From Department of Oncology-Pathology
Karolinska Institutet, Stockholm, Sweden

# THE ODYSSEY OF THE *MYC* TRANSCRIPT FROM THE NUCLEUS TO THE CYTOPLASM: THE MOLECULAR MECHANISM OF THE GENE GATING IN HUMAN CANCER CELLS

Mirco Martino


Karolinska Institutet

Stockholm 2023

All previously published papers were reproduced with permission from the publisher.
Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2023

ISBN 978-91-8017-023-9
Cover illustration: Tempera coloring with glazes, painted by Silvana Vasarri. The DNA is a fundamental molecule of the human body that contains all the information needed for the organism to function and live. The background is the artist free interpretation of the RNA-FISH picture in Fig. 2 of Paper III.

# The odyssey of *MYC* transcript from the nucleus to the cytoplasm: the molecular mechanism of the gene gating in human cancer cells

Thesis for Doctoral Degree (Ph.D.)

By

**Mirco Martino**

The thesis will be defended in public at J3:12 Nanna Svartz U220032500, Bioclinicum, NKS U2, Solnavägen 30, 171 64, Solna, Sweden, Friday, June 16, 2023 at 10:00

**Principal Supervisor:**
Professor Anita Göndör
University of Oslo
Department of Clinical Molecular Biology
and
Karolinska Institutet
Department of Oncology-Pathology

**Co-supervisors:**
Assistant Professor Barbara Scholz
Karolinska Intitutet
Department of Oncology-Pathology

Professor Galina Selivanova
Karolinska Institutet
Department of Microbiology, Tumor and Cell Biology

Associate Professor Andreas Lennartsson
Karolinska Institutet
Department of Biosciences and Nutrition

**Mentor:**
Associate Professor Vicente Pelechano Garcia
Karolinska Institutet
Department of Microbiology, Tumor and Cell Biology

**Opponent:**
Professor Piergiorgio Percipalle
New York University Abu Dhabi
Division of Science
Faculty of Biology
and
Stockholm University
Department of Molecular Biosciences, The Wenner-Gren Institute

**Examination Board:**
Associate Professor Claudio Cantù
Linköping University
Department of Biomedical and Clinical Sciences
Division of Medicine and Virology

Professor Maria Eriksson
Karolinska Institutet
Department of Biosciences and Nutrition

Associate Professor Marianne Farnebo
Karolinska Institutet
Department of Cell and Molecular Biology
and
Department of Biosciences and Nutrition

**Chairperson**:
Docent Ourania Kostopoulou
Karolinska Institutet
Department of Oncology-Pathology

To my entire family, the constellation that lights up my life

Alla mia intera famiglia, la costellazione che illumina la mia vita

*“The man who is blind to the beauties of nature has missed half the pleasure of life”*

Lord Robert Baden-Powell, the founder of the worldwide Scout Movement

*“L'uomo che è cieco alle bellezze della natura, ha perduto metà del piacere di vivere”*

Lord Robert Baden-Powell, fondatore del movimento mondiale dello Scoutismo

# Popular science summary of the thesis – English version

The genetic material of our body, the DNA, that contains all information needed for an organism to function and live, is contained inside the nucleus of cells. The human body is made of an average of more than 30 trillion cells, each of them containing DNA molecules of approximately 2m in length, if stretched end-to-end. If we could link altogether and stretch the DNA molecules of a single human body, it would make more than 150.000 round trips to the moon. The 2m DNA of a single cell is stored in a nucleus with an average diameter of 10μm (0,00001m). How is that possible?

To accomplish such a spectacular achievement, nature has evolved a quite simple system by which the DNA molecules are first rolled up around nucleosomes, a complex of proteins called histones. The arrangement of such nucleosome-DNA structures can agglomerate to each other forming complicated structures that densely compact the DNA within the nucleus. The most compacted forms of DNA molecules are the famous 23-chromosome couples visible in human cells during the mitotic process when the cells duplicate themselves and their genetic materials.

During the other phases of the cell cycle, which together are much longer than the mitotic phase, the DNA is structured in a less condensed form that can be divided into 2 main categories: a more compacted form called heterochromatin and a less compacted one called euchromatin. Heterochromatin contains generally genetic material not needed at that precise time to the cell, and for that reason, the information contained in it doesn't need to be read and interpreted. Conversely, euchromatin needs to be available continuously for reading, interpreting, and being transformed into protein or other information molecules needed for executing cellular functions.

For many years, it was thought that these 2 forms of DNA were positioned within the nucleus of the cells in a random structure, like a plate of spaghetti, and that the structure itself didn't have any specific function. Recent studies and observations have overturned this idea, and now it is clear that not only are the DNA structures very dynamic, highly regulated, and specific, but also that the deregulation of the three-dimensional architecture of the DNA can be the cause and consequence of many types of diseases, including cancer.

With the work in this thesis, we have uncovered a new system that the cells use to regulate the 3D structure of the DNA, a system that seems to be specific for at least one type of cancer cells.

My colleagues and I have observed that 2 pieces of a DNA, one that produces an important regulatory protein, MYC, that is very often overexpressed in cancer, and one that is a regulatory element, that doesn't produce any protein but is important to regulate the activity of the *MYC* gene, called the Oncogenic Super-enhancer (OSE), communicate with each other to control the expression of this protein. More precisely, when the *MYC* gene and the OSE are both located near the periphery of the nucleus and in close proximity to each other, the OSE piece of DNA, thanks to the help of other regulatory proteins that form a scaffold, can recruit a protein that is a component of the nuclear pores. The nuclear pore structures form channels that allow the information produced by the DNA, molecules called mRNAs (messenger RNAs), to travel out from the nucleus to become translated into proteins. In this way, the OSE can actively bring the *MYC* gene directly to the nuclear pores where its mRNA can quickly move out from the nucleus.

As these *MYC* mRNAs are degraded much faster inside the nucleus than outside of it, the novel fast nuclear export mechanism, which is called gene gating, enables the mRNAs produced by *MYC* near nuclear pores to escape the higher degradation rate in the nucleus and thereby increases total *MYC* mRNA and MYC protein levels, which, in turn, contributes to the maintenance of the cancer cell state.

In this thesis, I will describe the molecular mechanisms that regulate the gating of the *MYC* gene and the players involved in it, as well as discuss how such discovery will benefit our understanding of cancer development, progression and the design of new treatments for cancer.

# Sintesi della tesi ad uso divulgatorio – versione italiana

Il materiale genetico ereditabile del nostro corpo, il DNA, che contiene tutte le informazioni necessarie ad un organismo per poter funzionare e vivere, è contenuto all'intero dei nuclei delle cellule. Il corpo umano è costituito da più di 30 trilioni di cellule, ognuna delle quali contiene molecole di DNA di circa 2m in lunghezza, se allungate da cima a fondo. Se potessimo unire e allungare le molecole di DNA di un singolo corpo umano, tale molecola sarebbe lunga più di 150.000 volte la distanza andata e ritorno per la luna. I 2m di DNA di una singola cellula sono immagazzinati in un nucleo di un diametro di circa 10μm (0,00001m). Come è possibile tutto ciò?

Per compiere tale spettacolare impresa, la natura ha evoluto un semplice meccanismo per il quale il DNA è inizialmente arrotolato intorno ai nucleosomi, un complesso di proteine chiamate istoni. Tali strutture nucleosoma-DNA possono agglomerarsi tra loro formando complesse strutture che compattano densamente il DNA all'interno del nucleo. Le forme più compatte delle molecole di DNA sono le famose 23 coppie cromosomiche visibile in una cellula umana durante il processo mitotico quando la cellula duplica sé stessa e il proprio materiale genetico.

Durante le altre fasi del ciclo cellulare, le quali insieme sono molto più lunghe della fase mitotica, il DNA è strutturato in una forma meno condensata e può essere suddiviso in due principali categorie: una forma più compatta chiamata eterocromatina ed una meno compatta chiamata eucromatina. L'eterocromatina contiene generalmente materiale genetico non necessario alla cellula in quel preciso momento, e per questo motivo, l'informazione in esso contenuto non richiede di essere letta ed interpretata. Contrariamente, l'eucromatina necessita di essere continuamente disponibile per essere letta, interpretata e trasformata in proteine or altre molecole informative necessarie per compiere le funzioni cellulari.

Per molti anni, si è pensato che queste 2 forme di DNA fossero posizionate all'interno del nucleo delle cellule in una struttura totalmente casuale, come in un piatto di spaghetti, e che la struttura stessa non avesse alcuna specifica funzione. Recenti studi e osservazioni hanno rovesciato questa idea ed è ora evidente che, non solo le strutture del DNA sono molto dinamiche, altamente regolate e specifiche, ma anche che la deregolazione della architettura tridimensionale del DNA è la causa e conseguenza di molti tipi di malattie, compreso il cancro.

Con il lavoro di questa tesi, abbiamo scoperto un nuovo meccanismo che le cellule usano per, in qualche modo, regolare la struttura tridimensionale del DNA, un meccanismo che sembra essere specifico perlomeno di un tipo di cellule tumorali.

Io e i miei colleghi abbiamo osservato che due parti del DNA, una che produce una importante proteina regolatoria, MYC, che è spesso sovraespressa nei tumori, e un'altra che è un elemento regolatorio il quale non produce alcuna proteina ma è importante per regolare l'attività del gene *MYC*, chiamata il super-enhancer oncogenico (OSE), comunicano tra di loro per controllare l'espressione di questa proteina. Più precisamente, quando il gene *MYC* e il OSE sono entrambi localizzati vicino alla periferia del nucleo e in stretta prossimità tra di loro, la parte del DNA OSE, grazie all'aiuto di altre proteine regolatorie che formano un'impalcatura, può reclutare una proteina che è un componente dei pori della membrana nucleare. I pori della membrana nucleare formano canali che permettono all'informazione prodotta dal DNA, molecole chiamate mRNA (RNA messaggeri), di viaggiare al di fuori del nucleo per poter essere tradotta in proteine. In questa maniera, il OSE può attivamente portare il gene *MYC* direttamente sui pori nucleari dove il suo mRNA può velocemente uscire fuori dal nucleo.

Siccome questi mRNA di *MYC* sono degradati molto più velocemente all'interno del nucleo che al suo difuori, questo nuovo meccanismo di esportazione nucleare, il quale è chiamato gene gating (trasporto del gene), permette all'mRNA prodotto da *MYC* vicino ai pori nucleari di fuggire dalla più rapida degradazione all'interno del nucleo e quindi di aumentare i livelli totali dei mRNA di *MYC* e della proteina MYC, i quali, a loro volta, contribuiscono al mantenimento dello stato tumorale delle cellule.

In questa tesi, descriverò i meccanismi molecolari che regolano il gating del gene di *MYC* e i vari elementi coinvolti in esso e, discuterò anche come tale scoperta possa beneficiare la nostra comprensione circa lo sviluppo e la progressione tumorale e la progettazione di nuovi trattamenti per il cancro.

# Abstract

Pathological expression of the *MYC* oncogene is a common denominator in a wide range of cancers and is linked with abnormal cell proliferation. To achieve this status, the *MYC* gene benefits from being embedded in a region rich in enhancers and super-enhancers that are often absent in the normal cell counterparts. How those regions regulate *MYC* transcription and expression is, however, not well understood, although likely players include enhancer-binding factors, the 3D nuclear architecture and local eRNAs and ncRNAs.

In this thesis, two new models governing MYC expression have been identified. The first describes a posttranscriptional mechanism that is based on the gene gating concept proposed in 1985, while the second is based on the ability of the non-coding eRNA, *CCAT1* to promote *MYC* transcription, which paradoxically antagonizes the gating of *MYC*.

In **Paper I**, a model of gene gating mechanism of *MYC* in human cancer cells is proposed. Briefly, the Nucleopore Complex (NPC) member ELYS (or AHCTF1) recruits *MYC* and its distal Oncogenic Super Enhancer (OSE) to the nuclear pore in a β-catenin -dependent manner. This principle increases *MYC* expression post-transcriptionally by facilitating the nuclear export of its derived mRNAs and thus enabling the escape of *MYC* transcripts from the faster degradation rate in the nucleus compared to the cytoplasm.

In **Paper II**, a CTCF binding site within the non-coding gene, *CCAT1*, positioned within the OSE, was mutated using CRISPR technique. Expanded clones carrying the mutated CTCF binding site revealed that this site is essential for the canonical WNT-mediated gating of *MYC*. Normally ascribed an insulator function, this non-canonical feature of CTCF was essential for the recruitment of ELYS/AHCTF1 to the OSE, thereby effectuating its anchoring to the nuclear pore. In addition, this report shows that CTCF is essential for the WNT-mediated activation of the *CCAT1* gene.

In **Paper III**, the role of the OSE transcript *CCAT1* in the gating mechanism was further analyzed. siRNA-mediated knockdown of *CCAT1* eRNA expression revealed its dual function. While it promotes *MYC* transcription in the nuclear interior, it impedes the nuclear export of its derived mRNAs. We speculate that the *CCAT1* eRNA likely indirectly alleviates transcriptional pausing of *MYC* transcription. Conversely, transcriptional pausing is proposed to promote the migration of the *MYC* gene to the nuclear pores to provide a key switch in the nuclear export pathways of *MYC* mRNAs.

In summary, this work has identified two new models of *MYC* expression regulation in cancer cells, thereby providing opportunities for designing new pharmaceutical strategies targeting pathological expression of this central oncogene during cancer evolution.

# List of scientific papers

I. Barbara A Scholz*, Noriyuki Sumida*, Carolina de Lima, Ilyas Chachoua, **Mirco Martino**, Ilias Tzelepis, Andrej Nikoshkov, Honglei Zhao, Rashid Mehmood, Emmanouil G Sifakis, Deeksha Bhartiya, Anita Göndör, Rolf Ohlsson

WNT signaling and AHCTF1 promote oncogenic MYC expression through super-enhancer-mediated gene gating

*Nature Genetics* 51, 1723–1731 (2019). *Equal Contribution

II. Ilyas Chachoua*, Ilias Tzelepis*, Hao Dai*, Jia Pei Lim*, Anna Lewandowska-Ronnegren*, Felipe Beccaria Casagrande, Shuangyang Wu, Johanna Vestlund, Carolina Diettrich Mallet de Lima, Deeksha Bhartiya, Barbara A. Scholz, **Mirco Martino**, Rashid Mehmood, Anita Göndör

Canonical WNT signaling-dependent gating of MYC requires a noncanonical CTCF function at a distal binding site

*Nature Communications* 13, 204 (2022). *Equal Contribution

III. **Mirco Martino**, Felipe Casagrande, Jia Pei Lim, Ilyas Chachoua, Ilias Tzelepis, Anita Göndör

The long non-coding eRNA, CCAT1, regulates nuclear export pathways of MYC mRNA in colon cancer cells

*Manuscript*.

# Contents

# List of abbreviations

| | |
|---|---|
| 3C | Chromosome Conformation Capture |
| 3D | Three-Dimensional |
| 3D-SIM | Three-Dimensional Structured Illumination Microscopy |
| A/T | Adenine/Thymine |
| ABCs | ATP-Binding Cassettes |
| AHCTF1 | At-Hook Containing Transcription Factor 1 |
| AKT | Ak Strain Transforming |
| ALYREF | Aly/Ref Exoport Factor |
| APC | Adenomatous Polyposis Coli |
| ATAC-seq | Assay For Transposase-Accessible Chromatin Using Sequencing |
| ATPase | Adenosine 5'-Triphosphatase |
| b-TrCP | Beta-Transducin Repeat Containing E3 Ubiquitin Protein Ligase |
| BAC | Bacterail Artificial Chromosome |
| BAF | Brg1/Brm-Associated Factor |
| BRD4 | Bromodomain-Containing Protein 4 |
| Ca2+ | Calcium |
| CAP | Catabolite Activator Protein |
| Cas9 | CRISPR Associated Protein 9 |
| CBC | Cap-Binding Complex |
| CBP | Cyclic Adenosine Monophosphate Response Element Binding Protein (CREB) Binding Protein |
| Cby | Protein Chibby Homolog 1 |
| CCAT1 | Colon Cancer Associated Transcript 1 |
| CCAT1-L | Colon Cancer Associated Transcript 1 - Long |
| CCAT1-S | Colon Cancer Associated Transcript 1 - Short |
| CCAT2 | Colon Cancer Associated Transcript 2 |
| CDK9 | Cycline-Dedpedent Kinase 9 |

| | |
|---|---|
| ChIP | Chromatin Immunoprecipitation |
| ChIP-seq | Chromatin Immunoprecipitation - Sequencing |
| Chr8q24 | Band 24 Of The Long Arm Of Chromosome 8 |
| ChrISP | Chromatin In Situ Proximity |
| CK1 | Casein Kinase 1 |
| Co-IP | Co-Immunoprecipitation |
| CPB | Camp Response Element-Binding Protein |
| CREB | Camp Response Element |
| CRISPR | Clustered Regularly Interspaced Short Palindromic Repeats |
| Crm1 | Chromosomal Maintenance 1 |
| CTCF | Ccctc-Binding Factor |
| CTCFBS | Ctcf Binding Site |
| CTD | Carboxyl-Terminal Tail Domain |
| CTNNB | Catenin Beta-1 |
| CYTB | Cytochrome B |
| DamID | DNA Adenine Methyltransferase Identification |
| DAPI | 4′,6-Diamidino-2-Phenylindole, |
| DC | Destruction Complex |
| DNA | Deoxyribonucleic Acid |
| DNase-seq | Dnase I Hypersensitive Sites Sequencing |
| DSIF | 5,6-Dichloro-1-β-D-Ribofuranosylbenzimidazole (DRB) Sensitivity Inducing Factor |
| E-box | Enhancer Box |
| EGFR | Epidermal Growth Factor Receptor |
| ELYS | Embryonic Large Molecule Derived From Yolk Sac |
| ERK1/2 | Extracellular Signal-Regulated Kinases 1/2 |
| eRNA | Enhancer-RNA |
| EU | 5-Ethynyl Uridine |
| FG | Phenylalanine And Glycine Repeats |

| | |
|---|---|
| FISH | Fluorescent In-Situ Hybridization |
| GC | Guanine Cytosine |
| GCN5 | General Control Of Amino Acid Synthesis Protein 5 |
| GSK3 | Glycogen Synthase Kinase 3 |
| GTPase | Guanosine 5'-Triphosphatase |
| H3K27 | Lysine 27 Of Histone 3 |
| H3K27ac | Acetylated Lysine 27 Of Histone 3 |
| H3K27me3 | Three-Methylated Lysine 27 Of Histone 3 |
| H3K4me1 | Mono-Methylated Lysine 4 Of Histone 3 |
| H3K4me2 | Di-Methylated Lysine 4 Of Histone 3 |
| H3K4me3 | Tri-Methylated Lysine 4 Of Histone 3 |
| H3K4me3 | Three-Methylated Lysine 4 Of Histone 3 |
| H3K9me2 | Di-Methylated Lysine 9 Of Histone 3 |
| H3K9me3 | Three-Methylated Lysine 9 Of Histone 3 |
| HCECs | Human Colonic Epithelial Cells |
| HCT116 | Human Colorectal Carcinoma Cell Line |
| HER2 | Human Colonic Epithelial Cells |
| Hi-C | High-Throughput Cromosome Conformation Capture |
| HP1a | Heterochromatin Protein 1 Alfa |
| ICAT | Inhibitor Of B-Catenin And TCF-4 |
| IDR | Intrinsically Disordered Domain |
| Igf2 | Insulin-Like Growth Factor 2 |
| IgG | Immunoglobulin G |
| IGH | Immunoglobulin Heavy-Chain Locus |
| ISPLA | In Situ Proximity Ligation Assay |
| kb | Kilobase |
| KODs | H3k9me2-Only Domains |
| LAD | Lamin Associated Domain |
| LAS X | Leica Application Software X |
| LCR | Locus Control Region |

