.
- van de Linde, S., Löschberger, A., Klein, T., Heidbreder, M., Wolter, S., Heilemann, M., and Sauer, M. (2011). Direct stochastic optical reconstruction microscopy with standard fluorescent probes. *Nat. Protoc.* *6*, 991–1009.
- Lis, J.T., Mason, P., Peng, J., Price, D.H., and Werner, J. (2000). P-TEFb kinase recruitment and function at heat shock loci. *Genes Dev.* *14*, 792–803.
- Lorenzin, F., Benary, U., Baluapuri, A., Walz, S., Jung, L.A., von Eyss, B., Kisker, C., Wolf, J., Eilers, M., and Wolf, E. (2016). Different promoter affinities account for specificity in MYC-dependent gene regulation. *ELife* *5*.
- Lovén, J., Orlando, D.A., Sigova, A.A., Lin, C.Y., Rahl, P.B., Burge, C.B., Levens, D.L., Lee, T.I., and Young, R.A. (2012). Revisiting Global Gene Expression Analysis. *Cell* *151*, 476–482.
- Marshall, N.F., and Price, D.H. (1992). Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Mol. Cell. Biol.* *12*, 2078–2090.
- Mason, P.B., and Struhl, K. (2005). Distinction and Relationship between Elongation Rate and Processivity of RNA Polymerase II *In Vivo*. *Mol. Cell* *17*, 831–840.

- Massó-Vallés, D., Beaulieu, M.-E., and Soucek, L. (2020). MYC, MYCL, and MYCN as therapeutic targets in lung cancer. *Expert Opin. Ther. Targets* *24*, 101–114.
- McKeown, M.R., and Bradner, J.E. (2014). Therapeutic Strategies to Inhibit MYC. *Cold Spring Harb. Perspect. Med.* *4*, a014266–a014266.
- McMahon, S.B., Van Buskirk, H.A., Dugan, K.A., Copeland, T.D., and Cole, M.D. (1998). The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins. *Cell* *94*, 363–374.
- McMahon, S.B., Wood, M.A., and Cole, M.D. (2000). The essential cofactor TRRAP recruits the histone acetyltransferase hGCN5 to c-Myc. *Mol. Cell. Biol.* *20*, 556–562.
- McSwiggen, D.T., Hansen, A.S., Teves, S.S., Marie-Nelly, H., Hao, Y., Heckert, A.B., Umemoto, K.K., Dugast-Darzacq, C., Tjian, R., and Darzacq, X. (2019). Evidence for DNA-mediated nuclear compartmentalization distinct from phase separation. *ELife* *8*.
- Meyer, N., and Penn, L.Z. (2008). Reflecting on 25 years with MYC. *Nat. Rev. Cancer* *8*, 976–990.
- Michel, M., and Cramer, P. (2013). Transitions for Regulating Early Transcription. *Cell* *153*, 943–944.
- Missra, A., and Gilmour, D.S. (2010). Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *Proc. Natl. Acad. Sci.* *107*, 11301–11306.
- Muhar, M., Ebert, A., Neumann, T., Umkehrer, C., Jude, J., Wieshofer, C., Rescheneder, P., Lipp, J.J., Herzog, V.A., Reichholf, B., et al. (2018). SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. *Science* *360*, 800–805.
- Nair, S.K., and Burley, S.K. (2003). X-Ray Structures of Myc-Max and Mad-Max Recognizing DNA. *Cell* *112*, 193–205.
- Nie, Z., Hu, G., Wei, G., Cui, K., Yamane, A., Resch, W., Wang, R., Green, D.R., Tessarollo, L., Casellas, R., et al. (2012). c-Myc Is a Universal Amplifier of Expressed Genes in Lymphocytes and Embryonic Stem Cells. *Cell* *151*, 68–79.
- Nilson, K.A., Guo, J., Turek, M.E., Brogie, J.E., Delaney, E., Luse, D.S., and Price, D.H. (2015). THZ1 Reveals Roles for Cdk7 in Co-transcriptional Capping and Pausing. *Mol. Cell* *59*, 576–587.
- Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods* *6*, 917–922.
- Noe Gonzalez, M., Blears, D., and Svejstrup, J.Q. (2020). Causes and consequences of RNA polymerase II stalling during transcript elongation. *Nat. Rev. Mol. Cell Biol.*

- Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* *161*, 526–540.
- Orphanides, G., LeRoy, G., Chang, C.-H., Luse, D.S., and Reinberg, D. (1998). FACT, a Factor that Facilitates Transcript Elongation through Nucleosomes. *Cell* *92*, 105–116.
- Palangat, M., Renner, D.B., Price, D.H., and Landick, R. (2005). A negative elongation factor for human RNA polymerase II inhibits the anti-arrest transcript-cleavage factor TFIIS. *Proc. Natl. Acad. Sci.* *102*, 15036–15041.
- Palomero, T., Lim, W.K., Odom, D.T., Sulis, M.L., Real, P.J., Margolin, A., Barnes, K.C., O’Neil, J., Neuberg, D., Weng, A.P., et al. (2006). NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. *Proc. Natl. Acad. Sci.* *103*, 18261–18266.
- Park, P.J. (2009). ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* *10*, 669–680.
- Parua, P.K., and Fisher, R.P. (2020). Dissecting the Pol II transcription cycle and derailing cancer with CDK inhibitors. *Nat. Chem. Biol.* *16*, 716–724.
- Parua, P.K., Booth, G.T., Sansó, M., Benjamin, B., Tanny, J.C., Lis, J.T., and Fisher, R.P. (2018). A Cdk9-PP1 switch regulates the elongation-termination transition of RNA polymerase II. *Nature* *558*, 460–464.
- Parua, P.K., Kalan, S., Benjamin, B., Sansó, M., and Fisher, R.P. (2020). Distinct Cdk9-phosphatase switches act at the beginning and end of elongation by RNA polymerase II (Molecular Biology).
- Pei, Y., and Shuman, S. (2002). Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *J. Biol. Chem.* *277*, 19639–19648.
- Peukert, K., Staller, P., Schneider, A., Carmichael, G., Hänel, F., and Eilers, M. (1997). An alternative pathway for gene regulation by Myc. *EMBO J.* *16*, 5672–5686.
- Popov, N., Wanzel, M., Madiredjo, M., Zhang, D., Beijersbergen, R., Bernards, R., Moll, R., Elledge, S.J., and Eilers, M. (2007). The ubiquitin-specific protease USP28 is required for MYC stability. *Nat. Cell Biol.* *9*, 765–774.
- Prendergast, G.C., Lawe, D., and Ziff, E.B. (1991). Association of Myn, the murine homolog of Max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation. *Cell* *65*, 395–407.
- de Pretis, S., Kress, T.R., Morelli, M.J., Sabò, A., Locarno, C., Verrecchia, A., Doni, M., Campaner, S., Amati, B., and Pelizzola, M. (2017). Integrative analysis of RNA polymerase II and transcriptional dynamics upon MYC activation. *Genome Res.* *27*, 1658–1664.
- Price, D.H. (2010). Regulation of RNA Polymerase II Elongation by c-Myc. *Cell* *141*, 399–400.