| | |
|---|---|
| LDB1 | Lim Domain Binding |
| LEF | Lymphoid Enhancer-Binding Factor 1 |
| Lem2 | Nuclear Lamina-Associated Inner Nuclear Membrane Protein |
| LINE1 | Long Interspersed Nuclear Elementm1 |
| LLPS | Liquid-Liquid Phase Separation |
| LMDB1 | LIM Domain-Binding Protein 1 |
| LMNA | Lamin A/C |
| LMNB1 | Lamin B1 |
| LMNB2 | Lamin B2 |
| lncRNAs | Lon Non-Coding RNA |
| LOCK | Large Organized Chromatin K9-Modification |
| LRP | Low Density Lipoprotein Receptor-Related Protein 1 |
| MAX | Myc-Associated Factor X |
| MED12 | Mediator Complex Subunit 12) |
| MEK | Mitogen-Activated Protein Kinase Kinase |
| miRNA | Microrna |
| MIZ1 | Myc-Interacting Zinc Finger Protein 1 |
| mRNA | Messanger RNA |
| mRPC | Messanger-Ribonucleoprotein Complex |
| MTREX | Mtr4 Exosome Rna Helicase |
| MYC | V-Myc Myelocytomatosis Viral Oncogene Homolog |
| MYCL | V-Myc Avian Myelocytomatosis Viral Oncogene Homolog 1, Lung Carcinoma Derived |
| MYCN | Neuroblastoma-Derived V-Myc Avian Myelocytomatosis Viral Related Oncogene |
| ncRNA | Non-Coding Rna |
| NEAT1 | Nuclear Paraspeckle Assembly Transcript 1 |
| NELF | Negative Elongation Factor |
| NF-kB | Nuclear Factor Kappa-Light-Chain-Enhancer Of Activated B Cells |

| | |
|---|---|
| NONO | Non-POU Domain-Containing Octamer-Binding Protein |
| NPC | Nuclear Pore Complex |
| Nup | Nucleoporine |
| NXF1 | Nuclear Transcription Factor, X-Box Binding 1 |
| NXT1 | Nuclear Transport Factor 2 Like Export Factor 1 |
| OSE | Oncogenic Super-Enhancer |
| p-TEFb | Positive Transcription Elongation Factor B |
| PARP1 | Polyadp-Ribose Polymerase I |
| PAXT | Polya RNA Exosome Targeting |
| PCAT1 | Prostate Cancer Associated Transcript 1 |
| PCP | Planar Cell Polarity |
| PI3K | Phosphoinositide 3-Kinases |
| PIN1 | Peptidylprolyl Cis/Trans Isomerase |
| PKC | Protein Kinase C, |
| Pol II | Rna-Polymerase Ii |
| Pom | Nuclear Envelope Pore Membrane Protein |
| PRC | Polycomb Repressive Complex |
| PRC2 | Polycomb Repressive Complex 2 |
| qRT-PCR | Quantitative Real-Time Polymerase Chain Reaction |
| Ran | Ras-Related Nuclear Protein |
| RBP | RNA-Binding Protein |
| RNA | Ribonucleic Acid |
| RNF14 | Ring Finger Protein 14 |
| ROS | Reactive Oxygen Species |
| rRNA | Ribosomal RNA |
| Rsc | Remodeling The Structure Of Chromatin |
| scaRNA | Small Cajal Body RNA |
| SE | Super-Enhancer |
| SFPQ | Splicing Factor Proline And Glutamine Rich |
| SINE | Short Interspersed Nuclear Element |

| | |
|---|---|
| siRNA | Short Interfering RNA |
| SKIV2L2 | Superkiller Viralicidic Activity 2-Like 2 |
| SLAM-seq | Thiol(SH)-Linked Alkylation For The Metabolic Sequencing Of RNA |
| SMC | Structural Maintenance Of Chromosomes |
| snoRNA | Small Nucleolar RNA |
| SNP | Single-Nucleotide Polymorfism |
| snRNA | Small Nuclear RNA |
| SOX2 | Sry-Box Transcription Factor 2 |
| spCas9 | Streptococcus Pyogenes Cas9 |
| SPRY4 | Sprouty Rtk Signaling Antagonist 4 |
| SPT6 | Suppressor Of Ty 6 Homolog |
| SRSF1 | Serine/Arginine-Rich Splicing Factor 1 |
| SRSF2 | Serine/Arginine-Rich Splicing Factor 2 |
| SRSF3 | Serine/Arginine-Rich Splicing Factor 3 |
| SRSF7 | Serine/Arginine-Rich Splicing Factor 7 |
| SUN | Sad1p, UNC-84 |
| SUV39H1 | Suppressor Of Variegation 3-9 Homolog 1 |
| SWI/SNF | Switch/Sucrose Non-Fermentable |
| TAD | Topological Associated Domain |
| TALEN | Transcription Activator-Like Effector Nucleases |
| TBP | TATA-Binding Protein |
| TCF/LEF | T Cell Factor/Lymphoid Enhancer Factor Family |
| TEFb | Positive Transcription Elongation Factor B |
| TF | Transcription Factor |
| TFII | Transcription Factor Ii |
| TP63 | Tumor Protein P63 |
| TREX-2 | Three Prime Repair Exonuclease 2 |
| tRNA | Transfer RNA |
| WDR5 | Wd Repeat Domain 5 |

| | |
|---|---|
| WNT | Wingless-Type Mmtv Integration Site Family, Member 1 |
| WRE | Wint Response Element |
| XRN2 | 5'-3' Exoribonucleasi 2 |
| YY1 | Yin Yang 1 |
| ZFC3H1 | Zinc Finger C3h1-Type Containing |
| Zn-fingers | Zinc Finger |

# Table of Figures

# Introduction

The work of this thesis describes evidence of a new molecular mechanism governing the nuclear export of *MYC* transcripts into the cytoplasm.

Producing these data during these years of my doctoral education has been not always easy and the journey that brought me to this moment has been divided into several stages. When I had to think about a title that could summarize my results, the first word that came to my mind was "odyssey".

Perhaps my subconscious thinking of my trip to reach this milestone, in combination with my southern-Italian heritage, has played an important role in choosing this word, but after a deep reflection, I thought "odyssey" would perfectly describe the gene gating mechanism of *MYC* transcripts.

To begin this journey that discovered the secrets of gene gating that regulates the *MYC* expression, I will start describing the world in which this odyssey takes place.

The sea in which *MYC* sails is complicated and full of dangers. It's a very crowded and dynamic world, with many kingdoms (nuclear periphery and nuclear pores, transcription condensates, speckles, etc.), that will be described below in more detail, and powerful enchantresses, first among all CTCF that, like Circe, is able to change the shapes of what it touches and rule on the 3D nuclear architecture.

After that, I will talk about the different stages of the transcription journey, from its initiation and elongation, with all the regulatory mechanisms behind them, to the termination and the defense mechanism against the enemies that want to degrade the transcript before it is able to reach the final destination.

Two of the most dangerous monsters in this world, probably more dangerous for us readers who want to find a meaning in the story, than for *MYC* itself, are plasticity and stochasticity, the Scylla and Charybdis of the nucleus that would do their best to change the route of the trip and wave the sea.

Finally, I will talk about the main character of the story, that is of course *MYC*, one of the most studied oncogenes, upregulated in a vast type of cancers and, that, as for the Homeric Ulysses, is still surrounded by an aura of mystery. As I will describe it later, what exactly is the function of MYC oncoprotein in cancer cells is not yet well-understood. Whereas MYC binds almost all promoters, the effect on the transcription of its target genes is rather moderate, even though MYC remains an essential player in the oncogenesis process. Another common feature between Ulysses and MYC is

astuteness: MYC is able to resist many drug treatments by regulating the cellular drug-export mechanisms that will get rid of them.

Essentials for the success of the journey are a good boat and nice fellows: the oncogenic super-enhancer upstream of the *MYC* locus, active only in cancer cells, is a perfect scaffold for the transcript to travel with, and the eRNA *CCAT1*, a very friendly mate involved in the upregulation of *MYC* and in many cancerogenic processes, is essential to accomplish this adventure.

As the help of some Gods was crucial for Ulysses to navigate, in the same way, the WNT signaling pathway, from the above, overlooks many aspects of the life of a cancer cell and helps *MYC* in its gating.

I hope that after having read the description of the world and of the characters of this odyssey, you would also be excited in knowing more details of this great adventure, as I was, when I observed them for the first time, and as I'm still now while trying to describe this story.

# 1 Literature review

## 1.1 The crowded world of the nucleus: the 3D nuclear architecture

In the highest taxonomy rank, all organisms are divided into 3 domains: Bacteria, Archaea and Eukaryota. Humans, together with all animals, fungi, plants, and other unicellular organisms like amoeboids belong to the Eukaryote domain. The common denominator between them is the possession of a nucleus (from ancient Greek, eukaryote, "good kernel"), in which the genetic material is protected and separated from the other cellular compartments and functions.

It is generally thought that the nucleus is similar among all eukaryotes since the Last Eukaryotic Common Ancestor, but recent studies of nuclear membranes in different types of organisms have shown that the nucleus, as we know it today, has not only appeared recently in the evolution[1] but also that the nuclear membranes have evolved together with the nuclear pore complexes (NPCs)[1–3].

The nucleus is a quite crowded environment: it contains all the genetic material structured in a complex 3D architecture (Figure 1A) and all the enzymes and molecules needed for essential nuclear functions like DNA duplication, repair and RNA transcription and maturation. Containing all these elements in a relatively small space is what probably makes the nucleus such a special organelle in which unique features like compartmentalization and liquid–liquid phase separation have evolved to allow more complicated but also energetically more sustainable functions to exist[4].

### 1.1.1 Nuclear Compartmentalization

If we color cell nuclei with a simple DAPI (4',6-diamidino-2-phenylindole) staining, that binds adenine-thymine -rich regions of the DNA, it is possible to distinguish areas with different fluorophore absorption that seem to form compartments within the nucleus.

Whereas DAPI staining is mainly used to visualize nucleoli, more advanced microscopy techniques have enabled the detection of many more compartments (Figure 1A): speckles and paraspeckles, splicing hubs, Cajal bodies, histone locus bodies, transcription factories, etc. Each of these structures, which do not have a surrounding membrane, has specific functions and is formed by specific components.

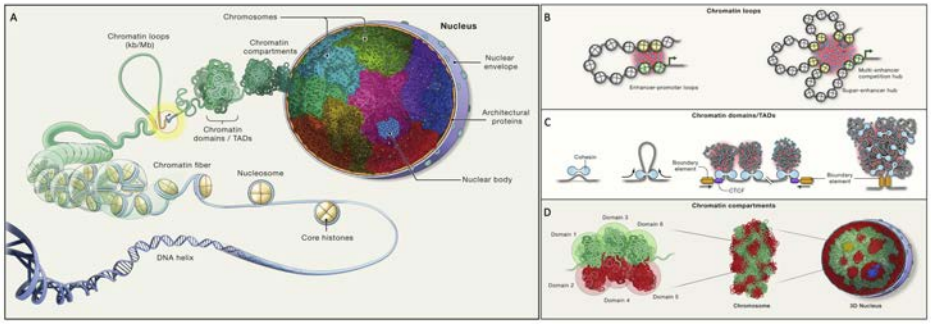


**Figure 1: The crowded nuclear environment**

(A) *The nucleus is mainly occupied by DNA, which can reach 2m in length in every single human cell. To be able to be contained in the much smaller nuclear space, the DNA is organized in a strictly regulated architecture. (B) Firstly, the DNA is rolled around histone proteins to form nucleosomes, which agglomerate to shape chromatin fibers. Within the same fiber, DNA can form loops to connect regulatory elements, like enhancers (in orange in A), to gene promoters (in blue in A) and regulate transcriptional activity. (C) Fibers, in turn, fold to form chromatin domains, or TADs that reduce the probability of interactions between specific elements. (D) TADs of similar structure assemble in chromatin compartments: A-type compartments are formed mainly by less compact euchromatin, and B-type compartments consist of dense heterochromatin. Furthermore, each individual DNA fiber forms a chromosome, visible during mitosis, which during interphase occupies a specific space within the nuclear volume, forming the so-called chromosome territory that extensively intermingles with other chromosome territories. Nuclear envelope and architectural proteins, like CTCF, are necessary for a correct 3D nuclear organization. In the DNA-free space, RNA molecules and proteins aggregate in nuclear bodies or compartments.*

*Modified from The Self-Organizing Genome: Principles of Genome Architecture and Function, Cell 2020 18328-4, https://doi.org/10.1016/j.cell.2020.09.014*

One of the most diffuse hypotheses explaining nuclear compartmentalization is based on the physical concept of phase separation and condensation: normally a molecule A and a molecule B would randomly mix with each other to maximize the entropy of the system, but if A has a higher affinity to similar other A molecules, the non-covalent interactions between A molecules will be energetically favorable. If those interactions happen at multiple independent sites of A, the avidity (the sum of the strength of multiple non-covalent molecular interactions) of such molecules will increase, allowing, when a critical concentration threshold is passed, the self-interacting A molecules to form a separated phase from B molecules, a condensate. This process is called concentration-dependent- phase separation[4]. From a biological perspective, these types of multivalent weak interactions can happen between the intrinsically disordered regions (IDRs) of proteins that allow them to interact simultaneously with other proteins, DNA, or RNA molecules[5].

To have a biological functional condensate, the molecules within the separated phase have to be able to dynamically form and destruct weak interactions as in a liquid-type of state, a phenomenon called liquid-liquid phase separation (LLPS)[4–7]. Those condensates will have distinct properties and physical states (liquid, solid or gel-like) than their surrounding[4–7].

In their review, Sabari et al.[5] claim that biological condensates have 3 main features: compartmentalization, the ability to separate proteins involved in a specific process from other activities that happen in the surroundings, selective partitioning, by which the condensates are able to select molecules only essential for its function and "expel" others even if they were produced within them, and concentration, that is the production of higher concentration of molecules involved in the condensate's processes.

It is important to highlight how in the field the terminology is still not properly uniform, and the terms condensates, compartments, phase separation, and liquid-liquid phase separation are often interchanged and mixed. More precisely, phase separation is a specific thermodynamic behavior that results in the formation of condensate, the LLPS is a specific subtype of phase separation that forms liquid-type of condensates which in turn are one, but not the only, speculated formation mechanism of nuclear compartments[4].

One of the first studies to observe the LLPS nature of a nuclear compartment, the P granules of *Caenorhabditis elegans*, was done by Brangwynne et al. [8] that show the dynamics of formation and disruption of those granules and their ability to undergo fission, division, and fusion.

An important component for phase separation and condensation to happen in a cell is the presence of a confined environment, such as, e.g., the nucleus, or the presence of "sticky surfaces" that can recruit molecules in specific territories, such as the nucleoli or the nuclear periphery with Lamin proteins, nuclear pore complexes, SUN- domain proteins, and non-coding RNAs acting as "Velcro"[4,9,10] that can bind diffusible molecules and be the seeding point of nuclear compartment formation.

The advantages of forming compartments within the nucleus are many: they allow for the concentration of enzymes or transcription factors that would otherwise be difficult to accumulate just by diffusion in a specific space; they mediate contacts between chromatin fibers, e.g., between enhancers and promoters that would be otherwise linearly far away from each other; they facilitate co-transcription of specific regions; they permit to couple different processes like transcription and splicing, and they seem to be very important in helping transcription factors to find their targets[4,5]. This was brilliantly shown by the work of Brodsky et al.[11], which describes how IDRs of transcription factors are necessary and sufficient to localize the majority of the promoter targets. Transcription factors (TFs) thus often bind common DNA motifs with high-affinity, but, nonetheless, they occupy specific sets of promoters. By only affinity and diffusion it would take several hours for a transcription factor to find the target. To solve this problem, Brodsky and co-authors[11] suggest that first the IDRs, of which TF are enriched

in, place the TF in a specific DNA region, compartment, surrounding the target. Once there, the TF DNA Binding Domain can then recognize with high-affinity the DNA motive, stabilizing in this way the binding[11].

Apart from the above-mentioned principles, the formation of several very specialized nuclear compartments is supported by the chromatin that works as a scaffold, folding the genome in a dynamic and specific 3D architecture[12]. It is important to highlight how all these structures are really dynamic, and they cannot be thought of as static stable organizations.

### 1.1.2 3D Genome organization

#### *1.1.2.1 The formation of A and B compartments*

The physical organization of the genome within the nucleus has been examined with 3D-DNA-FISH (Fluorescent *In-Situ* Hybridization), a microscopy-based method that allows targeting genomic regions with fluorescent oligo probe, and Hi-C variants that enable the identification of all-to-all cross-linked chromatin fiber contacts in large and small cell populations or single cells following chromatin digestion, ligation and sequencing steps.

3D-DNA-FISH has visualized that interphase chromosomes occupy specific volumes within the nucleus called chromosome territories[13] (Figure 1A) (also reviewed in Lemaître and Bickmore, 2015[14]) that extensively intermingle with each other at their boundaries[15].

Hi-C assays have confirmed the existence of chromosome territories by documenting high interaction frequencies between regions in *cis*, i.e. within the same chromosome, compared to *trans* interactions bridging distinct chromatin fibers on different chromosomes.

Importantly, Hi-C has uncovered that chromatin, at megabase scale, is divided in 2 compartments: A compartment, which is composed of gene-rich, high accessible regions with active transcriptional histone marks, and B compartment, which consists of compact, less accessible chromatin regions with repressive transcription marks. Regions belonging to the same type of compartment are more frequently in contact with each other and tend to compartmentalize[16] (Figure 1A and 1D).

Innovative approaches have thus elucidated the role of β-actin in regulating the communication and switches between these 2 compartments: β-actin knockdown thus increases general chromatin decompaction by facilitating a decrease in H3K9me3 and an increase in H3K4me3 levels and, as consequence, deregulated gene expression[17]. β-actin levels don't influence only local chromatin arrangements but also reversibly change nuclear 3D architecture by affecting the DNA-binding of the chromatin remodelers BAF/BRG1 that generally oppose the function of the Polycomb Repressive

Complex (PRC). PRC, together with heterochromatin protein 1-alpha (HP1α) is, in turn, involved in initiating the phase-separation processes to establish A and B compartments[17,18]. Moreover, switching from B to A compartment seems to be driven more by the transcriptional activity and gain in chromatin accessibility at long-range regulatory elements, probably guided by β-actin-mediated H3K27ac, than by a change in chromatin interactions[19].

Observing Hi-C data at higher resolution within A and B compartments, Dixon et al.[20] have observed that the genome is organized in what they call Topological Associated Domains (TADs) (Figure 1A and 1C): large, megabase-sized, evolutionally and, to a great extent, cell-type conserved local chromatin domains displaying 2-4 fold higher interaction frequency within the domain than between domains, which provide a framework for enhancer-promoter interactions (Figure 1A and 1B). The boundaries of such domains are enriched in chromatin architectural proteins like CTCF (Figure 1C) and the cohesin complex, tRNAs, housekeeping genes, and SINE (short interspersed nuclear element) retrotransposons, which have an essential role in their formation and preservation. TAD are divided into sub-TADs that tend to limit the activity of an enhancer or super-enhancer to a specific genomic region[21]. Although the formation of TADs is not well understood, the role of architectural proteins, such as CTCF and cohesin is well-established[22].

#### *1.1.2.2 CTCF: the "Circe enchantress" of the nuclear architecture*

CTCF, the CCCTC-binding factor, is considering the master regulator of chromatin architecture and genome organization[23]. It is an evolutionary conserved transcription factor that harbors 11 zinc fingers, of which zinc fingers 4-7 are essential for the distribution of CTCF within the genome and bind the core-DNA binding motif[22,24,25] in a methylation-sensitive manner[26].