- Price, D.H., Sluder, A.E., and Greenleaf, A.L. (1989). Dynamic interaction between a *Drosophila* transcription factor and RNA polymerase II. *Mol. Cell. Biol.* *9*, 1465–1475.
- Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science* *352*, aad9926–aad9926.
- Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. *Cell* *141*, 432–445.
- Richards, M.W., Burgess, S.G., Poon, E., Carstensen, A., Eilers, M., Chesler, L., and Bayliss, R. (2016). Structural basis of N-Myc binding by Aurora-A and its destabilization by kinase inhibitors. *Proc. Natl. Acad. Sci.* *113*, 13726–13731.
- van Riggelen, J., Muller, J., Otto, T., Beuger, V., Yetil, A., Choi, P.S., Kosan, C., Moroy, T., Felsher, D.W., and Eilers, M. (2010). The interaction between Myc and Miz1 is required to antagonize TGF- β -dependent autocrine signaling during lymphoma formation and maintenance. *Genes Dev.* *24*, 1281–1294.
- Roussel, M.F., and Robinson, G.W. (2013). Role of MYC in Medulloblastoma. *Cold Spring Harb. Perspect. Med.* *3*.
- ar-Rushdi, A., Nishikura, K., Erikson, J., Watt, R., Rovera, G., and Croce, C. (1983). Differential expression of the translocated and the untranslocated c-myc oncogene in Burkitt lymphoma. *Science* *222*, 390–393.
- Sabari, B.R., Dall’Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science* *361*, eaar3958.
- Sabò, A., Kress, T.R., Pelizzola, M., de Pretis, S., Gorski, M.M., Tesi, A., Morelli, M.J., Bora, P., Doni, M., Verrecchia, A., et al. (2014). Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis. *Nature* *511*, 488–492.
- Saldi, T., Cortazar, M.A., Sheridan, R.M., and Bentley, D.L. (2016). Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *J. Mol. Biol.* *428*, 2623–2635.
- Sampathi, S., Acharya, P., Zhao, Y., Wang, J., Stengel, K.R., Liu, Q., Savona, M.R., and Hiebert, S.W. (2019). The CDK7 inhibitor THZ1 alters RNA polymerase dynamics at the 5' and 3' ends of genes. *Nucleic Acids Res.* *47*, 3921–3936.
- Sansom, O.J., Meniel, V.S., Muncan, V., Pheasant, T.J., Wilkins, J.A., Reed, K.R., Vass, J.K., Athineos, D., Clevers, H., and Clarke, A.R. (2007). Myc deletion rescues Apc deficiency in the small intestine. *Nature* *446*, 676–679.
- Saponaro, M., Kantidakis, T., Mitter, R., Kelly, G.P., Heron, M., Williams, H., Söding, J., Stewart, A., and Svejstrup, J.Q. (2014). RECQL5 Controls Transcript Elongation and Suppresses Genome Instability Associated with Transcription Stress. *Cell* *157*, 1037–1049.