Initially, it was discovered as a transcriptional regulator of the chicken c-myc[27], and later as an insulator of chicken β-globin gene where it is placed between the gene and its enhancer element to block enhancer-promoter interaction[28]. Since then, it has become evident that CTCF can thus work as transcriptional activator, repressor and insulator by binding directly to gene promoters and enhancer or insulator elements[22]. More recently, CTCF was discovered to act as chromatin architectural protein and establish short- and long-range chromatin fibre interactions between various regulatory elements, demarcate TAD boundaries and isolate enhancer-promoter interactions within sub-TAD compartments[20,29] (Figure 1C).

How CTCF carries out its regulatory roles is not clear[22]. For the insulation function, it is thought that CTCF can bend the DNA element it binds to and organizes a specific 3D chromatin configuration that prevents the access of enhancer elements to the target promoters, if placed in between. In the case of the *Igf2/H19* imprinting control region,

CTCF bound to the maternal *Igf2/H19* imprinting control region thus blocks the access of upstream enhancer elements to the *Igf2* gene and simultaneously maintains DMR1, a silencer element targeting *Igf2*, in an active state to repress *Igf2* expression[30]. Moreover, the various functions of CTCF can be further boosted by the combinatorial use of its Zn-fingers to bind DNA as well as its RNA and protein-interactors in a site-specific manner[23]. CTCF has thus been shown to be able to both promote histone deacetylation[31] and the recruitment of Pol II and other transcription factors to facilitate the formation of transcription condensates[32,33].

Its role in the formation of TAD boundaries is less well-understood[22,33,34]. Bioinformatic analyses have thus revealed that the formation of the loop base is facilitated by 2 CTCF binding sites in a convergent motive orientation with respect to the interior of the loop[35] (Figure 1C). The reason for such specific orientation, that would be difficult to explain if the anchor sequences found each other only by diffusion mechanisms, can be better understood in the context of the loop extrusion model[36]. This model is used to explain features of genome architecture, including TAD formation and enhancer-promoter interactions[36]. According to the model, cohesin, a DNA binding multisubunit ATPase belonging to the family of structural maintenance of chromosomes (SMC) protein complexes, would initially form a small loop entrapping the DNA sequences *in cis* inside a ring structure formed by its 3 subunits[36]. The loop is proposed to grow until the extruding cohesin complex interacts with 2 CTCF molecules bound at the convergent CTCF binding sites, brought together *in cis* by this loop extrusion mechanism[36]. This model is able to explain why CTCF and cohesin are both found on loop and TAD boundaries and why cohesin accumulates at CTCF sites[36].

Surprisingly, acute removal of CTCF effects only marginally the mean levels of transcriptional activity, despite the alteration of the 3D nuclear architecture[37]. Although the transcription of some genes seems to be more affected than others by CTCF depletion[38], this observation raised the question that looping might not always be essential for enhancer-promoter interactions and CTCF might only modulate their existence[22]. In contrast to this observation, mutation of a CTCF binding site at a TAD boundary was sufficient to cause developmental defects [39], despite that the lack of CTCF binding only slightly altered the contact frequencies of enhancer-promoter contacts over TAD boundaries[39]. Moreover, reduced levels of CTCF have been shown to alter the cell-to-cell variability of gene expression[40]. In addition, the effect of CTCF on TAD-boundary strength might be locus-specific. Gong et al.[41] have thus shown that multiple CTCF sites are enriched at TAD boundaries with higher boundary strength, and that super-enhancers are, in general, insulated by such strong boundaries. Finally, altered transcription and disrupted 3D genome organization are typical features of cancers[22]. Some of those alterations have been linked to mutations in the CTCF gene or in some of the CTCF binding sites[22].

Post-translation modifications and interactions are also known to affect CTCF functions: its Poly(ADP-ribosyl)ation by PARP1 is, for example, required for its insulator function at the *H19* imprinting control region[42], whereas its direct interaction with PARP1 is needed for the positioning of circadian genes to the nuclear lamin[43]. In addition, noncoding RNAs can strengthen or disrupt CTCF binding to chromatin, and lncRNAs are thought to fine-tune the DNA binding ability of CTCF at specific genomic loci[22].

Although experimental evidence for the role of chromatin fibre interactions between distant regulatory elements and TAD organization is well established, it is less clear how such events on the primary chromatin fibre relate to the compartmentalized architecture of the 3D nucleus[33].

### 1.1.3 Nuclear periphery

The periphery of the nucleus is the region in proximity to the nuclear envelope. The envelope, is not only a delimitation structure that protects the nucleus from cytoplasmic activities, but it also plays a crucial role in many nuclear activities, from gene regulation to RNA transport and nuclear export, nuclear architecture and DNA duplication. The nuclear membrane is constituted of 2 individual lipid bilayers, the outer nuclear membrane, which is in continuation with the cytoplasmic endoplasmic reticulum, and the inner nuclear membrane, which contains several transmembrane proteins. The 2 layers are separated by a perinuclear space of around 50nm in human cells[44], and connected by nuclear pore complexes that work as gatekeepers of the nucleus. The inner nuclear membrane is connected with an intricate set of intermediate protein filaments called lamins. In mammals, there are 4 types of lamin proteins: lamin A and lamin C (A-type lamins), produced by alternative splicing of the *LMNA* gene, and lamin B1 and lamin B2 (B-type lamins) which are the products of the *LMNB1* and *LMNB2* genes, respectively. Alternative splicing of *LMNA* and *LMNB1* produces lamin C2 and lamin B3, which are germ-cell-specific isoforms. While A-type lamins are expressed asynchronously only in differentiated cells, B-type lamins are expressed early during embryonic development and they continue to be expressed in mature adult cells (already reviewed in 1988 by Larry Gerace and Brian Burke[45]).

Many diseases have been linked to malformation or mutation in the lamins, as reviewed by Muchir and Worman[46], including the premature aging syndrome called Hutchinson-Gilford progeria syndrome[47].

#### *1.1.3.1 Lamina Associated Domains*

The lamina is essential for shaping the 3D architecture of chromatin: it works as a scaffold to dynamically attach specific regions of DNA to form the so-called Lamina Associated Domains or LADs. Since the first microscopy observation of chromatin, it was obvious that in interphase, the genome is divided into a more dense and compact

structure called, heterochromatin, and into a more open architecture called euchromatin. Already with basic electron microscopy techniques in the early 70', it was observed that heterochromatin is mainly localized at the periphery of the nucleus, in contact with the lamin[48] (also reviewed by Lemaître and Bicknore[14]). This observation has been confirmed with many more sophisticated and modern techniques, including three-dimensional structured illumination microscopy (3D-SIM), that have been able to add more details to the pictures: the LADs are not uniform all around the nuclear envelope, but the area around the NPCs is generally chromatin-poor[49]. The high density of heterochromatin is the consequence of highly compact DNA domains formed by the deposition of histones carrying transcriptional silencing marks, such as di- or tri-methylated lysine 9 or lysine 27 of histone 3 (H3K9me2, H3K9me3, H3K27me3)[50]. Euchromatin is instead more open and accessible to transcription factors thanks to histones that are enriched with active transcription marks like tri-methylated lysine 4 of histone 3 (H3K4me3) or the high acetylated tails of histone 3 and 4[51]. Those are just some of the most well-established of the large amounts of modifications and their combinations histone tails can carry. The various modification combinations of histones form a much more complicated and yet not well-understood language, which is dynamically deposited, erased, read, and interpreted by many players within the nucleus, a language generally known as histones code[52].

In the last decades, many methods have been developed to study the specific genome sequences that form hetero- or euchromatin in a cell type-, differentiation stage- and environmental signaling-dependent manner. The Chromatin Immunoprecipitation-sequencing (ChIP-seq) assay, for example, enables the immunoprecipitation of an antibody against, in this context, a histone modification and the chromatin around it that would be sequenced[53], DNase-seq, in which the DNase I hypersensitive sites, nucleosome-free chromatin regions that are subjected of DNase I cleavage, would be able to be sequenced[54]; and ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) where hyperactive Tn5 Transposases is used to tag with adaptor for next-generation-sequencing open chromatin region[55]. Another method used to study more specifically the LADs is the DNA adenine methyltransferase identification (DamID), in which a protein of interest, for example, LaminB, is fused with a DNA-methyltransferase that methylates in position 6 any adenine on the DNA region in which the target protein will bind. Since the methyl6adenine is not a natural DNA modification in eukaryotes, the chromatin regions in which the target protein will bind could be easily detected by methylation-sensitive restriction enzyme digestion, sequencing and bioinformatic analysis[56].

Since LADS are mainly occupied by heterochromatin, they also largely overlap with the so-called LOCKs, Large Organized Chromatin K9-modifications[57], regions of the chromatin enriched by H3K9me2. Wen et al.[57] have thus shown that H3K9me2 LOCKs

represent 30% of the entire genome of mammalian cells, are evolutionally conserved, acquired during development in a tissue-specific manner to coordinate cell type-specific gene repression, and are perturbed in cancer cells. LOKCs seems also to play an important role in the reconstruction of LADs after cell division acting as "*a 3D architectural mitotic guidepost*[58]".

Later studies have shown that LADs could be divided into 2 categories: constitutive LADs that are A/T rich-regions, gene-poor, more evolutionally conserved, and work as the "skeleton" of the nuclear architecture[59]; and facultative LADs that are more cell type-specific and richer in genes, which act by constraining in a transcriptional silent space developmental genes that have to be inactivated in differentiated cells[60,61].

A recent study has uncovered LOCKs with active marks, which are prominent during early differentiation and can also be used to distinguish undifferentiated cells from differentiated ones, and LOCKs with silencing marks other than H3K9me2, such as H3K9me3 and H3K27me3[62]. An interesting sub-group of LOCKs are the bivalent LOCKs prominently present in undifferentiated cells. Bivalent LOCKs carry both active, H3K4me1 or H3K4me3, and repressive, H3K27me3, marks similarly to bivalent genes marked by both H3K4me3 and H3K27me3, protecting the underlying DNA promoters from DNA methylation during long-term silencing[63]. The bivalent LOCKs tend to overlap with Topologically Associated Domains (TADs)[62], are bound by chromatin architectural proteins such as the transcription factor CTCF[62], and in differentiated cells, they are replaced by only the silencing H3K9me3 mark[62].

LADs form a dynamic environment, which can have functions beyond repressing gene-poor domains and developmentally regulated chromatin regions. An interesting contribution of LADs to the expressivity of the genome in adult cells is the formation of a temporary silencing platform for specific genes in certain circumstances: Zhao et al.[43] have observed, for example, that several circadian clock-regulated genes move, in a transcriptionally active state, from the center of the nucleus to the periphery where they acquire silencing marks and become temporally transcriptional inactive before being released back to the nuclear interior in a rhythmic manner, under the control of CTCF and the DNA repair protein PARP1. Interestingly, repressed circadian genes are released to the nuclear interior while being marked by H3K9me2, despite that in other cellular contexts the acquisition of H3K9me2 and H3k9me3 marks have been sufficient to reposition a locus to the nuclear periphery and to delay replication timing[64–67].

A recent study has also observed a role for lamin A/C in splicing[68]. Tammer et al. have thus shown how the outcome of the splicing is influenced by the genome architecture: at the nuclear periphery where genes tend to have a lower GC- and higher AT-content, errors in alternative splicing might result in exon skipping[68]. Whereas in the nuclear

interior, which is enriched in GC-rich genes, splicing errors are linked to intron retention[68].

Another very interesting recent discovery by Caballero et al.[69] is that the inner nuclear membrane protein Lem2 coordinates, in yeast, the RNA degradation of meiotic transcripts and non-coding RNAs by interacting with and targeting exosome subunits to the nuclear periphery.

These examples clearly show not only the dynamics of the nuclear periphery but also the importance of the 3D nuclear architecture, which is essential in all aspects of gene duplication and regulation.

#### *1.1.3.1 Nuclear Pores*

The nucleus is connected to the cytoplasm by one of the most complicated protein macromolecular complexes in the cell called Nuclear Pore Complexes (NPCs). NPCs consist of more than 1000 subunits and have a mass of around 110MDa[70]. They localize at around 800Å wide-pores formed by the fusion of the outer and inner nuclear membranes[70]. One of the features that distinguish NPCs from other cellular channels is the ability to protect the nucleus from free diffusion of large molecules and at the same time allow transport of cargo in their native folded state[70]. The passive diffuse barrier is created by the phenylalanine and glycine-enriched (FG) IDRs repeats. This allows molecules lower than 40kDa to passively diffuse across it, while bigger cargos need facilitated transport, achieved by nuclear transport factors, to efficiently cross the barrier[70].

Since their first discoveries in 1959[71] by electron microscopy, the structures of NPCs have been intensively studied by X-ray crystallography, proteomics, and structural biology methods, and recently also by Artificial Intelligence-based structure prediction[72]. Those studies have shown that each NPC is composed of around 30 nucleoporins or Nups (the exact number seems to be cell-type dependent), some of which surround symmetrically the transportation channel, forming a core of three cylinders: one within the nuclear envelope and one each in the outer and inner nuclear membranes. This core works as a scaffold for the asymmetrical binding of the cytoplasmic and nuclear nucleoporins which form, respectively, the cytoplasmic filaments and the nuclear basket[73]. The structure of the nuclear pore and that of each Nups are evolutionarily conserved, despite poor sequence conservation. This feature has brought some authors to suggest that proto-NPCs have been the first form of the nuclear envelope, and that the current nuclear membrane was a later addition[1]. Nup133 and Nup153 are the main Nups responsible for the NPC reassembly after cell division[74,75]: Nup133 seems to be essential for the reconstitution of the NPC basket and to recruit Nup107 and ELYS (or AHCTF1), while Nup153 is responsible to recruit Nup98, Nup62, and Pom121. ELYS seems to be the only nucleoporin with an AT-hook-DNA-binding domain[76].

Like other components of the nuclear periphery, also the NPCs are very dynamic entities involved in many other processes than only being a transport channel of transcripts and molecules between the nucleus and cytoplasm. Actually, many nuclear Nups are mobile and have been found inside the nucleus in contact with specific chromatin regions. Nup98, for example, has been shown to not only bind enhancer regions of *Hox* genes and ecdysone-inducible genes in *Drosophila* but also to be responsible for the stabilization and transcriptional memory of enhancer-promoter loop and their anchoring to the NPC[77]. Other 2 independent studies[78,79] have also observed similar behaviour in different mammal cells: Nup98 and Nup153 have been found connected to super-enhancer regions to the locus control region of β-globin. Their absence is also responsible for a drastic change in the transcriptional activity of SE-regulated targets. Both Nup98 and Nup153 contain many IDRs that could help in forming phase-separated compartments between super-enhancers and their target genes[77,80]. In *Drosophila*, Nup98 seems to interact in the interior of the nucleus with highly acetylated and transcriptionally active chromatin, while a small percentage of Nup98 interactions happen at NPCs, at the nuclear membrane, with short, transcriptionally inactive regions[81]. In addition, Nup98 works as a chromatin architectural protein, similarly to CTCF that also physically interacts with Nups. Indeed, Nup98 is abundant at TAD and sub-TAD domain boundaries, suggesting its involvement in their formation[82,83]. In addition, Nup93 is a component of the inner-ring sub-complex of the NPC and seems to interact with both inactive Polycomb Response Elements[84] (PRCs) and H3K27ac-enriched active chromatin region[79,84].

Other components of the NPCs, like TREX-2, Nup210 and Nup153, are also involved in transcription regulation: TREX-2 binds to and regulates the assembly of the mediator complex[85], and promotes the formation of tissue-specific transcription complexes, whereas Nup210 induces genes needed for differentiation[86,87], and Nup153 promotes the silencing of developmental genes by mediating the recruitment of the PRC1[88]. Whereas ELYS ChIP peaks overlap with that of other Nups, they seem to be more robust than the peaks detected in Nup98, Nup93, and Nup97[77,84]; potentially indicating that the binding of other Nups to chromatin are mediated by ELYS. Of interest, ELYS is able to bind directly to nucleosomes[89] and is responsible to recruit activating chromatin remodeler complexes like PBAP (Polybromo-containing Brahma-associated proteins) when recruited to a genomic locus by another Nup[90]. The extensive interactions of Nups with genomic loci have models proposing that LADs might be interspaced with domains near the NPCs of active developmental or inducible genes[90].

Nucleoporins seem to be also sensitive to environmental stress signals, such as Nup133, which is downregulated in mice by hyperoxia[91].

### 1.1.4 The Nuclear Interior

From the nuclear periphery to the nuclear interior there is an increase in gene-dense chromatin regions[14]. For that reason, the interior of the nucleus is mainly a transcriptionally permissive and dynamic environment, in which a huge amount of transcripts need to be transcribed, processed, and exported. Recent studies have shown that many of those functions happen within specialized nuclear compartments or bodies[4,7,12].

#### *1.1.4.1 Transcriptional condensates*

The concept of transcription factories is quite old and comes from observations by electron microscopy and immunostaining of clusters of stable RNA-Polymerase II (Pol II) in specific regions of the nucleus[92]. A recent *in-vivo* study with super-resolution microscopy has confirmed the existence of such entities, but it also documented their dynamic nature which enables rapid cellular responses to external stimuli[93]. Another essential component of RNA transcription, the Mediator Complex, also forms large dynamic condensates when recruited by transcription factors to large or clustered enhancer elements, such as the so-called super-enhancers[94]. The Pol II and Mediator condensates interact with each other and seem to be involved hierarchically in different phases of the transcription process[94]. The main role in the creation of the transcription factor/mediator condensates is played by the Activation Domains of transcription factors, which are enriched in IDRs[95]. IDRs are essential for the selective occupation of super-enhancer-associated genes, as shown by Zamudio et al.[96] for the canonical WNT-signalling transducer β-catenin. Indeed, β-catenin is recruited on cell-type specific super-enhancer/mediator condensates also in the absence of the domain responsible for the interaction with TCF/LEF family DNA-binding WNT effector TFs, needed to bind DNA[96]. The combination of the condensate-mediated concentration of the signaling factors and the DNA-binding ability is, according to the authors, required for the specificity and the high level of gene activation during signaling cascades, such as WNT[96].

Recent studies have also shown an important role of nuclear actin in the formation of Pol II transcription condensates, especially in response to environmental cues, such as serum deprivation or interferon-γ: under these stimulations, dynamic polymerization and depolymerization of nuclear actin promote long-lasting, larger, and more dynamic Pol II condensates, enabling an increase in the basal initiation transcription rate[97].

#### *1.1.4.2 Splicing condensates, Speckles and Paraspeckles*

The phosphorylation of the Serine 2 of the carboxyl-terminal tail domain (CTD) of Pol II is the event that initiates the elongation phase of the transcription at promoters containing Serine 5 phosphorylated paused Pol II[98]. Recent discoveries have shown that CTD phosphorylation is also responsible for switching the preference of the Pol II from

clusters of Mediator transcriptional condensates towards the splicing condensates, enriched of splicing factors, the IDRs of which, together with the IDRs of the phosphorylated Pol II, drive the condensate formation[99].

It has been observed that highly-transcribed Pol II genes and their pre-mRNA are organized around particular nuclear compartments called speckles (reviewed in Lamond and Spector[100]). Those are nuclear bodies enriched in splicing factors and RNA involved in the splicing process[4,100]. They are generally free of DNA and mainly work as a supplier of splicing factors to the transcription site[100,101]. The inner core is composed of arginine-rich mRNA splicing factors like SRSF1 and SRSF2 and other kinases and phosphatases needed for the splicing, while chromatin and pre-mRNAs are localized at the periphery[4,100]. The distance between transcription condensates and speckles is regulated and follows an "economy of scale" by which the higher transcription rate, the shorter the distance between the 2 bodies. In this way, highly-transcribed genes are also subjected to higher co-transcriptional splicing without the need to modify the overall concentration of splicing factors in the cell[4,102].

Recently, splicing and DNA-repair mechanisms have been linked, and the two processes seem to be interconnected[103]. In that regard, speckles could also be important compartments storing factors for DNA-repair mechanisms[104].