References

- Schaub, F.X., Dhankani, V., Berger, A.C., Trivedi, M., Richardson, A.B., Shaw, R., Zhao, W., Zhang, X., Ventura, A., Liu, Y., et al. (2018). Pan-cancer Alterations of the MYC Oncogene and Its Proximal Network across the Cancer Genome Atlas. *Cell Syst.* *6*, 282–300.e2.
- Schermelleh, L., Carlton, P.M., Haase, S., Shao, L., Winoto, L., Kner, P., Burke, B., Cardoso, M.C., Agard, D.A., Gustafsson, M.G.L., et al. (2008). Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science* *320*, 1332–1336.
- Schmitz, R., Ceribelli, M., Pittaluga, S., Wright, G., and Staudt, L.M. (2014). Oncogenic Mechanisms in Burkitt Lymphoma. *Cold Spring Harb. Perspect. Med.* *4*, a014282–a014282.
- Schüller, R., Forné, I., Straub, T., Schreieck, A., Texier, Y., Shah, N., Decker, T.-M., Cramer, P., Imhof, A., and Eick, D. (2016). Heptad-Specific Phosphorylation of RNA Polymerase II CTD. *Mol. Cell* *61*, 305–314.
- Seoane, J., Pouponnot, C., Staller, P., Schader, M., Eilers, M., and Massagué, J. (2001). TGF β influences Myc, Miz-1 and Smad to control the CDK inhibitor p15INK4b. *Nat. Cell Biol.* *3*, 400–408.
- Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* *15*, 284.
- Shetty, A., Kallgren, S.P., Demel, C., Maier, K.C., Spatt, D., Alver, B.H., Cramer, P., Park, P.J., and Winston, F. (2017). Spt5 Plays Vital Roles in the Control of Sense and Antisense Transcription Elongation. *Mol. Cell* *66*, 77–88.e5.
- Shilatifard, A., Conaway, R.C., and Conaway, J.W. (2003). The RNA Polymerase II Elongation Complex. *Annu. Rev. Biochem.* *72*, 693–715.
- Shou, Y., Martelli, M.L., Gabrea, A., Qi, Y., Brents, L.A., Roschke, A., Dewald, G., Kirsch, I.R., Bergsagel, P.L., and Kuehl, W.M. (2000). Diverse karyotypic abnormalities of the c-myc locus associated with c-myc dysregulation and tumor progression in multiple myeloma. *Proc. Natl. Acad. Sci.* *97*, 228–233.
- Slobodin, B., Han, R., Calderone, V., Vrieling, J.A.F.O., Loayza-Puch, F., Elkon, R., and Agami, R. (2017). Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. *Cell* *169*, 326–337.e12.
- Smith, K.P., Byron, M., O’Connell, B.C., Tam, R., Schorl, C., Guney, I., Hall, L.L., Agrawal, P., Sedivy, J.M., and Lawrence, J.B. (2004). c-Myc localization within the nucleus: Evidence for association with the PML nuclear body. *J. Cell. Biochem.* *93*, 1282–1296.
- Sodir, N.M., Kortlever, R.M., Barthet, V.J.A., Campos, T., Pellegrinet, L., Kupczak, S., Anastasiou, P., Swigart, L.B., Soucek, L., Arends, M.J., et al. (2020). MYC Instructs and Maintains Pancreatic Adenocarcinoma Phenotype. *Cancer Discov.* *10*, 588–607.

- Sohal, D.S., Nghiem, M., Crackower, M.A., Witt, S.A., Kimball, T.R., Tymitz, K.M., Penninger, J.M., and Molkenin, J.D. (2001). Temporally regulated and tissue-specific gene manipulations in the adult and embryonic heart using a tamoxifen-inducible Cre protein. *Circ. Res.* *89*, 20–25.
- Solomon, D.L.C., Amati, B., and Land, H. (1993). Distinct DNA binding preferences for the c-Myc/Max and Max/Max dimers. *Nucleic Acids Res.* *21*, 5372–5376.
- Soucek, L., Whitfield, J., Martins, C.P., Finch, A.J., Murphy, D.J., Sodik, N.M., Karnezis, A.N., Swigart, L.B., Nasi, S., and Evan, G.I. (2008). Modelling Myc inhibition as a cancer therapy. *Nature* *455*, 679–683.
- Takahashi, K., and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* *126*, 663–676.
- Tesi, A., Pretis, S., Furlan, M., Filipuzzi, M., Morelli, M.J., Andronache, A., Doni, M., Verrecchia, A., Pelizzola, M., Amati, B., et al. (2019). An early Myc-dependent transcriptional program orchestrates cell growth during B-cell activation. *EMBO Rep.* *20*.
- The FANTOM Consortium, Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., et al. (2014). An atlas of active enhancers across human cell types and tissues. *Nature* *507*, 455–461.
- Thomas, L.R., Wang, Q., Grieb, B.C., Phan, J., Foshage, A.M., Sun, Q., Olejniczak, E.T., Clark, T., Dey, S., Lorey, S., et al. (2015). Interaction with WDR5 Promotes Target Gene Recognition and Tumorigenesis by MYC. *Mol. Cell* *58*, 440–452.
- Trumpp, A., Refaeli, Y., Oskarsson, T., Gasser, S., Murphy, M., Martin, G.R., and Bishop, J.M. (2001). c-Myc regulates mammalian body size by controlling cell number but not cell size. *Nature* *414*, 768–773.
- Veloso, A., Kirkconnell, K.S., Magnuson, B., Biewen, B., Paulsen, M.T., Wilson, T.E., and Ljungman, M. (2014). Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res.* *24*, 896–905.
- Vervoorts, J., Lüscher-Firzlauff, J.M., Rottmann, S., Lilischkis, R., Walsemann, G., Dohmann, K., Austen, M., and Lüscher, B. (2003). Stimulation of c-MYC transcriptional activity and acetylation by recruitment of the cofactor CBP. *EMBO Rep.* *4*, 484–490.
- Vo, B.T., Wolf, E., Kawauchi, D., Gebhardt, A., Rehg, J.E., Finkelstein, D., Walz, S., Murphy, B.L., Youn, Y.H., Han, Y.-G., et al. (2016). The Interaction of Myc with Miz1 Defines Medulloblastoma Subgroup Identity. *Cancer Cell* *29*, 5–16.
- Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P. (2018a). Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* *560*, 607–612.
- Vos, S.M., Farnung, L., Urlaub, H., and Cramer, P. (2018b). Structure of paused transcription complex Pol II-DSIF-NELF. *Nature* *560*, 601–606.

- Walz, S., Lorenzin, F., Morton, J., Wiese, K.E., von Eyss, B., Herold, S., Rycak, L., Dumay-Odelot, H., Karim, S., Bartkuhn, M., et al. (2014). Activation and repression by oncogenic MYC shape tumour-specific gene expression profiles. *Nature* *511*, 483–487.
- Wei, Y., Resetca, D., Li, Z., Johansson-Åkhe, I., Ahlner, A., Helander, S., Wallenhammar, A., Morad, V., Raught, B., Wallner, B., et al. (2019). Multiple direct interactions of TBP with the MYC oncoprotein. *Nat. Struct. Mol. Biol.* *26*, 1035–1043.
- Wiese, K.E., Walz, S., von Eyss, B., Wolf, E., Athineos, D., Sansom, O., and Eilers, M. (2013). The role of MIZ-1 in MYC-dependent tumorigenesis. *Cold Spring Harb. Perspect. Med.* *3*, a014290.
- Windhager, L., Bonfert, T., Burger, K., Ruzsics, Z., Krebs, S., Kaufmann, S., Malterer, G., L'Hernault, A., Schilhabel, M., Schreiber, S., et al. (2012). Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. *Genome Res.* *22*, 2031–2042.
- Wissink, E.M., Vihervaara, A., Tippens, N.D., and Lis, J.T. (2019). Nascent RNA analyses: tracking transcription and its regulation. *Nat. Rev. Genet.* *20*, 705–723.
- Wolf, E., and Eilers, M. (2020). Targeting MYC Proteins for Tumor Therapy. *Annu. Rev. Cancer Biol.* *4*, 61–75.
- Wolter, S., Löschberger, A., Holm, T., Aufmkolk, S., Dabauvalle, M.-C., van de Linde, S., and Sauer, M. (2012). rapidSTORM: accurate, fast open-source software for localization microscopy. *Nat. Methods* *9*, 1040–1041.
- Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* *97*, 41–51.
- Yin, X., Giap, C., Lazo, J.S., and Prochownik, E.V. (2003). Low molecular weight inhibitors of Myc–Max interaction and function. *Oncogene* *22*, 6151–6159.
- Zeid, R., Lawlor, M.A., Poon, E., Reyes, J.M., Fulciniti, M., Lopez, M.A., Scott, T.G., Nabet, B., Erb, M.A., Winter, G.E., et al. (2018). Enhancer invasion shapes MYCN-dependent transcriptional amplification in neuroblastoma. *Nat. Genet.* *50*, 515–523.
- Zeller, K.I., Jegga, A.G., Aronow, B.J., O'Donnell, K.A., and Dang, C.V. (2003). An integrated database of genes responsive to the Myc oncogenic transcription factor: identification of direct genomic targets. *Genome Biol.* *4*, R69.
- Zhang, N., Ichikawa, W., Faiola, F., Lo, S.-Y., Liu, X., and Martinez, E. (2014). MYC interacts with the human STAGA coactivator complex via multivalent contacts with the GCN5 and TRRAP subunits. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* *1839*, 395–405.
- Zhang, Y., Xue, W., Li, X., Zhang, J., Chen, S., Zhang, J.-L., Yang, L., and Chen, L.-L. (2016). The Biogenesis of Nascent Circular RNAs. *Cell Rep.* *15*, 611–624.