Another compartment that visually resembles speckles but the function of which is still not clear is the paraspeckle[105]. Paraspeckles are condensates formed mainly by 2 proteins, the essential paraspeckle protein SFPQ (Splicing Factor Proline and Glutamine rich), and NONO (Non-POU domain-containing octamer-binding protein), as well as the long-non-coding RNA NEAT1 (Nuclear Paraspeckle Assembly Transcript 1)[105,106]. Paraspeckles, and specifically NONO, have been linked to many diseases and cancers[106,107]. In a recent review, McCluggage and Fox argue that paraspeckles can work as global sensors of cell stress and help to move the cells between different states: from stress condition to homeostatic or between two different differentiation or programming states[106].

#### *1.1.4.3 Other compartments: nucleoli, histone locus bodies, and Cajal bodies*

One of the first nuclear bodies ever observed are the nucleoli, essential compartments responsible for ribosome biogenesis. As reviewed by Lafontaine et al.[108], they contain small nucleolar RNAs (snoRNAs), ribosomal DNA, and protein like RNase needed for the process[108]. It is divided into three liquid-like phases: the center is a fibrillar compartment containing the ribosomal RNA (rRNA) transcription molecules, which is surrounded by the dense fibrillar components where the rRNA is processed, and finally the granular compartment where the ribosomes are assembled[108]. The separation into three distinct phases probably is needed to ensure that each biogenesis phase is completed before each intermediate product is moved to another compartment[4].

Other very specialized nuclear bodies are the histone locus bodies and the Cajal bodies. In the first, the histone mRNA biogenesis occurs, a process quite different from the classical mRNA transcription, since histone mRNAs need snRNA like U7 and regulatory proteins to be processed, and are generally not polyadenylated[4]. An exception are the replication-independent histones variants or histones that work as replacement variants in non-dividing long-lived cells that have been shown to be polyadenylated[109].

Cajal bodies are responsible instead for the maturation of snRNAs, which process requires special modifications like methylation and pseudouridylation. They contain snRNAs, small Cajal body RNAs (scaRNAs), and proteins like coilin[4]. ScaRNA2 is involved also in DNA-repair pathways, giving Cajal bodies a role in this important process[110] and a possible connection to speckles.

The creation of these really specialized nuclear sub-compartments enables cells to concentrate molecules that otherwise would easily diffuse within the nucleus, e.g., ncRNAs, and to increase the rate of identification between regulators and targets[4,5], especially in cases of low-expressed regulators, such as scaRNAs[4].

## 1.2 The stages of the odyssey: from RNA transcription to export

As discussed above, the majority of nuclear functions happen within specialized, dynamic subcompartments or condensates[4,7]. In many of the sub-compartments, RNA molecules are the main products that need to be produced and/or modified, carried out by enzymes and factors recruited to the compartment[4,111]. To be able to perform all its functions, a cell thus needs to read and interpret the information deposited in its genome[112]. That is the main role of the transcription that produces transportable complementary RNA units. If those units will be later translated into proteins, they are called messenger RNAs (mRNAs), otherwise, they belong to the big family of non-coding RNAs (ncRNAs)[113].

Eukaryotes have 3 RNA Polymerases[114]: Pol II is responsible for all mRNAs and the majority of ncRNAs. Pol I is involved in the transcription of large ribosomal RNAs, while Pol III is needed for transfer RNAs (tRNAs) and other small non-coding RNAs[115]. Although the molecular mechanisms of Pol II transcription are well-characterized[116], it still fascinates scientists how specific transcription programs can be executed and distinct chromatin loci be recognized by the transcription machinery in a such crowded and complex environment like the nucleus[117,118].

### 1.2.1 Transcription Initiation

The first phase of the transcription is the initiation, where Pol II and other factors bind the gene promoter forming the preinitiation complex responsible for melting the DNA strands and initiating the synthesis of short transcripts that generally are aborted until

they don't reach the minimum size that allows the transcription complex to enter in the second phase[119].

#### *1.2.1.1 Transcription Factors*

The deposition of the Pol II complex on the promoter is regulated by transcription factors, the main factors responsible for the execution of very specific and highly regulated transcriptional programs[118].

To better understand the function of TFs, is useful to divide them into 4 categories as suggested by Pope and Medzhitov in their review[117]: Class A TFs regulate housekeeping genes, or ubiquitous constitutively expressed genes that are expressed similarly in all cell types and are responsible for essential functions of the cell life[120]. The promoters of these genes are generally nucleosome-free and enriched in unmethylated CG, forming the so-called CpG islands[121]. Class B TFs are responsible to activate or repress the transcription of primary response ubiquitously inducible genes upon specific cellular signals[122]. Stress, inflammation, or specific metabolic conditions will activate a signaling pathway, like NF-κB, Toll-like receptors, and ROS (Reactive Oxygen Species) sensors, which cascades will terminate in the activation of specific TFs belonging to this class and to the activation or repression of their target genes[117,122]. Class C TFs, on the other hand, are not generally expressed in the cells, and their expression and induction are often the results of the activation of the primary response genes by Class B TFs[117,122]. Their targets are secondary response inducible genes.
Subclass C1 TFs induce their targets after being activated by a broad range of signals in most types of cells, e.g., c-MYC, c-Fos, and JunB[117,122]. Their targets are secondary response ubiquitously inducible genes, and their main function is to act as amplifiers of specific transcription programs[123].
Subclass C2 TFs are activated in most cell types, but only by specific signaling pathways[117]. Their targets are a small group of secondary response ubiquitously inducible genes, and their main role is to control the specificity of the signals. Some members of this class are, e.g., E2F family of TFs that are activated by mitogenic signals and control cell-cycle genes[124]. Subclass C3 contains TFs that are expressed under specialized signals unique to a specific type of cell. Their target genes are indeed secondary response cell-type-specific genes, that generally require chromatin remodeling before being able to be activated[125]. Finally, Class D TFs are mainly involved in cell differentiation and cell-type specific expression programs[117]. Their targets are generally cell-type specific constitutive genes, so they are not in need of signals to be activated. They are generally responsible to activate cell-type specific enhancers for class B and class C TFs[117,126]. However, it is important to keep in mind that the categories above are only illustrative, and TF expression, more than simply an on/off mode, should be described as a continuum with peaks if certain conditions are in place[12,117] (Figure 2B, page 27).

How TFs could rapidly find their target is still under debate. One of the hypotheses is termed the facilitated diffusion[127], by which first, TFs identify genomic neighbourhoods at high speed by 3D rapid diffusion. This step is followed by the slowly identification of the high-affinity targets within the neighbourhoods. Whereas the IDR domains are mainly responsible for the first rapid scan and to localize the correct region, the DNA-binding domains are responsible for finding the specific DNA sequence[11,118].

#### *1.2.1.2 Enhancers and Super-enhancers*

TFs play important roles in the regulation of enhancers[126,128], which are by definition "*DNA sequences able to activate gene expression over large genomic distance, independent of sequence orientation*"[129] and regulate cell type- and differentiation stage-specific gene expression[130]. Although the first sequence with this characteristic was found in 1981 in simian virus 40 genome - an element able to increase human β-globin transcription independently of the relative position to the gene promoter[131] – it is still very difficult to provide a precise molecular definition to enhancer elements[129]. Typically, an active enhancer element is nucleosome-free, therefore sensitive to DNase I treatment, and is enriched in TF binding sites, transcriptional co-activators, Mediator complex binding, and in activating chromatin modifications such as H3K27 acetylation[129,132,133]. They thus share many features with promoters, and while many enhancer elements can act as promoters, some promoters can function as enhancers for other genes[134]. Although the differences between enhancers and promoters are not well defined, promoters tend to be enriched in H3K4me3[135]. However, the identification of enhancers based only on chromatin features does not seem to be sufficient, because many of these characteristics have been found also in candidate enhancers that don't stimulate the expression of a reporter gene *in-vivo*[129]. Other common enhancer-binding proteins are the chromatin architectural and looping factors CTCF and the cohesin complex[136], master regulators of the 3D genome organisation[23]. CTCF, with or without the cohesin complex[137], is thus strictly required for a sub-set of CTCF-dependent -enhancer-promoter loops[138,139], and regulates boundary strength at 80% of the TAD domains[138,139].

How enhancer-target promoter interactions are specified during differentiation, is, still an unsolved question. The traditional view that enhancers establish stable contacts with target promoters *via* the Mediator complex that works as a bridge by binding TFs loaded at both extremities of the loop[129] has been challenged by new observations[140]. Experiments based on live-imaging microscopy have shown that, upon activation, the diffusion rate and thus the mobility of both enhancers and promoters drastically increases, suggesting that the observed higher interactions frequencies between the 2 entities are rather due to an increased number of stochastics encounters than the formation of a stable enhancer-promoter complexes[140]. Interestingly, another study has documented, by super-resolution 3D DNA-FISH and chromosome conformation capture techniques, that activation of the Sonic hedgehog pathway correlates with a higher

enhancer-target promoter distance as a consequence of the assembly of a large chromatin-bound protein complex and the catalytic activity of poly (ADP-ribose) polymerase 1 (PARP1)[141]. As the authors of the study describe, PARP1, which is generally known to be involved in DNA-repair mechanisms[142], has recently been linked to transcription regulation[143], the insulator properties of its binding partner CTCF[42], while the product of PARP1, PAR chains, have been seen as possible seeds that promote phase separation by liquid de-mixing[144]. Indeed, Pol II, the Mediator complex and other transcriptional co-activators, and chromatin-remodeling proteins, such as BRD4, are all enriched in IDRs[4]. A current model[136] thus proposes that transcription hubs might be pre-assembled, similarly to condensates, onto enhancers and later loaded onto the target promoter once that is transiently looped to the enhancer *via* structural proteins like LDB1 (LIM domain-binding protein 1) and YY1 (Yin Yang 1). Such a loop is then envisaged to become larger and more dynamic[136]. Such a scenario would also explain how the same enhancer can co-activate genes that are 200nm distant from each other[136,145].

In 2013, a special sub-group of enhancers were discovered in murine embryonic stem cells[146] and human cancer cells[147], termed super-enhancers (SEs). In both cases, super-enhancers have been defined as large cluster of enhancer units highly enriched in master developmental TFs in the first case, and oncogenic TFs together with BRD4 in the second, as well as high levels of Mediator occupancy[146,147]. Common feature of SEs and oncogenic SEs (OSEs) is the ability to robustly increase the transcription rate of a specific set of target genes: development-related, cell fate-specifying genes in case of SEs and oncogenes in case of OSEs[148,149]. Since then, super-enhancers have been observed in many different cell types[149], but a clear structural definition has not been made, because many of their features overlap with those of classical enhancers. It is still under debate if super-enhancers should be considered a separate entity or a subclass of enhancers, and how they are related to other transcription-controlling regions such as stretch enhancers and Locus Control Regions (LCRs), as reviewed by Pott and Lieb[150].

Current literature highlights several common features of super-enhancers (reviewed in[148,150]), which suggests that these large regulatory regions are more than just the sum of their individual enhancer units[150]: apart from their large size that usually extends beyond 10kb, SEs and OSEs are enriched one magnitude more than normal enhancers in cell-type specific or developmental-specific TFs, Mediator Complex, Pol II, chromatin remodeling proteins, like BRD4, and H3K27ac, H3K4me2, and H3K4me1 histone marks and display increased chromatin accessibility[148,150]. OSEs cancer-specific are enriched at oncogenes to ensure robust expression of cancer-relevant genes. Both SEs and OSEs are moreover enriched in disease-associated variants and disease-associated SNPs (single-nucleotide polymorphisms)[149]. Several SEs can overlap with other large-scale regulatory elements like LCRs, and both SEs and OSEs are much stronger activators of

transcription than regular enhancers, an ability that is very sensitive to the down-regulation or inhibition of TFs and BRD4[151]. Finally, SEs and OSEs tend to be located near CTCF binding sites that provide a "porous barrier" to demarcate SE range[152]. Moreover, CTCF binding within SEs is linked with their hierarchical organisation and the emergence of hub enhancer units within SEs[153].

Another characteristic that distinguishes SEs from classical enhancers is that enhancer-RNAs (eRNAs) produced within SEs tend to be similar to lncRNAs in that they are generally long, complete, spliced, and polyadenylated, as opposed to the short, bidirectional, capped eRNAs generated within regular enhancers[154].

#### *1.2.1.3 Enhancer-RNAs*

Although eRNAs mark enhancers and SEs/OSEs, they are discussed in a separate paragraph due to their emerging importance in the regulation of enhancer functions[155]. eRNAs are thus RNA transcripts produced bidirectionally or unidirectionally by Pol II on active, H3K4me1- and H3K27ac-enriched enhancers or SE/OSE regions[156]. The majority of eRNAs produced within regular enhancers are short, unspliced, nuclear, non-polyadenylated, 5-capped, and more sensitive to exosome degradation than normal mRNAs[154] – a feature that led to models arguing that they might be only the noisy consequence of Pol II recruitment[134]. As opposed to regular enhancers, a subset of SEs and OSEs produce long-enough eRNAs to become polyadenylated that can work similarly to lncRNA[157].

Several models have been proposed to provide explanations for the evolution and function of eRNAs. Based on the results of Parakal et al.[158] showing that the lncRNA *Lockd* doesn't have any function in *cis* or in *trans*, Espinosa [159] has thus speculated that lncRNAs evolved from simple eRNAs that, by chance, obtained regulatory *trans* activity that has been advantageous for the cell/organism and therefore selected upon during evolution.

Several lines of evidence support a regulatory role for eRNAs. For example, the expression of eRNAs tends to precede the expression of target genes and their level correlates with that of their targets[160,161]: downregulation or exogenous overexpression of certain eRNAs thus results in inhibition or stronger activation of the target gene[160,162]. How eRNAs perpetrate their action is not yet very clear, but several observations and hypotheses have been made.

An important mechanism involves the ability of certain eRNAs to bind cohesin subunits to thereby facilitate the formation and the stabilization of enhancer-promoter chromatin loop[163]. Similarly, eRNAs have been described to activate the kinase activity of the Mediator complex towards H3 serine 10 phosphorylation and to facilitate the chromatin localisation of Mediator subunits to thereby affect chromosome folding[161].

Moreover, eRNAs can interact directly with histone acetyltransferase CPB/P300 that increases the level of activating H3K27ac around enhancers and promoters[164]. In addition, it has been proposed that similarly to other nascent RNAs, eRNAs might be able to bind the Polycomb Repressive Complex 2 (PRC2), a transcriptionally repressive chromatin remodeling complex, to antagonize its binding to chromatin and thereby reduce the repressive H3K27me3 mark[165].

As we discussed for other nuclear compartments, RNA molecules have a key role as a scaffold for the creation of LLPS condensates[4]. In the same way, eRNA might work to create phase-separated enhancers, as illustrated by Arnold, Wells and Li[166].

Finally, a surprising observation suggests that eRNAs could work as spatiotemporal decoys, regulating the transcription elongation phase by binding of the RNA recognition motif of the E subunit within the NELF (Negative Elongation Factor) complex that is responsible to induce Pol II pausing[167]. NELF thus binds Pol II and nascent transcripts to pause the transcription process until the phosphorylation of the Pol II CTD and of NELF itself by the positive transcription elongation factor p-TEFb - a step needed for the initiation of the elongation phase[98]. eRNA might modulate this process, facilitating the release of NELF by recruiting it on itself, resulting in a largely and more precisely timed response compared to the basic transcription[167].

On the note, also other non-coding RNAs have been shown to play a role in the regulation of transcription initiation: Dueva and collogues, have, for example, observed how single-stranded RNA can promote the opening of chromatin by attenuating the histones' electrostatic interactions *via* a rapid, passive, and sequence-independent mechanism. RNA removal has thus caused histone precipitation *in-vitro*, which has rapidly re-dissolved upon adding RNA to the solution. Further supporting this new principle, an interaction has been verified between histone H2B and LINE1 (Long Interspersed Nuclear Element 1) RNA[168].

Given the role of eRNAs in the function of regular enhancers, SEs and OSEs, it is not surprising that the expression of many eRNAs have been documented to be perturbed in tumors[149,169] and linked with oncogene expression and tumor development[134]. An example is represented by the eRNA transcribed within the oncogenic colorectal super-enhancer regulating *MYC*, termed Colon Cancer Associated Transcript 1 (CCAT1)[162,170]. Recent analyses of, the clinical utility of eRNAs has revealed their importance in predicting prognosis and providing potential novel therapeutic targets[171].

### 1.2.2 Transcription elongation and RNA maturation

A fundamental step in the transcription process is the phosphorylation of the serine 2 of the CTD domain of Pol II by the cyclin-dependent kinase 9 (CDK9) subunit of p-TEFb[98]. This reaction is thus needed to induce structural conformational changes in the enzyme,

which allows the continuation of transcript elongation[98]. During the initiation phase, Pol II is often paused as a consequence of several challenges that need to be addressed before continuing, as reviewed by Gonzalez et al.[172]. These challenges include: obstacles in chromatin conformation, specific DNA sequences like AT-reach regions that form weak DNA-RNA hybrid that stall and promote backtracking of Pol II, correction of transcription errors, presence of negative regulating factors, such as DSIF and NELF, and DNA-damage[172]. Without elongation inputs, Pol II pausing can lead to transcription backtracking, arrest, or termination[173]. The loading of elongation factors like p-TEFb, TFII, and SPT6 is thus needed to induce the transcription elongation phase[98]. Furthermore, Shao and Zeitlinger[174] have shown that Pol II pausing regulates new transcription initiation frequency on the same gene, and they speculate that this mechanism could regulate the bursting frequency (rapid loading of Pol II transcribed protein) of transcription by maintaining open conformation and, at the same time, allowing enhancer-promoter contacts to fine-tune the total transcription rate. In addition, certain TFs can regulate specifically transcription elongation instead of initiation[98]. Loss of c-MYC in murine embryonic stem cells, for example, reduces the elongation ability of Pol II and the total level of phosphorylated-serine 2-Pol II without interfering with its recruitment on promoters or the level of phosphorylated-serine 5-Pol II, probably *via* affecting the activity of p-TEFb, to which it is able to bind[175].

Transcription elongation has been reported to be strictly linked with transcript maturation[98,176]. In order to be functional and exported into the cytoplasm, pre-mRNAs need to receive a 7-methylguanosine cap to their 5′-end, need to be spliced, i.e. their introns need to be removed and the exons ligated, sometimes they also need to undergo editing, by, e.g., deamination of adenosines and cytosines, and their 3′-end has to be cleaved and linked with a poly-adenosines tail. As reviewed by Bentley[176], elongation and mRNA maturation are strictly coupled mechanisms that influence each other: slow elongation, for example, can favour alternative splicing and induce specific RNA folding. On the contrary, slow splicing can slow down elongation and regulate the transcription rate. However, many questions remain without explanation, such as what determines if the splicing happens co-transcriptionally or post-transcriptionally and can the signaling pathways influence such mechanisms?

### 1.2.3 Transcription termination, mRNA degradation, and transport

#### *1.2.3.1 General mechanisms regulating transcription termination, mRNA degradation and nuclear export*

When the Pol II complex reaches the end of a gene, the 3′-end cleavage and polyadenylation complex (CAP) is recruited[177]. The cleaved and polyadenylated transcript will form an R-loop that invades the DNA duplex causing the slowdown of the transcription and the release of the transcript from the transcription complex[177,178]. The

Pol II complex will continue the transcription also after the release of the transcript until close chromatin conformation will signal the release of the complex from the DNA helix[177]. Alternative polyadenylation sequences, transcription errors, or DNA breaks are some of the causes of premature termination, a mechanism that might control transcription in case of cellular stress or virus infection and in cancers, as reviewed by Proudfoot[177].