7 APPENDIX

7.1 Acknowledgements

The first and biggest acknowledgment goes out to Prof. Dr Elmar Wolf for putting faith in me and giving a chance to work in his research group. The unending patience, insight and mentoring provided by him shall be difficult to replace in my future. Big thanks also go out to Prof. Dr Martin Eilers whose experience, support and encouragement in all the aspects of the doctoral work was instrumental in making this piece of work what it is now. For all the helpful advice on experiments and thesis work, I would like to thank Prof Dr Alexander Buchberger for having the patience to guide a newcomer in the field. I would also like to thank Prof. Dr Reinhard Lührmann for his support and commitment as a thesis committee member.

All member of the Wolf lab, especially Julia, Jessica, Bikash, Ashwin, Nevenka, Markus and André, have been a source of never-ending support during endless paper submissions, revisions and during lab-work in general, to which I shall remain grateful.

The senior members of every lab are a constant source of wisdom, but I also learned a lot from the students whom I had the pleasure of supervising: Pranjali and Beril. Many thanks go out to both of you for supporting the challenging projects which couldn't have been completed without your optimism and hard work.

All member of the Eilers lab, especially Carsten, Steffi, Gabriele, Theresa and Giacomo, have been the go-to people for all my queries, scientific and otherwise. I think the patience they showed while answering some of my amazingly weird questions, probably deserves a separate chapter, and I am thankful to all of them. Many thanks again to Giacomo, Dimitrios and Gabriele for taking time out of their lives and offering corrections on this thesis.

For all the work presented here, I've taken the help of many scientists outside my department. This was instrumental in getting to know new technologies and scientific concepts. For this, I am grateful to members of the Patrick Cramer lab and the Markus Sauer lab, especially Lisa Jung, Patrick Eiring, Marvin Jungblut and Sören Döse.

Special thanks to Dirk Rieger, Christian Stigloher, Marcus Behringer (Biocenter Imaging Facility), Wolfgang Feichtinger (Sequencing Facility) and to Peter Gallant, for his ever-omniscient expert advice.

Appendix

Life as a doctoral student is filled with many unforgettable moments, and I appreciate the effort put in sharing those moments by the members of “*Kinonacht*”, “*Weinkenner*”, “*WhiskyAmigos*”, “*Weinwanderung*”, “*Feirabendbier*” and “*ein anderes, bitte*” groups! The team spirit shown here will remain incomparable no matter where I go in the future. Special acknowledgments to Theresa, Irem and Gabriele for introducing me to Silvaner. I am sure someday I will make friends with it.

Finally, and majorly, this endeavour wouldn't have been the same without the patience and affection of my partner in life, Anastasia – I promise to fix my definition of “almost done”. Many thanks go out to my family for their support in all my efforts.

7.2 Publication List

Narain A., Bhandare P., Adhikari B., Erhard F.*, **Baluapuri A.***, Wolf E*., Targeted protein degradation reveals a direct role of SPT6 in POL2 elongation and termination, *under revision* (*corresponding author)

Hartmann O., Reissland M., Maier C.R., Fischer T., Prieto-Garcia C., **Baluapuri A.**, Schwarz J., Schmitz W, Garrido-Rodríguez M., Pahor N., Davies C.C., Orian A., Wolf E., Schulze A., Calzado M.A., Rosenfeldt M.T., Diefenbacher M., Implementation of CRISPR/Cas9 genome editing to generate murine lung cancer models that depict the mutational landscape of the human disease., **Front. Cell & Dev. Biol.**, 2021, 9, 201

Endres T., Solvie D., Heidelberger J., Andrioletti V., **Baluapuri A.**, Ade C., Muhar M., Eilers M., Vos S., Cramer P., Zuber J., Beli P., Popov N., Wolf E., Gallant P., Eilers M. Ubiquitylation of MYC couples transcription elongation with double-strand break repair at active promoters, **Mol Cell**, 2021, 81(4), 830-844.E13

Adhikari B., Bojilovic J., Diebold M., Schwarz J.D., Hofstetter J., Schröder M., Wanior M., Narain A., Vogt M., Dudvarski Stankovic N., **Baluapuri A.**, Schönemann L, Eing L, Bhandare P, Kuster B, Schlosser A, Heinzlmeir S, Sottriffer C, Knapp S, Wolf E. PROTAC-mediated degradation reveals a non-catalytic function of AURORA-A kinase., **Nature Chem Bio**, 2020, 16(11), 1179-1188

Jessen C, Kreß J.K.C., **Baluapuri A.**, Hufnagel A., Schmitz W., Kneitz S., Roth S., Marquardt A., Appenzeller S., Ade C.P., Glutsch V., Wobser M., Friedmann-Angeli J.P., Mosteo L., Goding C.R., Schilling B., Geissinger E., Wolf E., Meierjohann S. The transcription factor NRF2 enhances melanoma malignancy by blocking differentiation and inducing COX2 expression., **Oncogene**, 2020, 39(44), 6841-6855

Rajeeve K., Vollmuth N., Janaki-Raman S., Wulff T., **Baluapuri A.**, Dejure F., Huber C., Fink J., Schmalhofer M., Schmitz W., Sivadasan R., Wolf E., Eisenreich W., Schulze A., Seibel J., Rudel T. Reprogramming of host glutamine metabolism during Chlamydia trachomatis infection and its key role in peptidoglycan synthesis, **Nat Microbiology**, *in press*

Baluapuri A., Wolf E., Eilers M., Target gene-independent functions on MYC oncoproteins, **Nature Rev. Mol. Cell Bio.**, 2020, 1-13

Cossa G., Roeschert I., Prinz F., **Baluapuri A.**, Vidal R.S., Schülein-Völk C., Chang YC., Ade C.P., Mastrobuoni G., Girard C., Wortmann L., Walz S., Lührmann R., Kempa S., Kuster B., Wolf E., Mumberg D., Eilers M., Localized Inhibition of Protein Phosphatase 1 by NUA1 Promotes Spliceosome Activity and Reveals a MYC-Sensitive Feedback Control of Transcription., **Molecular Cell**, 2020, 1322-1339.e11

Baluapuri A., Hofstetter J., Stankovic N.D., Endres T., Bhandare P., Vos S.M., Adhikari B., Schwarz J.D., Narain A., Vogt M., Wang S-W., Düster R., Jung L.A., Vanselow J., Wiegerin A., Geyer M., Maric H., Gallant P., Walz S., Schlosser A., Cramer P., Eilers M., Wolf E., MYC recruits SPT5 to RNA polymerase II to promote processive transcription elongation, **Molecular Cell**, 2019, 16(74), 674-687

Herold S., Kalb J., Büchel G., Ade C.P., **Baluapuri A.**, Xu j., Koster J., Solvie D., Carstensen A., Klotz C., Rodewald S., Schülein-Volk C., Dobbstein M., Wolf E., Molenaar J., Versteeg R., Walz S., Eilers M., Recruitment of BRCA1 limits MYCN-driven accumulation of stalled RNA Polymerase, **Nature**, 567 (7749), 545-549

Lorenzin F, Benary U, **Baluapuri A**, Walz S, Jung LA, von Eyß B, Kisker C, Wolf J, Eilers M, Wolf E, Different promoter affinities account for specificity in MYC-dependent gene regulation, **eLife** 2016;5: e15_16

Pattshull G., Walz S., Schwab M., Rühl E., Wolf E., Ade C., **Baluapuri A.**, Kneitz S., Rosenwald A., von Eyss B., Gaubatz S., The Myb-MuvB complex is required for YAP-dependent transcription of mitotic genes, **Cell Reports**, 27 (12), 3533-3546. e7