Short, unspliced, and unmodified transcripts are targets of the co-factors of 3′–5′ exo- and endo-nucleolytic RNA exosome that is responsible for their degradation (reviewed by Schmid and Jensen[179]). One of the first checkpoints where the exosome apparatus intervenes is during transcription pausing or pre-termination near the transcription start site[180]. Here, factors involved in the decapping of RNA - by the recruitment of exosome, and the 5′–3′-exonuclease XRN2 - are responsible to degrade the nascent transcript still connected to the Pol II complex[181]. A common cause of transcription pre-termination is the presence of cryptic polyadenylation sites that cause premature cleavage and polyadenylation[179]. Generally, the effects of those sites are dampened by the U1 snRNAP[182], allowing the transcription to continue. If the transcription is terminated due to low levels or inefficient U1 activity, unspliced and not properly polyadenylated transcripts are produced, which are efficiently degraded by the exosome similarly to the promoter upstream transcripts[183,184]. While the co-transcriptional degradation processes are well characterised, it is less well understood how the immature transcripts that escape the transcription site are degraded[185,186]. Some of them are possibly degraded in the cytoplasm, whereas others are probably decapped and targeted by XRN2[179,185]. Importantly, even fully mature transcripts are subjected to nuclear decay, as shown by observations that the zinc finger ZFC3H1 that recognizes polyA interacts with the exosome co-factor MTR4 (also known as SKIV2L2 or MTREX) to form the polyA-tail exosome targeting (PAXT) connection[186]. As a consequence, nuclear RNA decay targets all RNAs, although there are some mechanisms in place to escape it[179]. An example is provided by lncRNAs that can form compartments or complex secondary structures or bind to RNA-binding proteins (RBPs) or chromatin to get protected from the exosome[179]. Coding transcripts, on the other hand, escape rapid decay by being exported out the nucleus and by forming ribonucleoprotein complexes (mRNPs) with RBPs that will temporarily protect them from decay[187]. Indeed, nuclear export and decay seem to strictly balance each other: MTR4 and ALYREF (Aly/REF export factor), an mRNA nuclear export adaptor, compete to bind the cap-binding complex (CBC) to, respectively, recruit the transcript to the exosome for the degradation or to the NPCs for nuclear export[188]. ALYREF is generally more abundant in the nucleus than MTR4 that is mainly localized in the nucleoli. Moreover, as ALYREF binds efficiently only spliced and mature transcripts, a combination of kinetic competition and binding affinity is what defines the destiny of a transcript[188].

To be efficiently exported into the cytoplasm, mature transcripts have to be "marked" by a series of proteins to form a complete mRNP[187]. As reviewed by De Magistris[187], at the begging of transcription, during the capping, the first export protein to be attached is UAP56 - a member of the Transcription Export Complex (TREX), which binds the CBC. UAP56 is an ATPase helicase that is required for structural rearrangements and for the assembly/disassembly of RNPs[189]. UAP56 also recruits THO and ALY, the other 2 members of the TREX complex[187]. ALY, which seems to be also involved in the splicing process together with other splicing complex factors like SRSF3 and SRSF7, is needed for the recruitment of NXF1/NXT1, the fundamental component of RNPs that is responsible to dock the transcript to the NPC by binding the FG domain of the nucleoporins[187,190]. Whether or not the binding to nucleoporins happens at the nuclear membrane or in the nuclear interior to mobile nucleoporins, is still debated[187,191]. It is, however, clear that a complete NPC with nuclear basket, inner circle, and cytoplasmic filaments is needed for the mRNA nuclear export to happen[187].

Interestingly, snRNAs and rRNAs seem to have a slightly different nuclear export mechanism, similar to the one used by proteins containing nuclear export signals[192]. This mechanism requires the binding of the karyopherin protein Crm1 and the small GTPase Ran[193].

An alternative export mechanism, which inspired the work of this thesis, was proposed by Blobel in 1985 and defined as the gene gating principle[194].

#### *1.2.3.2 The Gene Gating principle*

In his opinion paper about the gene gating hypothesis[194], Günter Blobel starts from the assumption that all cells have identical 3D chromatin structures, similarly to metaphase chromosomes[194], and that those structures are destroyed and remade in cells undergoing cell-cycle progression and differentiation[194]. He then proposes that fixed 3D positions of DNA fibers within the nucleus would provide the signal enabling the reassembly of those 3D structures when needed[194]. More precisely, protein-coding active genes would be positioned close to specific NPCs in a way that all cells at the same cell cycle and differentiated state would have a specific gene or set of genes in the proximity of an NPC with fixed 3D coordinates. Apart from this structural perspective, Blobel also speculates that all transcripts of a gene or set of genes might be linked to specific NPCs for the nuclear export of their products, and thus they need to be actively gated to those NPCs to couple transcription and nuclear export, creating structural and functional asymmetry in eukaryotic cells[194].

Although Blobel`s idea, which has been quite futuristic for that time, collides with our current knowledge about the highly dynamic nature of the nuclear environment, it can be still considered partially valid according to recent discoveries[77,195,196]: in *Saccharomyces cerevisiae*, and to some extent also in *Caenorhabditis elegans*, it has

observed that certain inducible genes that regulate for example stress response or are under developmental control need to be relocated to NPCs to be transcriptional activated (reviewed by Burns and Wente[196]). In *Drosophila*, Nup98 is frequently found on promoters and enhancer elements, which seems to be involved in transiently stabilizing the loop between them[77]. Before the results presented in this thesis, in mammals including humans nucleoporins have mainly been linked to architectural functions, transcriptional activation or repression, and export mechanism without evidence for an active gene gating mechanism to NPCs[197]. However, a gating-similar mechanism facilitating the rhythmic recruitment of circadian genes to LADs for gene repression has been proposed[43].

Considering the size and the complexity of mammalian nuclei compared to that of other lower organisms[198], gene gating mechanisms could thus have evolved with different functions or different mechanisms that make their observation more elusive.

## 1.3 Plasticity and Stochasticity: Scylla and Charybdis of the Nucleus

In a such complex environment as the nucleus, in which essential and highly regulated cellular activities take place, such as cell type- and differentiation state-specific gene transcription, one could expect an extremely ordered and stable environment, similar, if not identical, in all cells of the same type[194]. However, dynamicity is a well-described phenomenon[7] in the nucleus where compartments are constantly created and disassembled and interactions formed and released[136].

Adaptation to environmental change thus requires dynamic transcription patterns and underlying dynamic nuclear processes enabling the cell to respond to signaling, metabolic and mechanical pathways responsible to communicate environmental changes to the nucleus[12,199,200]. While inducible TFs are generally among the first responders to signaling cascades[117], the integration and modulation of the signals over time is facilitated by mechanisms evolved to write, erase and read heritable and reversible chromatin and DNA modifications[200]. In this way, chromatin itself has been proposed to be considered as a digital-to-analogic converter that transforms on and off signals into continuous waves, the amplitude and intensity of which could be adjusted over time, as postulated by Badeaux and Shi[201] (Figure 2B).

With this strategy, evolution has thus selected upon a system in which the same DNA sequence could give rise to alternative gene activity states, which mechanisms are nowadays defined as epigenetics[200]. This term was introduced for the first time by Waddington[202], who also proposed a landscape model for epigenetic mechanisms during development and introduced the concept of developmental plasticity: cells with higher differentiation potential would have the possibility to choose from several developmental paths to follow, but once the decision is made towards a path, i.e. differentiation towards a specific mature cell type, they would no longer be able to go

back, similarly to a ball rolling down on a slope decorated with well-defined canals as a consequence of gravity[203,204]. The robustness of the process against perturbations is termed canalisation, which refers to the epigenetic force that doesn't allow a differentiated cell, which contains the same DNA sequence as the progenitor, to return back to the previous developmental stage or trans-differentiate into other cell types. The model has been during the years modified to include recent observations on the ability of heritable and reversible chromatin states and 3D nuclear architectures[205] to balance phenotypic plasticity and canalization, as well as discoveries like inducible pluripotent cells[206]. A more modern definition of the concept of plasticity includes also the ability of the cells to respond to external stimuli by changing their state, as opposed to robustness - an ability to be able to keep the current state despite internal or external perturbations[12].

Current observations on 3D nuclear structures and activity, including single-cell studies, have documented the highly plastic and dynamic nature of this environment (Figure 2): TADs and chromatin domains slightly differ between cells of the same type[207], alleles of the same genes can have different transcriptional states[208], and transcription on the same site works generally in burst instead of a continues mode[209,210] (Figure 2B). This heterogeneity and plasticity within the same cell or between a population of cells of the same-type will contribute to a faster response to external stimuli compared to cells locked into permanent states[12,211].

Heterogeneity is also linked with another fundamental principle of the epigenome: stochasticity[12,212,213], resulting from the probabilistic nature of chromatin compartmentalization, transcription, or any other nuclear activity, while regulation mechanisms are in place to increase or reduce such probability[12,212] (Figure 2C).

Incorporating stochastic principles in chromatin-based processes and nuclear functions makes it easier to understand the relationship between genome structure and genome function (Figure 2A): genome function alone is thus not enough to shape genome structure and genome structure itself cannot totally modulate genome function, but certain functions increased the probability to form specific structures and certain structures increase or decrease the probability of certain functions to take place[12]. At the same time, the features observed in a cell population have to be considered the ones with the highest probability to happen in each cell, rather than unique and common to all the cells[213](Figure 2C).

Although plasticity and stochasticity are fascinating principles, they make the trip to understand the molecular mechanism of gene transcription and regulation very complicated and hostile. Special approaches and precautions, like single-cell analysis or deep-sequencing, need thus to be applied when embarking in this adventure[214].

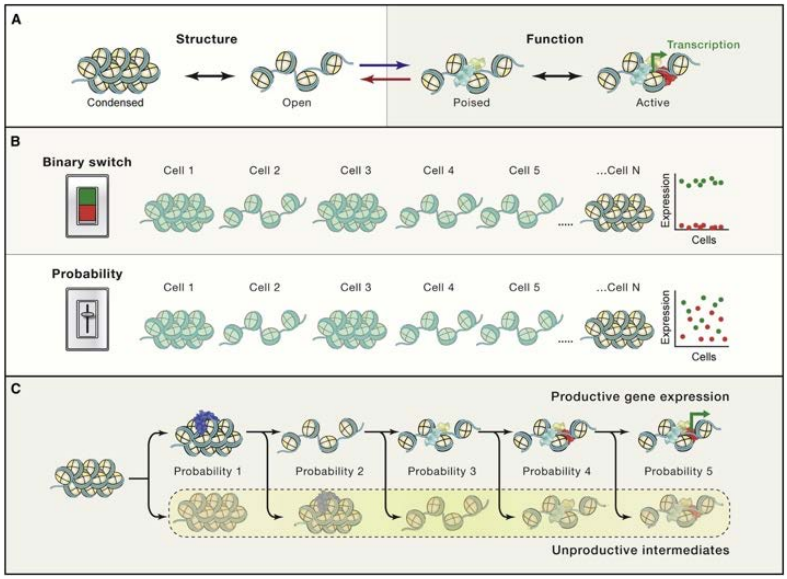


**Figure 2: Plasticity and stochasticity of the genome**

*(A) The regulation of both the structure and the function of the genome follows probabilistic behaviors, which underlies the plastic nature of certain cellular phenotypic features. Structure and function are strictly connected to each other: open chromatin status poised the chromatin locus for transcription initiation by facilitating the association of transcription factors, and transcription itself promotes chromatin decondensation. (B) The chromatin structure has a more probabilistic modulator function than deterministic binary switch behavior. The expression level of a specific gene in individual cells of a cell population was thus shown to be heterogeneous rather than following a 50% "on" or "off" scheme, indicating the probabilistic nature of the expressivity of the genome. The success of a transcriptional program was thus proposed to be the result of many productive events in equilibrium with the probability of as many unproductive reverse events. (C) The probability of stable interaction with chromatin remodelers is in equilibrium with their transient transition while they diffuse through the nucleus (Probability 1). The success of the first productive event increases the chance of the second one, i.e. the maintenance of an open chromatin state (probability 2), but the unproductive event of a return to the compact status can still take place. In the same way, the open chromatin increases the probability of association of early transcription factors (Probability 3), which in turn increases the chance of the recruitment of the RNA Pol II (Probability 4) and of the transcription activation (Probability 5). But all those events are in equilibrium with opposite unproductive intermediates: disassociation of early transcription factors, loss of late transcription factors, or dissociation of RNA Pol II. This probabilistic nature is what generates the stochastic and plastic nature of gene activation, and, in a similar way, of other cellular functions.*

*From The Self-Organizing Genome: Principles of Genome Architecture and Function. Cell 183, 28–45 (2020), https://doi.org/10.1016/j.cell.2020.09.014.*

## 1.4 The war against cancer

New single-cell studies and mathematical approaches that take into consideration the plasticity and stochasticity of biological processes have slowly started to change the dogma of the somatic mutation theory of cancer, i.e. that mutations and/or expression change in oncogenes and tumorsupressor would be the only needed and sufficient causes of tumor formation and progression[215].

It is, indeed, clear now, that cell populations in general possess a certain heterogeneity that is not linked only to genetic variations[12,212,216]. Such heterogeneity can result either

from non-genetic phenotypic heterogeneity among an isogenic group of cells harboring stable phenotypic states or from phenotypic plasticity that allows cells to adopt transiently different phenotypic states[217]. Epithelia-mesenchymal transition, drug resistance, cell proliferation, and formation of primary tumors and metastases are all heavily dependent on phenotypic plasticity and not always explainable by only phenotypic heterogeneity[217].

In the original Waddington landscape model adapted to describe tumor development, new genetic mutations would create phenotypic heterogeneity, i.e. new valleys, thereby increasing the entropy of the system, and enabling the emergence of new paths for the balls to run into as cancerous states[205,217]. On the other hand, phenotypic plasticity increases the entropy, the noise, of the system by allowing the balls to enter in transit paths that they normally would not be able to enter, and thereby make their behaviour much more dynamic[217,218]. If environmental or other signals would, later, reduce the entropy of the newly created valleys, they might become the new dominant states[217]. This type of behaviour is selected upon in nature to allow cells to rapidly respond to changes in the surrounding environment, but, at the same time, it was proposed to stochastically enable cells to enter into possible cancerous states[217].

Phenotypic heterogeneity generally increases during aging, where somatic mutations and epimutations tend to accumulate in the cells of an individual[219,220]. Although this phenomenon has been long considered to be a contributing factor to aging itself, no strict evidence has yet been produced, even if new observations with advanced technologies are still pointing in that direction[219]. Epimutations and perturbed epigenetic plasticity emerging during the aging process might thus represent an underlying mechanism contributing to the observed increased cancer risk[221] during aging[220,222].

At the molecular level, epigenetic regulation is considered to be one of the main mechanisms underlying both plastic heterogeneity during development and pathologically increased plasticity in cancer: genetically identical cells can thus have distinct reversible and heritable epistatus that can cause differences in, for example, the transcriptional noise or busting frequency of specific genes[215,217]. Pathologically increased phenotypic plasticity in cancer has been proposed to be linked to stochastic erosion of LOCKs[223,224] that overlap with large hypo-methylated blocks[225] observed in cancer and upon cancer predisposing stimuli and aging, which carry the most "variably expressed" portion of the cancer genome involving genes with cancer-relevant functions[226]. Such heterogeneity, arising from stochastic variations in unstable, plastic epigenetic states, has thus been proposed to drive cancer evolution under changing selection pressure[226].

Another source of pathologically increased phenotypic plasticity in cancer was proposed to be linked to the intrinsically disordered proteins, enriched in intrinsically disordered regions[215]. These factors were thus proposed to cause "conformational noise"

in different cells and, depending on the local conditions, create complexes with several different partners or condensates of different types[4,215].
Finally, also the tumor microenvironment appears to contribute to an increase in phenotypic plasticity of the tumor cells[215] by, e.g., increased variability of tumor-associated cells that can activate pro- or anti-cancer signaling or by influencing the oxygenation and vascularization of the tumor[215].

Alterations in the level of the *MYC* oncogene or the activity of the WNT signaling pathway are two of the most common features in cancer cells[227,228] - both of which have an essential role in increasing phenotypic plasticity[229,230] and giving cancer cells evolutionary advantages[227,228], allowing tumor maintenance and progression[227,228].

### 1.4.1 MYC: the Odysseus of the Cancer epic

#### *1.4.1.1 MYC and its role in tumor development*

One of the most common deregulated genes in the majority of cancer types are *MYC* genes[227], coding for a family of 3 transcription factors MYC, MYCN, and MYCL, of which the first one is probably the most studied[227]. The causes of pathological MYC expression[231,232] include chromosomal translocations, copy number changes, genetic mutations, epigenetic alterations, but also deregulation of many signaling pathways impinging on *MYC* expression, such as WNT. As a versatile TF, MYC is essential to promote cancer growth and maintenance[123,227,231,233], as its inhibition in tumours with altered MYC expression generally causes tumor regression[233]. Apart from effects on cell proliferation, the underlying mechanisms include inducing cellular senescence[234], modification of the tumor microenvironment including involution of the vasculature[235], and reactivating immune recognition[236].

Structurally, the oncoprotein MYC has, at its carboxy-terminal, a highly-conserved helix-loop-helix and a leucine zinc-finger domain that allow recognition of Enhancer-box (E-box) sequences and dimerization with other proteins containing the same domain, like MAX that is essential for MYC function[237], or MIZ1 that, complexed with MYC, works as a transcriptional repressor of MYC targets[238].

In addition, MYC contains other six high-conserved sequence regions called boxes, which are able to form a complex with a high number of different proteins influencing in MYC stability, chromatin remodeling and modification, promoter affinity, and chromatin association[231]. This feature helps MYC to recognize specific targets, as exemplified by its interaction with WDR5 at certain promoters[239].

What makes MYC such a challenge to target in cancer is that it seems to affect - activate or repress[231] - a huge number of heterogeneous set of genes, genes transcribed by all three RNA-Polymerases, including tRNAs, genes belonging to different transcriptional programs, cell-adhesion, cell-cycle, mitosis, apoptosis, translation, and

many other[231]. Importantly, however, more surprising is the fact that the effect of MYC on its target's expression is generally below two-fold[231]. Several models have tried to explain how MYC could be such a powerful agent without drastically perturbing the expression of its targets. One model is built on evidence showing that MYC regulates a specific set of genes, in spite of binding to a large fraction of the genome[238,240]. Another model, instead, views MYC as a global transcriptional amplifier that increases the overall transcription rate of its already active enhancer and promoter targets[123,241,242] – an interpretation that could explain the different tumor type-specific effects of MYC on gene expression[123,241]. In line with this model, apart from recruiting chromatin remodeling factors and transcription factors, as well as promoting transcription condensate, recent studies have shown that MYC is able to recruit transcription elongation factors to regulate the passage of Pol II from non-productive to the productive mode and balance promoter-proximal pausing and the bursting frequency of the transcription process[175,231,243]. Finally, a third model attempts to combine the first 2, proposing that the effect of MYC is promoter-affinity dependent, and promoters with higher affinity and higher amount of MYC are more affected by MYC binding[244].

Interestingly, an array of post-translational modifications has been identified to affect not only MYC stability, DNA binding, and function, but also its sub-nuclear localization[245]. For example, MYC protein can be phosphorylated at serine 62 upon growth stimuli, which facilitates its dimerization with PIN1. Its interaction with PIN1, in turn, causes the re-distribution of MYC to the basket region of the nuclear pore complexes, where it recruits histone acetyltransferase GCN5, leading to the upregulation of the transcription of the genes involved in proliferation and migration pathways[246].

Finally, the challenge in targeting MYC in cancer is compounded by its ability to increase chemoresistance by upregulating the ATP-binding cassettes (ABCs), cytoplasmic membrane transporters involved in expelling drugs from the cells[247]. In a fascinating study on HER2+ breast cancer cells, it has been shown that the HER2 kinase inhibitor Lapatinib induces the tumorsupressor FOX2 that, unexpectedly, also upregulates *MYC* expression by recruiting histone methyltransferases and histone acetyltransferases to the *MYC* gene. *MYC* overexpression, in turn, increases the level of ABCs and the resistance of the tumor cells to the drug in a negative feedback loop[248].