Schmidt S., Gay D., Uthe F.W., Denk S., Paauwe M., Matthes N., Diefenbacher M.E., Bryson S., Warrander F.C., Erhard F., Ade C.P, **Baluapuri A.**, Walz S., Jackstadt R, Ford C., Vlachogiannis G., Valeri N., Otto C., Schülein-Völk C., Maurus K., Schmitz W., Knight J.R.P, Wolf E., Strathdee D., Schulze A., Germer C., Rosenwald A., Sansom O.J., Eilers M., Wiegering A. A MYC–GCN2–eIF2 α negative feedback loop limits protein synthesis to prevent MYC-dependent apoptosis in colorectal cancer. **Nature Cell Biology**, Vol 21, 1413–1424

Büchel G, Carstensen C, Walz S, Mak K, Rodewald S, Roeschert I, Eoin L, Sumara O, Herold S, **Baluapuri A**, Poon E, Kwok C, Chesler L, Rickman DS, Dobbstein M, Bayliss R, Eilers M, Association with Aurora-A controls N-MYC-dependent pause release of RNA polymerase II during the cell cycle, **Cell Reports**, 2017, 21, 3483–3497

Gerlach JM, Furrer M, Gallant M, Birkel D, **Baluapuri A**, Wolf E, Gallant P, The PAF1 complex component Leo1 recruits Drosophila Myc to promoters, **PNAS**, 2017 114 (44) E9224-E9232

Briese M, Saal L, Appenzeller S, Moradi M, **Baluapuri A**, Sendtner M., Whole transcriptome profiling reveals the RNA content of motor axons, **Nucleic Acids Res.** 2015, 44(4)

7.3 Curriculum vitae

APOORVA BALUAPURI

Personal Data

Born on 7.10.1986 in Bhopal, India.

Contact Data

Department of Biochemistry and Molecular Biology

Biocenter, University of Würzburg

Am Hubland

97074 Würzburg

☎ +49-931-3181821

✉ apoorva.baluapuri@biozentrum.uni-wuerzburg.de

Training and Studies

since 01. 2016	Cancer Systems Biology Lab , <i>University of Würzburg</i> <i>Thesis Title – Molecular Mechanisms of MYC's impact on Transcription</i>
10. 2014 – 10. 2016	M.Sc (FOKUS Life Sciences) , <i>University of Würzburg</i> Final Grade: 1.3
07. 2006 – 08. 2010	B.Tech (Bachelor of Technology) , <i>Biotechnology</i> <i>Vellore Institute of Technology, Vellore, India</i> Final Grade: 1.3
05. 2004 – 05. 2006	BA (Indian Classical Music) <i>Pracheen Kala Kendra, Chandigarh, India</i>

Working Experience

08. 2010 – 08. 2014	System Engineer , <i>Tata Consultancy Services LTD.</i> , <i>Chennai, India</i>
---------------------	---

Awards and Stipends

01. 2016 – 12. 2018	German Excellence Initiative Fellowship , Graduate School of Life Sciences, <i>University of Würzburg</i>
12.2017 – 12.2019	Add-on Fellowship , <i>Joachim Hertz Stiftung, Hamburg,</i> <i>Germany</i>
07.2019	Biocenter Award for Scientific Excellence , <i>University of</i> <i>Würzburg</i>
10.2019	Best Scientific Talk Award , EUREKA-2019, Graduate School of Life Sciences, <i>University of Würzburg</i>
10.2020	Outstanding Tutoring and Mentorship Award , Faculty of Biology & Service Center for Innovations in Learning and Studies (ZiLS), <i>University of Würzburg</i>

7.4 Affidavit

I hereby confirm that my thesis entitled "*Molecular Mechanisms of MYC's impact on Transcription Elongation*" is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

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

Place, Date

Signature

Appendix