To be able to target MYC for tumor treatment, it is essential not only to unravel the mechanism of its action but also to explore how pathological MYC expression is achieved in tumor cells[232,249–251]. Hypermethylation and hyperacetylation are some of the most common epigenetic changes observed at the *MYC* locus in cancer cells[232]. Other common mechanisms underlying the deregulation of *MYC* expression are mutations or epigenetic changes in its regulatory elements, enhancers, and super-enhancers[232,250]: in fact, one of the first evidence linking the development of cancer to modifications of DNA

regulatory elements was the demonstration of the recurrent translocation of the *MYC* locus to the immunoglobulin heavy chain enhancer region in Burkitt lymphoma[251].

*MYC* is indeed located on a ~2Mb "gene desert" region forming a single TAD in the band 24 of the long arm of chromosome 8 (Chr8q24), within which *MYC* is surrounded by cell type-specific enhancers and super-enhancers delimitated in sub-TADs[252]. Alterations of these regulatory sequences are found in several types of cancers, and they are often the main cause of cancer growth[251].

#### *1.4.1.2 Regulation of MYC expression: the roles of the OSE and CCAT1 ncRNA*

The "gene desert" area around *MYC* has for a long time been associated with cancer hallmarks and cancer-associated single nucleotide polymorphisms (SNPs), amplifications and translocations[253–255]. The molecular function of these genetic changes has only recently been understood, when scientists realized the presence of many regulatory elements such enhancers and, later on, super-enhancers in the region[256].

Some of those regulatory elements have physiological functions also in normal development, but they might become upregulated and altered in cancer cells to sustain high MYC expression[257,258]. An example is illustrated by the contribution of the hematopoietic enhancer cluster to not only the physiological transcriptional regulation of *MYC* in mouse and human hematopoietic stem cells and progenitors, but also, in its altered form with more prominent chromatin accessibility, to the high level of *MYC* transcription needed for the maintenance of leukemia cells in mice[257].

There are several other regions that can acquire cancer-specific genetic or epigenetic mutations and become active enhancers or super-enhancers without any known function during normal development[251,252,259], as illustrated by the emergence of prostate, breast, or colorectal cancer super-enhancers regulating *MYC*. Those alterations promote loop formation between the newly-formed oncogenic enhancers (OE) or oncogenic super-enhancers (OSE) and *MYC* promoter, causing its transcriptional upregulation[252].

Complicating the picture, such regulatory loops between *MYC* and its enhancers might require a CTCF-occupied enhancer docking site[250]. Schuijers and collogues[250] have thus shown that a region proximal to the *MYC* promoter and its OSEs interact to form loops between the regions located far apart on chromosome 8[250]. A CTCF-occupied region proximal to the *MYC* promoter, which is generally hypomethylated in cancer, has thus turned out to be important for the formation of such long-distance interactions, working as an enhancer-docking site for *MYC* regulation[250]. According to the authors, other oncogenes might be regulated in a similar way, and epigenetic editing could inactivate those enhancer-docking systems[250].

Many studies have observed an eRNA transcribed from the colorectal super-enhancer region 515kb upstream of the *MYC* locus[254,260–264]. Initially, the spliced and poli-

adenylated lncRNA has been identified as a highly specific, easily detectable biomarker for colorectal cancer and other tumors[260]. As it is not expressed in normal cells, it was named Colon Cancer Associated Transcript-1 (*CCAT1*)[260].

Later results have shown that *CCAT1* is an eRNA[262] containing 2 exons and is transcribed in 2 forms: a long version, *CCAT1-L* of 5200bp that is localized to the nucleus and associated with chromatin, and a short form, *CCAT1-S* of 2600bp that is mainly cytoplasmatic and is probably derived from the long-form, since its expression is strictly correlated with that of the long-form[262]. *CCAT1* is transcribed from a 150kb long super-enhancer located 515kb upstream of the *MYC* locus, with which it forms a loop[262]. The same authors have also observed that the down-regulation of *CCAT1* causes downregulation of *MYC* mRNAs, and that only its overexpression *in cis*, achieved by TALEN technology, and not *in trans*, where it is overexpressed by the transfection of an expression plasmid, causes upregulation of *MYC*[262]. Moreover, in an engineered cell line for increased *CCAT1* expression, *CCAT1-L* has been shown to accumulate at the site of its transcription and bind CTCF to increase its presence in the OSE, and thus consequently modulate super-enhancer – *MYC* promoter looping[262]. In summary, Xiang and colleagues have for the first time shown that *CCAT1-L* might promote tumorigenesis of colorectal cancer by upregulating *MYC* expression, a process that likely involves its ability to modulate CTCF binding to the super-enhancer which it is transcribed from and, in this way, modulate the looping between this regulatory region and the *MYC* locus[262].

Following this study, *CCAT1* has been linked with proliferation, invasion, migration, drug resistance, and survival in several types of cancers including gastric cancer, lung cancer, breast cancer, and many others as reviewed by Liu et al.[265].

Other than promoting super–enhancer-promoter loops, *CCAT1* has been shown to form a complex with transcription factors and activate other super-enhancers in *trans*, as shown by Jiang and collegues[266]: in squamous cell carcinoma, the master transcription factors TP63 and SOX2 drive *CCAT1* expression, which, in turn, forms with them a complex that binds and activates an EGFR super-enhancer, leading to the activation of two signaling pathways, i.e. that of MEK/ERK1/2, and PI3K/AKT, that jointly promote tumor development[266].

Furthermore, *CCAT1* overexpression in esophageal squamous cell carcinoma has been shown to recruit Polycomb proteins, such as the PRC2 complex, and SUV39H1 to regulate histone methylation at *SPRY4* and, in doing so, promote tumor growth and migration[267].

Finally, in the cytoplasm, *CCAT1* seems to work as sponge RNA for miR-7, a microRNA that downregulates HOXB13 to thereby facilitate tumor survival and metastasis[267]. The function of competitive endogenous RNAs, or sponge RNAs, seems to be one of the

most common functions of *CCAT1* in many different types of tumors, a mechanism that promotes many tumors' survival mechanisms[265].

Super-enhancers and their eRNAs are generally the targets of developmental signaling pathways that regulate their transcriptional activity and chromatin state[268]. The oncogenic colorectal super-enhancer has, e.g., 4 binding sites for TCF4, the terminal transcription factor of the WNT cascade, one of the essential upregulated pathways in many tumors[169].

### 1.4.2 WNT signalling in tumor development

WNT is one of the most commonly altered signaling pathway in malignant tumors[228]. First discovered to regulate developmental processes in mice and Drosophila[269], WNT has been intensively studied, as it controls transcription activities of essential genes involved in the cell cycle, stemness, differentiation, proliferation, and morphology[270].

In mammals, there are 19 WNT genes with individual as well as overlapping functions during development, which can activate 3 distinct pathways[269]: the canonical β-catenin-dependent pathway, the WNT/PCP (Planar Cell Polarity) pathway involved in the organization of the plane of cells within a tissue, and the protein kinase C (PKC) WNT-dependent pathway that is responsible to increase the intracellular level of Calcium ($Ca^{2+}$) and the activation of its downstream effector[228,269].

Whereas the two non-canonical, β-catenin-independent, pathways are not well-characterized despite their links to cancer metastasis[228], much more data have been accumulated for the "canonical" β-catenin cascade and its links to several cancer types[228,271].

In absence of WNT ligands, in the "off-state", β-catenin is thus phosphorylated by two kinases, first by CK1, followed by GSK3, in its N-terminal that contains a series of serine/threonine motives[269,270]. CK1 and GSK3 are part of the destruction complex (DC) together with Axin that works as a scaffold and directly binds β-catenin and APC[269,270]. Phosphorylated β-catenin interacts with F-box-containing protein E3 ubiquitin ligase β-TrCP that, in turn, ubiquitinates β-catenin and promotes its proteasomal degradation[270]. In epithelial cells, β-catenin plays a major role also in adhesion junctions *via* acting as a binding partner of several cadherins, but this function is independent of the signaling one[269,270].

When WNT ligands bind the cellular transmembrane receptor Frizzled, i.e. in the "on-state", another component of the DC, Disheveled (Dvl), interacts with the receptor, causing the DC complex to re-localize to the cellular membrane[269]. Re-localization of the DC complex to the cell membrane induces a change in its activity, and, as a consequence, the accumulation of un-phosphorylated β-catenin[269]. Although how exactly this happens is not yet clear, one model proposes that ubiquitination is blocked

at the cell membrane, causing the DC complex to get saturated with phosphorylated β-catenin, and allowing free β-catenin to accumulate in the cytoplasm[269]. Alternatively, the β-catenin inhibitory domain of APC might play a role in downregulating β-catenin phosphorylation[269]. Finally, the WNT ligands can co-binding the co-receptor LRP, which causes receptor phosphorylation and interaction with the DC complex as well as the inhibition of GSK3, thereby promoting β-catenin stabilization[270].

When free to accumulate, β-catenin translocates into the nucleus where it binds to the transcription factor family TCF/LEF and transforms them from transcriptional repressors to transcriptional activators[270] of genes containing WNT Response Element (WRE) motives[269]. In the "off-state", TCFs are bound to Groucho, and together, by recruiting histone deacetylases, repress genes by establishing compact chromatin[269,272]. The β-catenin/TCF complex is further regulated by competitors, such as ICAT or Cby[273,274] that will impede the formation of the complex, and also by TCF/LEF isoforms produced by alternative splicing that lack β-catenin binding domain and will therefore compete for the binding of WRE without the transactivation power of β-catenin[275]. In addition, the β-catenin-TCF complex can be stabilized by other factors, such as RNF14, which seems to be also crucial for the survival of colon cancer cells[276].

To achieve its role as a transcriptional activator, β-catenin can interact with many other co-factors, including Pontin52 that co-bind at the same time the transcription factor TBP[277]; the acetyltransferase CREB, role of which seems to be promoter-specific[278]; Brg-1, a component of mammalian SWI/SNF and Rsc chromatin-remodeling complexes[279]; the Mediator complex subunit MED12[280], and many other as reviewed by Söderholm and Cantù[269]. Recent observations have, furthermore, emerged showing that β-catenin can also interact with developmental- or cell type-specific transcription factors that can cooperate or compete with TCF4 to supply β-catenin a DNA-binding domain[269,281]. Although these discoveries collide with the definition of "canonical" WNT signaling, they can be used by the cell to regulate very specific transcriptional programs[269,281]. This would also justify the observation of tissue-specific β-catenin effects[282].

Considering the heterogeneity of co-factors that β-catenin can bind to, it is not surprising that deregulation of these pathways can have enormous consequences on development and are linked to numerous types of cancers[228,270,271,283], especially a subtype of colorectal cancer that lacks the subunit of the destruction complex APC[228]. This causes the continuous stimulation of WNT signaling and the upregulation of genes including *MYC*[249] that are essential for tumor maintenance and growth[227,231].

# 2 Research aims

The role of the nuclear architecture in the regulation of transcriptional programs during development and in cancer cells is extensively studied[5,221,284], although many details of the picture are still missing. For example, the classical model of enhancer-mediated transcriptional regulation by loop formation is not always able to properly describe what has been observed by advanced technologies[36,37]. Moreover, very little is known about how enhancer-promoter contact frequencies and functional outcomes relate to structural hallmarks of the nuclear architecture. Super-enhancers and eRNAs have thus often been linked to a variety of functions besides the regulation of target gene transcription[166,169], and the role of the oncogenic super-enhancers has not yet been completely understood[149,252]. Despite the increasing amount of evidence showing the direct and functional binding of certain mobile nucleoporins to chromatin at regulatory elements, such as developmental super-enhancers and oncogenic super-enhancers, in mammalian cells[79,81], before the work presented in this thesis, the gene gating mechanism has been proved to exist only in lower organisms[194,196]. Interestingly, our group has previously observed that the circadian genes that rhythmically visit the nuclear periphery remain active at the lamina for several hours before gradual transcriptional attenuation[43], suggesting that they might land at a transcriptionally permissive environment, such as the nuclear pores. Focusing on *MYC*, one of the most commonly deregulated oncogenes in many types of tumors[227] that displays circadian expression in many different model systems[285–287], we have thus aimed at exploring the following questions:

- How are enhancer-promoter interactions integrated in the sophisticated landscape of the 3D nuclear architecture? More specifically, does the gene-gating phenomenon exist in human cells at the *MYC* locus (Paper I) and what are the underlying molecular mechanisms (Paper II)?
- What is the role of the OSE eRNA, *CCAT1*, in the regulation of *MYC* transcription and gating (Paper III)?

# 3 Materials and methods

The methods described in the papers included in this thesis can be divided into 2 main groups: technologies that analyze a cell population and single cells, respectively. As discussed later, both approaches are essential to generate a complete overview of the observations described in the papers that includes an assessment of features, such as plasticity and stochasticity.

In this paragraph, some of the methods used to achieve the results presented in this work will be described in general terms with the purpose of highlighting the principles and the advantages or disadvantages associated with each technique. Detailed protocols can be found in the material and methods sections of each individual paper.

## 3.1 Cell culture

The human colon cancer cell line, HCT116, is frequently used as a model system for colon cancer, while human colon epithelial cells or HCECs, were used as reference of a normal system.

HCT116 cells were donated kindly by Professor B. Vogelsten (Johns Hopkins Medical Schools and Sydney Kimmel Comprehensive Cancer Center) and maintained in a McCoy 5A Modified Medium containing GlutaMAX, 1% Penicillin/Streptomycin, and 10% Fetal Bovine Serum at 37°C and 5% $CO_2$. This cell line has a trisomy of chromosome 8 and an altered WNT pathway, with an autocrine loop and a mutation of β-catenin that impedes its degradation[288,289].

The HCECs were purchased from Sciencell and maintained in Colonic Epithelial Cell Medium.

The CTCFBS mutated clones E3 and D4 were generated by the electroporation of ribonucleoprotein of spCas9 and specific guide RNA, and donor DNA, as illustrated in Paper II. The clonal selections were made by seeding single cells in multiwell plates.

The use of cell lines gives the advantage of having a controlled, more homogeneous environment, easier to manipulate and control. However, these conditions do not represent a complete simulation of events within an organ or organism. Thus, the three-dimensional perspective with the organization of different cell types and their exposure to microenvironmental cues are not recapitulated. For the future, the use of organoid or 3D-culture and biopsies from patients will provide an important step for a closer approximation of *in-vivo* conditions to further validate the discoveries of this work.

To avoid any side effects influencing cell viability, for example, the treatment of cells with drugs were optimized to generate the most profound effect with the lowest possible dose and incubation. For BC21, which inhibits β-catenin/TCF4 complex, it was,

moreover, important to maintain the presence of the drug also following cell lysis to avoid reformation of the complex during the extraction procedures.

siRNA transfections were performed by using the lipofection approach and included the optimization of the amount of siRNA, lipofection reagents, and incubation time to achieve the highest possible effect on targeted RNA expression. For transcription products retained in the nucleus, such as the eRNA *CCAT1*, longer treatment for 72h was necessary to achieve efficient downregulation.

## 3.2 Single-cell methodologies

In recent years, many technologies have been developed to transform classical molecular biology techniques, like western blot or qRT-PCR, into single-cell versions. However, this scaling-down approach generally requires advanced protocol and expensive instrumentations. Importantly, they don't offer any information related to the cell or nuclear architecture. Conversely, a majority of single-cell methods described in this work are based on fluorescence microscopy, which has the benefit of allowing the visualization of spatial intensity features within single cells.

### 3.2.1 Nodewalk

This methodology was optimized by our group and described in detail in other publications[214,290]. Even though it is not a single-cell technique, it represents an ultrasensitive 3C (Chromatin Conformation Capture)- based technique able to quantitate chromatin fiber interactions in as little as 7 cells. Briefly, ligated DNA fragments, corresponding to interacting chromatin fibers, after a tagmentation process made by a modified Tn5 transposase, are converted into RNA molecules by *in-vitro* transcription using primers complementary to the adapter sequence and containing T7 RNA polymerase promoter and Illumina P5 sequences that will be added to the converted product. Following these steps, it is possible to use any primers against the desired target (bait) to be able to sequentially identify its interactors by sequencing. This principle of "walking" on the interactors can be used to build a highly interconnected network of chromatin fiber interactions.

In Papers I and II, the *MYC* and the OSE alleles were used as baits to analyze their interactomes.

### 3.2.2 3D DNA- and RNA-FISH

The RNA/DNA fluorescent *in-situ* hybridization techniques are essential for the determination of the position of DNA loci and their derived RNA transcripts in relation to the nuclear architectures, in single cells[14,291]. These approaches are straightforward in that they are based on oligonucleotide probes containing modified fluorescent nucleotides that hybridize to the complementary target within fixed cells. The

formaldehyde fixation of living cells ensures that 3D structures are recapitulated despite their permeabilization using detergents.

The main difference between the RNA- and DNA-FISH techniques is that the former does not depend on a denaturation step. The authenticity of the RNA-FISH was additionally tested by RNase treatments. To further ascertain that the RNA-FISH signal is correct, the same sample was denatured at 80°C for 40 minutes in a solution containing 50% formamide. This step not only removes the DNA:RNA hybrids constituting the RNA FISH signal but also separates the DNA strands to enable detection of the distribution of the individual template generating the identified transcript within the nuclear architecture. Here, we used small PCR-made fluorescent probes complementary to 8-10kb regions of *MYC* promoter and gene body, *CCAT1* or the OSE, in combination with labeled Bacterial Artificial Chromosomes (BACs) complementary to the same region. This approach further ensured the authenticity of the FISH signal by allowing the identification of any off-targets.

The hybridization time, denaturation conditions and concentration of the probes were all optimized to significantly reduce off-targets and background signals while producing optimal signal-to-noise ratios.

### 3.2.3 ISPLA and ChrISP

To be able to visualize possible interactions or close physical proximities within single cells, we relied on the *in-situ* proximity ligation (ISPLA)[292] and chromatin *in-situ* proximity (ChrISP)[293] assays. These techniques represent very powerful methodologies since they allow the screening of potential *in-situ* proximities between 2 targets. While ISPLA identifies proximities between two proteins, the ChrISP technique allows the visualization of the proximity between two chromatin loci or between a chromatin locus and a protein with a resolution down to 162Å (≈16,2nm)[293]. They thus complement the information generated by the biochemical Co-Immunoprecipitation (Co-IP) and Chromatin immunoprecipitation (ChIP) techniques by providing information of its spatial distribution within the cell.

For ISPLA, the epitopes of the two proteins representing candidates for interactions are targeted by primary antibodies produced in two different hosts, one mouse, and one rabbit. Following washes, the primary antibodies are recognized by secondary antibodies containing oligonucleotides, R+ (priming anti-rabbit), and M- (non-priming anti-mouse). Those oligonucleotides are complementary to backbone and splinter oligonucleotides that will be able to anneal to them only if they are in sufficiently close proximity to each other. The annealed complex is then stabilized by introducing a ligation step that generates a circular DNA molecule that enables detection by a rolling circle amplification reaction. The hybridization of fluorescently labelled small detection oligonucleotides complementary to the amplified splinter allows the visualization of the

proximal targets. It is important to keep in mind that the stereochemistry of all players and the occupancy of the epitopes with the ISPLA secondary antibodies might impede the visualization of the two primary targets.

The optimization of the primary antibody concentrations is essential to guarantee no saturation of the signal and a good signal-to-noise ratio, as well as no background. As a negative control, we routinely use omission of either the primary antibodies or the M-mouse antibody.

In ChrISP, the first step is to proceed with a DNA-FISH protocol to hybridize the 2 targets with digoxigenin- or biotin-labeled probes that represent the epitopes of the primary antibodies. If the aim is to examine the proximity between a protein and a chromatin locus instead of two chromatin loci, one of the primary antibodies will be targeted against the digoxigenin-labelled DNA-FISH probe and the other against an epitope of the protein of interest, such as a histone modification, NUP or a transcription factor. The steps following the addition of the primary antibodies follow closely the ISPLA protocol. The main difference is that the ligated circular DNA is not amplified. Instead, it is visualized by the fluorescently labeled splinter. This omission of the rolling circle amplification is essential since densely packed chromatin impedes the efficiency of the rolling circle amplification and thus generates a bias in data.

In both cases, the staining of the cells with DAPI or other cytoplasmic/nuclear staining markers facilitates the analyses of the distribution of the proximity signals within the cellular architecture

### 3.2.4 Fluorescent Widefield Microscopy

To visualize all the fluorescent signals described, we used widefield fluorescence microscopy represented by a Leica DMi8 inverted microscope equipped with a DFC9000 camera and a Thunder Imaging System. To attain the most optimal resolution, a HC PL APO 63X oil objective with a numerical aperture of 1.4, one of the highest available in the market, was selected.

Compared to a confocal system, the widefield microscopy[294] provides much more flexibility, in terms of the number of fluorophores that could be imaged together (this system was designed to be able to simultaneously collect 6-7 fluorophores), higher acquisition speed, less photobleaching since it is based on a LED-illumination system, less-parameters to optimize and generally cheaper[294]. The resolution is independent of the type of system, since it is a pure physical property linked to the magnification and numerical aperture of the objective (and the refractive index of the medium for the axial resolution), and to the wavelength of the fluorophore, that is, the same objective and same fluorophore would give exactly the same resolution in both confocal and widefield microscopy.

The reason why confocal is very often preferred is because it gives a better signal-to-noise ratio, since it is able to remove the out-of-focus light thanks to the pinhole technology[294].

The Thunder technology incorporated in the system used for this work resolves the problem by applying a computational clearing (patent by Leica Microsystems) that removes the out-of-focus light without altering the information of the pictures or introducing artifacts[295] and, at the same time, allowing to benefit of all the advantages of a widefield system.

To be able to have 3D information, optical sections were acquired using LAS X (Leica Application Suite X, Leica Microsystems) software-optimized z-axis intervals that avoid under- or over-sampling (Nyquist Sampling).

## 3.3 Cell population methodologies

When RNA, DNA, fixed protein/DNA complexes, etc are extracted from a cell population, the obtained results represent a snapshot of events happening within the cell population. At the same time, the data does not distinguish whether the specific observed events are equally happening in all the cells, or if they are happening with much higher frequencies in small subsets in the population. Nonetheless, they constitute indispensable techniques to be able to analyze overall changes within a population following specific treatments or culturing conditions.

### 3.3.1 Nascent RNA and Export Assay

The key methodology that the majority of the work described in the thesis is based on is represented by the dynamics of the nuclear export of newly synthesized RNA to the cytoplasm.

To this end, we pulse-labeled cultured cells with 5'-Ethynyl-Uridine (EU) to partially replace uridine in newly transcribed RNA. This moiety will later allow covalent bonding to an azide-containing molecule like biotin that can be used to specifically purify labeled RNA. This principle is possible thanks to the extraordinary power of the click-it chemistry[296], which was awarded the Nobel Prize in Chemistry in 2022. Streptavidin magnetic beads are successively used to pull down the biotin-EU-labeled newly synthesized RNA that, following conversion into cDNA, can be analyzed for specific transcripts by quantitative Real-Time Polymerase Chain Reaction (qRT-PCR).

To distinguish newly synthesized RNA from the total RNA pool, the timing of the EU labeling is critical. Transcription is generally a fast process that is completed within minutes, so to be able to catch such a phenomenon, the EU labeling must negotiate the transport of the nucleotide into the cell, its conversion into triphosphates before the EU molecules have the potential to reach the nucleus and the transcriptional process. We

have determined that to encompass these events and yield RNAs that have not yet been exported into the cytoplasm, the labeling period should last for only 30 minutes.

Following the 30 minutes pulse, excess EU in the medium is washed away to chase the labeled RNA for periods of up to 1 hour. During this period, the newly synthesized mRNA will exit the nucleus in a manner reflecting nuclear processes, such as gene gating. At this point, the cells are harvested and the cytoplasmic and nuclear compartments purified using the Thermo Fisher Scientific PARIS kit that allows differential lysis of plasma and nuclear cell membranes by nonionic detergents.

RNA extracted from the cytoplasmic and nuclear fractions are then subjected to the click-it reaction, the biotin pull-down, the cDNA retrotranscription, and the qRT-PCR. This approach enables the quantification of the amount of newly synthesized RNA in the nucleus and in the cytoplasm and thus the calculation of the export rates. These are usually defined by determining the ratio between cytoplasmic and nuclear levels of EU-labeled RNA representing products from individual genes.

To account for possible contamination between the two fractions, the export rate formula is integrated with calculations that estimate the presence of *MYC* intronic sequences in the cytoplasmic fraction and the presence of the mitochondrial *CYTB* transcript in the nuclear compartment, respectively.

To control the efficiency of the entire purification protocol, an *in-vitro* transcribed EU-labeled luciferase mRNA is added as a spike-in tracer before the click-it reaction and used to normalize the recovery of the EU-labeled RNA in qRT-PCR data before applying the export rate formula.

Comparing total RNA and nascent RNA expression and export rate, it is possible to obtain a detailed picture of the effect of a specific treatment or condition on the transcription and nuclear export process of a specific target. To normalize overall levels of mRNA expression, we generally use mRNA levels from the housekeeping *TBP* gene. However, to normalize nascent RNA levels representing transcription rates, we instead use 18S rRNA levels as a reference. The reason is that the nascent *TBP* RNA levels are too low to faithfully represent a normalizing factor.

### 3.3.2 CoIP and ChIP

Co-immunoprecipitation (Co-IP)[297] and Chromatin immunoprecipitation (ChIP)[298] are biochemical techniques used to study protein-protein and protein-chromatin interactions, respectively.

For the Co-IP the starting point is the cell lysate (or the purified nuclear fraction if the target is a nuclear protein). However, the conventional ChIP technique requires that the

protein-chromatin interaction is first fixed with formaldehyde before purifying the nuclear compartment.

The basic principle for the two methods is otherwise similar in that an antibody against an epitope of the selected target is mixed with the lysate and incubated to allow the binding. Following that, the mixture is incubated with immunoglobulin magnetic beads to capture the antibody-protein complex on a magnetic rack.

For the Co-IP analyses, the purified complex is analyzed by Western Blots using an antibody against the predicted interaction partner. Alternatively, the purified complex can be analyzed by mass-spectrometry to identify all possible interactors.

Conversely, for the ChIP analyses, the purified complex is first subjected to reverse-crosslinking to separate the protein from the chromatin followed by the purification of the DNA. Using specific primers individual regions of interest can be tested for their ability to interact with the protein of interest by q-PCR. Following comparisons to the input material and no-antibody control, the calculated percentage of recovery indicates the amount of target protein that binds the specific DNA region of interest. Additionally, the purified DNA can be analyzed by next-generation sequencing to observe genome-wide interactions with the protein target of choice (ChIP-seq).

The optimization of antibody concentration, the stringency of the washing procedure, fixation protocol conditions (for ChIP) represent essential steps to guarantee reproducible results. Moreover, it is important to include controls to avoid misinterpretations. In both cases, a no-antibody control, or immunoglobulin G (IgG) control, where the lysate is mixed with a no-antibody or IgG-control antibody, is needed to check non-specific background signals.

For Co-IP, it is important to use Western Blot to quantify also the recoveries in comparison to input material, which serves to determine the intensity of interactions.

For ChIP, it is important to include in the q-PCR analysis primers for sites where the target is not supposed to bind, such as alpha-satellite DNA.

Both techniques benefit from the use of positive controls that are usually represented by proteins known to interact with the target (in the Co-IP) or primers for a region known to have a strong binding with the target (in the ChIP).

### 3.3.3 Simple Western analyses

To quantitatively identify protein epitopes in purified Co-IPs, a Simple Western (Wes and Jess instruments, Protein Simple, Bio-Techne) methodology has been optimized and implemented[299].

Compared to classical Western blot methods, this technology allows the use of very low input amounts and antibodies – thus a particularly useful feature for Co-IP samples. Moreover, it is fully automated, except for the preparation of the plate, more sensitive and much faster, to allow runs of a full experiment from electrophoresis to antibody incubation and chemiluminescence reading in a little bit more than three hours.

It is a capillary-based technology: the sample, buffers, and primary and secondary antibodies are loaded manually in a specific plate that is later loaded into the instrument together with a capillary cartridge. Here, completely automatically, the separation matrix and then stacking matrix are aspirated in each capillary followed by electrophoretic separation of the samples. Next, UV light is applied to fix the proteins to the wall of the capillaries with the matrix being washed out prior to the immunoblotting and chemiluminescence steps. The results are shown as electropherograms and virtual blot-like image, on which it is possible to calculate molecular weight and signal intensity (as the area under the peak at a specific molecular weight position).

The technology allows also for multiplexing, using a combination of chemiluminescence and fluorescence channels, re-plexing, and normalization with total protein using specific kits.

Once the optimization of the concentration of the input material and of the primary antibody has been implemented, the Simple Western technology is a powerful method to quantify precisely and rapidly, with less error-prone steps, and low amounts of proteins in lysates.

## 3.4 Further notes

On some occasions, it has been necessary to combine single-cell and cell population methodologies. This is exemplified in Paper III, which describes the need to grow cells on a cover slip submerged in six-well-plates. Following treatments, the cells on the cover slip were fixed for RNA/DNA FISH analyses, while the remaining cells in the wells were isolated for subsequent determination of treatment efficiencies (by qRT-PCR, for example). In this way, the treatment was uniform to all the cells within the same time period in the same well.

## 3.5 Statistical analysis

Most of the statistical analysis was performed on at least 3 biological replicates. In DNA-FISH experiments, the distance of each allele was considered an independent experiment and the measures of the biologically independent replicas were pulled all together to calculate the statistics. A normality test was used to control the data distribution and decide the appropriate statistical analysis: when it passed, an unpaired two-tailed t-test was performed to calculate the p-value; otherwise, a nonparametric

Kolmogorov-Smirnov unpaired t-test was used. For fold change analysis, in Paper III, it was considered to be more appropriate to use a one-sample test by which it is possible to compare the hypothetical value of the controls (100%), with the mean of at least three independent biological replicas.

## 3.6 Ethical Considerations

No ethical permits were required for the work in this thesis, since neither animal nor patient material has been used.

Nevertheless, some ethical considerations, such as the type of reasoning often used by researchers and the problem of the background theories, are important to discuss.

Induction is the most common reasoning modality used by scientists working in empirical sciences like biomedicine and molecular biology, contrary to the scientists involved in formal sciences, like mathematics and physics, which use mainly a deductive type of reasoning.

In induction reasoning, if all the premises are fulfilled, the conclusion is considered verified and true; it goes from specific to generalized conclusions. If, for example, a specific behavior has been observed in a specific cell line, it is generally thought that all cell lines of the same type, with the same condition, would have the same behavior. Statistical analysis would only be used to assess the statistical significance of the behavior observed in that cell line, but not how generalized that behavior is. The induction reasoning is very common in biology and alternatives are not really possible: scientists think that biological processes such as DNA replication and transcription, are the same in all cells of the same type, in all humans when the same conditions have been met. To determine if that is an absolute truth is, however, impossible. Those conclusions are based on the fact that previous studies have shown that cells with certain features are all similar between them, but again, also in that case an induction reasoning was applied: that form a circle with no end, which is the main problem of this type of reasoning.

The work in this thesis, however, has also contributed to show that plasticity and stochasticity are basic components of any biological system, making it almost impossible to use induction reasoning for biological observations.

Having said that, current methodologies and technologies don't allow ignoring the induction reasoning, but scientists can and should always acknowledge in their work the limitations of such type of generalized thinking. At the same time, it should be limited by strengthening as much as possible the premises of their conclusions and by including assays that take into account the roles of plasticity and stochasticity in their

observations. That is possible by employing single-cell methodologies and personalized medicine approaches.

Another common problem in empirical sciences is the background theories: researchers base their own observation on the fact that previous ones have been accepted by the community and because of that they form a type of absolute truth[300]. That has been true for example with the "fundamental dogma of biology: one gene, one transcript, one protein", thanks to which for years non-coding DNA has been called junk DNA, while now we know how much important and how many functions such type of DNA have. If the fundamental dogma would not so hardly be impressed on the scientists' minds, the discoveries related to ncRNAs might have been done much earlier.

Another example is the trust researchers have in instruments, like microscopy, and basic methodologies, like fixation procedures: for years it has been believed that chromatin forms the famous 30nm fiber, and only recently has such a model been discarded, and the 30nm fiber is now mainly considered an *in-vitro* artifact[301]. It is, thus important for researchers to always have a critical mind and question their own and others' works.

Another ethical consideration related to the work here presented is personalized medicine[302,303]. To approach a specific disease of a specific patient as unique could be the game changer for curing complex diseases like cancer: analyzing the heterogeneity of the cancer cells in that patient and finding certain features, like for example the overexpression of *CCAT1* or a higher percentage of cells with active *MYC* gating, instead of other deregulation processes, could help in designing more specific and beneficial treatment.

Research in new technologies and methodologies that would reduce the cost and the logistics to adopt more universally personalized medicine should be favored in order to find the molecular mechanism underlying cell heterogeneities and justify the use of personalized approaches.

# 4 Results

## 4.1 The *MYC* gene is gated in human colon cancer cells (Paper I)

### 4.1.1 The OSE and *MYC* interact with different components of the NPC

For gene gating to take place, enhancer - gene complexes have to be recruited to the nuclear pores[194] – a step that likely involves the binding of NUPs to chromatin. To determine if the enhancer interactome impinging on *MYC* and *MYC* itself was bound to NUPs we compared the network of *MYC* interactors identified by the Nodewalk technique to NUP153, one of the mobile nucleoporins, genome-wide binding data[79]. These results showed that Nup153 bound not only the *MYC* promoter region but also other *MYC* interactors representing enhancer regions flanking *MYC* locus, including one of its most frequent interactors, the colorectal super-enhancer. Surprisingly, no overlap was found between NUP153 and *MYC* interactors representing constitutive LADs, pointing to the possibility that *MYC* transiently interacts with these regions when in the proximity of NPCs at the nuclear periphery.

To explore whether *MYC* was physically recruited to the NPCs, we first explored whether it contacted also "stable" components of the NPC. To explore if NUP133, a "stable" nucleoporin belonging to the NPC ring, would bind regulatory regions at the *MYC* locus and its enhancers, we performed ChIP qPCRs covering key enhancer and promoter elements. Surprisingly, we observed no binding on the *MYC* promoter, but an enrichment of NUP133 binding on a region corresponding to the colorectal super-enhancer, one of the most frequent interactors of *MYC* harboring a CTCF binding site. This oncogenic super-enhancer (OSE) emerges during cancer development and is present in HCT116 colon cancer cells but not in the primary human colon epithelial cells (HCEC). Importantly, also the NUP133 binding in this region appears to be cancer-specific, as we couldn't find NUP133 occupancy on OSE correspondent region in HCEC.

To confirm that the binding of NUP133 to the OSE takes place at the nuclear periphery, we have applied the ChrISP assay to quantitate the proximity between the OSE locus and NUP133 within the nuclear architecture with a modified protocol that includes tyramide signal amplification. In this way, we could not only confirm that NUP133 binds mainly to the OSE and only in HCT116 cells but not in HCEC, but also that such interaction happens primarily near the nuclear periphery.

In summary, we showed a division of work between the *MYC* promoter and the OSE in their interaction with the NPC components: while *MYC* primarily binds NUP153, a mobile component of the nuclear basket of the NPC, probably already in the nuclear interior, the OSE, in cancer cells, bind the NPC ring, mainly at the nuclear periphery.

### 4.1.2 OSE and *MYC* travel together towards the nuclear periphery

An essential step of the gene-gating process is the movement of the target locus and its enhancer to the nuclear periphery, and their engagement in productive interactions in proximity to the NPCs. To explore the distribution of OSE-*MYC* proximities, we have performed 3D-DNA-FISH with small probes targeting the *MYC* and the OSE loci and calculated their distances to the nuclear periphery corresponding to the edge of DAPI staining. The results showed that in HCT116, but not in HCEC, the OSE is generally closer to the periphery than *MYC*, and, interestingly, the difference (c value) between the distance of *MYC* to the periphery (b value) and the distance of the OSE to the periphery (a value) is drastically reduced as they approached positions near the periphery (c value near 0; c=b-a), suggesting that the OSE-*MYC* proximities are highest at the lamina. To confirm this possibility, we performed ChrISP analysis between the OSE and *MYC* regions, which showed the highest accumulation of positive signals at a distance starting from 0,5μm to the periphery.

To test if this phenomenon was specific to the OSE, we performed a similar ChrISP experiment between *MYC* and another enhancer (EnhD) that does not bind NUPs, is more proximal to *MYC*, and strongly interacts with *MYC* in Nodewalk analysis. Surprisingly, in this case, the positive ChrISP signals highlighting the proximity between *MYC* and EnhD, where primarily located in the nuclear interior.

To correlate these data with the transcriptional activity of the *MYC* gene, we performed RNA-FISH with probes specific for *MYC* intron 1 or for the entire *MYC* gene (exons + introns). We observed that while the unspliced form was mainly localized at the nuclear interior, there were two peaks in splicing events, one at the nuclear interior and one at the nuclear periphery.

All together these results suggest that while EnhD is involved in the transcriptional activation of *MYC* in the nuclear interior, the OSE is, probably primarily involved in tethering the active *MYC* alleles to the NPC when it is located in the proximity of the nuclear periphery.

### 4.1.3 *MYC* gating increases the cytoplasmic concentration of *MYC* transcripts

To assess the possible function of the recruitment of *MYC* to NPCs, we analyzed by qRT-PCR the total and nascent amount of *MYC* transcripts in both HCT116 and HCEC. Surprisingly, the data showed that, as expected, the total *MYC* mRNA accumulation was higher in HCT116 than HCECs. However, the level of nascent transcripts, as analyzed by 5'-ethynyl uridine (EU)-labeling and pull-down, was lower in cancer cells than in the normal counterparts, suggesting that the higher *MYC* expression in HCT116 is not caused by increased transcription compared to HCEC.

To explore the potential contribution of the gating process to higher *MYC* levels, first we explored if proximity to NPCs in HCT116 cells would accelerate the nuclear export rate of *MYC* transcripts compared to HCECs. To examine the nuclear export rate of newly synthesized *MYC* transcripts, before extracting the EU-labeled newly synthesized RNA, cells were fractionated to cytoplasmic and nuclear fractions, and the two fractions were analyzed separately over a time course of 1 hour (Figure 3A). We observed that the ratio between cytoplasmic and nuclear fractions of newly synthesized *MYC* transcripts at 1 hour after chase was approximately 5-fold higher in HCT116 cells compared to HCECs, indicating an increased nuclear export rate in the cancer cells (Figure 3B).

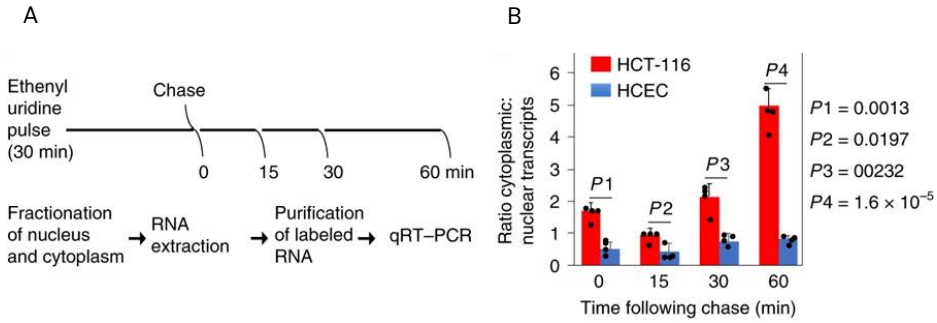


**Figure 3: *MYC* gating increases the export rate of *MYC* transcripts in the cytoplasm of cancer cells**

*(A) shows the scheme of the pulse-chased experiment used to calculate the export rate of MYC transcripts from the nucleus to the cytoplasm: following incubation with ethenyl-uridine for 30 minutes, the cells were chased for 0, 15, 30, or 60 minutes, as indicated in (A). Following harvesting, the nuclear and cytoplasmic compartments were fractionated, and the newly synthesized EU-labeled RNA extracted and purified. (B) shows the results presented as ratios between newly synthesized MYC mRNA detected in the cytoplasm and nucleus, respectively. HCT116 = colon cancer cell line; HCEC = normal colon epithelial cells. The data clearly indicates that MYC-gating increases the export rate in cancer cells, while this principle is not active in normal cells.*

*Modified from: WNT signaling and AHCTF1 promote oncogenic MYC expression through super-enhancer-mediated gene gating. Nat Genet 51, 1723–1731 (2019), https://doi.org/10.1038/s41588-019-0535-3.*

To explore how an increased export rate in cancer cells would contribute to higher total cellular *MYC* levels, we analyzed the total, cytoplasmic and nuclear decay of *MYC* mRNAs by blocking the transcription elongation by Actinomycin D treatment. Importantly, while we observed no difference in the decay of *MYC* mRNAs between HCT116 and HCEC, in both cell types the results showed a much faster decay in the nucleus than in the cytoplasm.

In light of these observations, we hypothesized that the OSE-mediated gating of *MYC* in cancer cells results in increased nuclear export rate of *MYC* transcripts to the cytoplasm, escaping in this way the faster decay in the nucleus and allowing the accumulation of *MYC* mRNAs in the cytoplasm. Importantly, modeling experiments confirmed that the increased nuclear export rate of *MYC* transcripts over time can solely explain the difference in total *MYC* mRNA levels between HCT116 cells and HCECs.

### 4.1.4 AHCTF1 and β-catenin regulate the OSE-mediated gating of *MYC*

AHCTF1 (or ELYS) is an essential nucleoporin for re-assembling the NPCs after cell division and the only one with a specific DNA-binding domain[76,90]. With these properties, we hypotesized that AHCTF1 might be involved in the OSE-mediated recruitment of *MYC* to the NPCs.

To verify this possibility, we used siRNA strategy to knockdown AHCTF1 and we observed that both the export rate of newly synthesized *MYC* transcript and the binding and well as proximity between the OSE and NUP133 were drastically reduced in the absence of AHCTF1, while the total transcriptional rate and the OSE-*MYC* polarized orientation at the nuclear periphery (c-value) were nearly unchanged. These data led us to conclude that AHCTF1 is needed to anchor the OSE to the NPC ring.

To investigate further the potential factors regulating the gating of *MYC*, we considered that WNT signaling is one of the pathways generally upregulated in cancer cells and responsible for MYC upregulation. Moreover, HCT116 cells have an autocrine WNT3a loop. In addition, we have also observed the presence of 4 TCF4-binding sites in the OSE in close proximity to the NUP133 binding site. To explore the hypothesis that WNT could be involved in the regulation of *MYC* gating, we treated HCT116 cells with BC21, a drug that specifically targets the WNT canonical pathway by inhibiting β-catenin-TCF4 complex formation[304]. Using ISPLA, we showed that not only β-catenin and TCF4 are in close proximity with AHCTF1, but also that such proximity is counteracted by BC21 treatment. Co-immunoprecipitation data have confirmed a similar trend, but with BC21 affecting the TCF4-AHCTF1 complex, indicating that β-catenin might mediate the indirect binding of AHCTF1 and TCF4. The β-catenin-mediated AHCTF1 binding was, moreover, essential to anchor the OSE to NUP133, as shown by the reduced ChrISP proximity signal between NUP133 and the OSE upon BC21 treatment.

We have also analyzed the binding patterns of these factors on the most prominent of the 4 TCF-binding sites, WRE520: β-catenin, AHCTF1, and NUP133 all have a strong binding to this site according to ChIP results, and their bindings are reduced upon BC21 treatment. After verifying that BC21 treatment affects both the newly synthesized *MYC* mRNA nuclear export rate to the cytoplasm and total *MYC* mRNA accumulation, but not the transcription rate of *MYC*, we conclude that β-catenin, and therefore WNT signaling, regulate *MYC* levels post-transcriptionally, *via* mediating the biding of AHCTF1 to the OSE and, in doing so, controlling the OSE-mediating gating of active *MYC* to the NPCs (Figure 4).

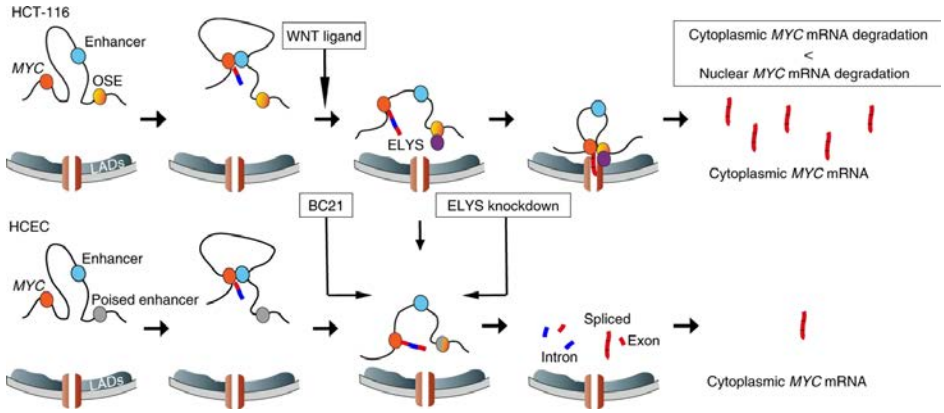


**Figure 4: OSE-mediating *MYC* gating**

*In the presence of an oncogenic super-enhancer (OSE) and an active WNT pathway, as represented in HCT116 colon cancer cells, the MYC gene is gated to the nuclear pore complex at the nuclear periphery in an ELYS-dependent mechanism. In this way, the transcripts can escape the faster nuclear degradation rate and accumulate in the cytoplasm by fast export through the nuclear pores. Moreover, there is a "division of labor" between gating and non-gating enhancers, the latter of which tends to increase MYC transcription in the nuclear interior. In the absence of the OSE, like in the normal HCEC cells, or in the presence of BC21 or ELYS knockdown, the export rate is drastically reduced and MYC transcripts are subjected to the more rapid nuclear decay.*

*From: WNT signaling and AHCTF1 promote oncogenic MYC expression through super-enhancer-mediated gene gating. Nat Genet 51, 1723–1731 (2019), https://doi.org/10.1038/s41588-019-0535-3*

## 4.2 A non-canonical CTCF function regulates the OSE-mediated *MYC* gating (Paper II)

### 4.2.1 The OSE possesses a CTCF-binding site

To investigate further the molecular mechanism of the gating of *MYC* transcripts in HCT116 cells, we focused our attention on CTCF, the master regulator of 3D-nuclear architecture that is involved in enhancer-promoter loop formation, the regulation of *MYC* levels[250,305] and the recruitment of circadian genes to the lamina[43]. Indeed, the OSE has a CTCF binding site (CTCFBS) located within the *CCAT1* eRNA - a region that showed high-frequency interaction with the *MYC* gene in Nodewalk assays.
To study the role of CTCF in the gating mechanism, we mutated the specific binding site by CRISPR-Cas9 technology and chose 2 clones, D3 and E4, for the follow-up experiments. We validated the efficiency of the mutations by CTCF ChIP, and verified the absence of any off-targets by bioinformatic tools and CTCF ChIP-seq analyses.

### 4.2.2 CTCF binding to the OSE confers a growth advantage to HCT116 cells

To evaluate the effect of the mutated CTCFBS on cell proliferation, we co-cultured WT and D3 or WT and E4 cells for 2 weeks, and documented the growth advantage of WT cells compared to either of the 2 mutated clones. Taking into consideration our previous results and the role of WNT-signaling in the gating process, we examined any possible

connection between the CTCF binding site and β-catenin/TCF4 binding at the OSE. Treating the cells with BC21 thus drastically reduced the growth advantage of WT cells over the mutant clones, indicating that CTCF binding to the OSE in cancer cells gives a growth advantage, likely by providing a platform to integrate the effects of the canonical WNT-signaling pathway.

### 4.2.3 CTCF regulates the nuclear export rate of *MYC* transcripts

To assess if the removal of CTCF from the OSE would interfere with the OSE-mediated gating of *MYC* and the nuclear export of *MYC* transcripts, and how these effects would be connected with the role of β-catenin in the gating process, we repeated in the D3 and E4 clones the export assay in the presence or absence of BC21. We observed an intense reduction of the nuclear export rate of *MYC* transcripts in the two mutant clones compared to the WT cells. However, BC21 did not reduce further the export rate of *MYC* transcripts in the mutant clones, indicating that the effect of WNT was mediated by the CTCFBS. In line with the effect of gating on total *MYC* mRNA accumulation, the total amount of MYC transcripts was also reduced in the mutant clones, but not the transcriptional rate of *MYC,* as measured by newly synthesized RNA quantification.

We repeated the same experiments with similar results also for the *FAM49B* gene that, according to Nodewalk analysis, is also an interactor of the OSE and its protein is functionally connected with MYC.

In summary, we conclude that CTCF binding to the OSE confers a proliferative advantage by increasing the nuclear export rate of *MYC* transcripts.

### 4.2.4 Both CTCF and β-catenin are needed to efficiently recruit AHCTF1 to the OSE

To understand further the mechanism of *MYC* gating, we scrutinized in further details the relationship between CTCF and the previously identified players in this process: β-catenin, AHCTF1, and NUP133.

By using co-immunoprecipitation analysis, we could conclude that CTCF interacts with all three components, but we also observed a very high recovery of AHCTF1 in CTCF-CoIP, higher than the recovery of CTCF itself, indicating the possibility that a subpopulation of CTCF molecules interacted with oligomers of AHCTF1. Moreover, AHCTF1 interaction with CTCF seemed to be also influenced by β-catenin, since BC21 treatment reduced both the recovery of AHCTF1 in the CTCF-CoIP experiment and the binding of it to the CTCFBS, as quantified by ChIP.
This data suggests a possible collaboration between CTCF, β-catenin and AHCTF1 in the OSE-mediated gating of *MYC*.

To further explore whether the effects of the CTCFBS mutation on the recruitment of AHCTF1 to the OSE is caused by the lack of CTCF binding to the OSE, we knocked down CTCF by siRNA strategy and analyzed the binding of AHCTF1 to the CTCFBS. Similarly to the situation in the two mutant clones, D3 and E4, we observed reduced binding of AHCTF1 to the OSE, suggesting that indeed it was CTCF that collaborated with β-catenin for the recruitment of AHCTF1 to the OSE.

To explore the role of AHCTF1 in the distribution of the OSE and *MYC* within the nuclear architecture, we performed 3D-DNA-FISH analyses of these regions in siAHCTF1-treated HCT116 cells. The results confirmed that the OSE requires AHCTF1 to complete its last 0,7μm travel toward the nuclear periphery. Interestingly, CTCF-AHCTF1 and CTCF-NUP133 ISPLA signals indicated that the proximity between these factors peaked around 1μm from the periphery – raising the question whether AHCTF1 and NUP133 might be loaded onto the OSE at this sub-nuclear position to facilitate its recruitment to NPCs. Moreover, this process required β-catenin, since BC21 treatment reduced the ISPLA signals. Finally, as controls we showed that BC21 didn't impact the recruitment of CTCF on the OSE CTCFBS, neither did the mutated CTCFBS in D3, or E4 affect the binding of TCF4 and β-catenin to the OSE.

All these data strongly advise for a model in which both CTCF and β-catenin binding to the OSE is needed to efficiently recruit and stabilize AHCTF1 on the OSE to allow the latter to reach the NPC at the nuclear periphery.

Interestingly, although in the D3 and E4 clones, both the number of OSE and *MYC* alleles at the periphery and the coordination between their recruitment to the periphery (indicated by the c value approaching 0) are reduced, Nodewalk analysis showed no difference in OSE-*MYC* interaction frequencies, excluding the possibility that CTCF is directly involved in the loop formation. Indeed, several interaction points were discovered between *MYC* and the OSE region outside the context of the CTCFBS, potentially mediated by factors other than CTCF. CTCF might thus have only an indirect role in mediating OSE-*MYC* proximity by enabling their recruitment to the more crowded environment of the nuclear periphery.

### 4.2.5 *CCAT1* expression correlates with the recruitment of the OSE to peri-nuclear positions

Recent publications have shown that, in an engineered cell line with high *CCAT1* expression, *CCAT1* seems to mediate the OSE-*MYC* interactions, in part by recruiting more CTCF to the OSE, and facilitates *MYC* expression. In our HCT116 model system, we observed only the presence of the long form of CCAT1, CCAT1-L, which is localized exclusively in the nucleus.

As opposed to previous findings in an engineered cell line[262], a combination of *CCAT1* RNA-FISH and 3D-DNA-FISH analyses with small probes covering the OSE and *MYC* loci was not able to show any correlation between *CCAT1* expression, which peaks proximal to but not precisely at the nuclear periphery, and the OSE-*MYC* proximity, peaking around 0,6μm from the nuclear periphery and at the lamina. On the contrary, we have found a correlation between high *CCAT1* RNA FISH signals and the proximity of the OSE allele DNA-FISH signals to perinuclear positions, approximately 1μm from the periphery, indicating a possible function of *CCAT1* in giving directionality of the OSE movement towards peri-nuclear positions.

As *CCAT1* transcription and expression required a functional CTCFBS, as shown by the qRT-PCR in WT and CTCFBS-mutant clones of total and newly synthesized transcripts, we explored whether or not WNT-signaling, which exerts its effect in collaboration with the CTCFBS, plays a role in *CCAT1* expression. Indeed, BC21 treatment reduced *CCAT1* expression, but only in WT cells and not in the CTCFBS-mutant clones, indicating that its effect impinges on the CTCFBS.

Data accumulated to this point supported a model in that WNT-and CTCF-induced *CCAT1* expression facilitates the recruitment of the OSE to perinuclear positions (0,7-1,5μm from the nuclear periphery), where the OSE and *MYC* regions acquire polarized orientation with the OSE being closer to the periphery, and CTCF together with β-catenin facilitate the loading of AHCTF1 and NUP133 onto the OSE – factors that are necessary to recruit the OSE-*MYC* complex to NPCs and the gating of *MYC* transcripts. At the same time, the *CCAT1* expression is drastically reduced at the lamina, indicating that it does not participate in the final steps of the anchoring of OSE-*MYC* complex to NPCs. This process is reduced or absent in HCECs or in mutant clones where CTCF is unable to bind efficiently the correspondent CTCFBS within the OSE, or in case of WNT-pathway inhibition (Figure 5).

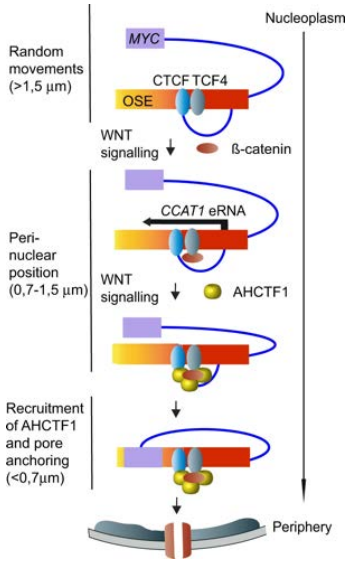


**Figure 5: Model showing the role of the CTCFBS in the *MYC*-gating mechanism**

*Following random and possibly also directed movements in the interior of the nucleus, the OSE reaches a position near the nuclear periphery. In parallel, WNT induces the expression of CCAT1, in concert with the CTCF binding site neighboring the TCF4 binding site. At this position AHCTF1 is recruited to the OSE in a manner mediated by both CTCF and the β-catenin/TCF4 complex. This step is essential for the ability of the OSE to reach the nuclear pore. During this transit at a position much closer to the nuclear periphery (<0,7 μm), CCAT1 expression is reduced, concomitant with the juxtaposition of the MYC and OSE regions. The entire process requires a functional CTCF binding site within the intron of the CCAT1 gene.*

*Modified from: Canonical WNT signaling-dependent gating of MYC requires a noncanonical CTCF function at a distal binding site. Nat Commun 13, 204 (2022), https://doi.org/10.1038/s41467-021-27868-3.*

## *4.3 CCAT1* antagonizes the OSE-mediated gating of *MYC* (Paper III)

### 4.3.1 Knockdown of *CCAT1* reduces *MYC* transcription but increases the export rate

Since it was difficult to conclude if the effects of *CCAT1* expression on the gating mechanism in the CTCFBS-mutated clones were direct or a consequence of the lack of CTCF binding, we explored the effects of knocking down *CCAT1* by siRNA for 72h.

As expected, according to other reports in the literature, *MYC* expression was reduced. We have also observed a reduced transcriptional rate, quantified by qRT-PCR of EU-labeled newly synthesized *MYC* mRNAs. Interestingly, also *CCAT1* nascent RNA was reduced, suggesting a dual effect of the siRNA: degrading the target by RNA interference mechanism and affecting its transcription rate.

What was unexpected, however, was that the nuclear export rate of *MYC* transcripts was drastically increased in the *CCAT1* knock-down cells (Figure 6A), contrasting the previously suggested role of *CCAT1* in the gene gating process.
These results thus pointed to a complex function of the *CCAT1* eRNA: it amplified the expression of *MYC*, but at the same time, antagonized the gating mechanism.

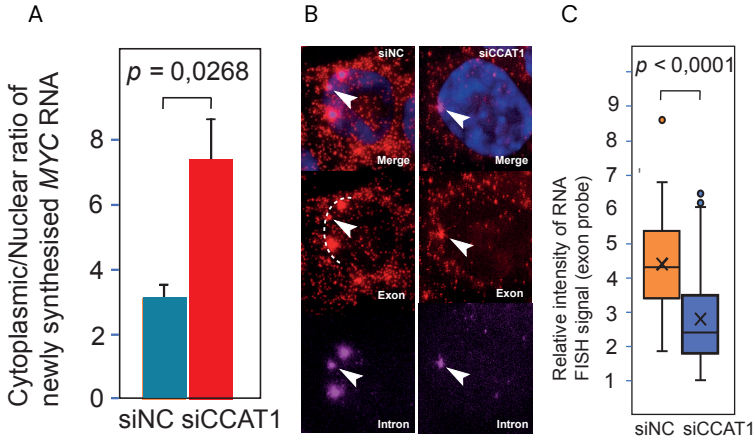


**Figure 6: *CCAT1* interferes with gated *MYC* mRNA export by promoting *MYC* transcription**

*CCAT1 interferes with gated MYC mRNA export by promoting MYC transcription. The knockdown of CCAT1 expression by siRNA significantly increases the nuclear export rate of MYC transcripts, as quantified by the newly synthesized RNA export assay (A), while reducing MYC transcription, as exemplified by RNA-FISH images in (B), quantified in (C). Moreover, is possible to visualise in (B) that CCAT1 promotes the formation of transcriptional condensates at the perinuclear position. Following knockdown of CCAT1 expression these condensates disappear while increasing the frequency of less actively transcribed MYC alleles on the nuclear periphery (arrow). In (B), MYC exon= red, MYC intron= purple DAPI= blue*

### 4.3.2 *CCAT1* promotes the formation of *MYC* transcript condensates

To verify how the expression and distribution of *MYC* transcripts within the nuclear space was affected by the *siCCAT1* treatment, we performed an RNA-FISH with specific probes for the *CCAT1* eRNA, *MYC* exons, and *MYC* introns. Taking into consideration our previous result showing that *CCAT1* expression signal was almost completely excluded within 0,7μm from the nuclear periphery, we expected to find two different populations of *MYC* alleles: one *CCAT1* eRNA-dependent population in the interior of the nucleus and one *CCAT1* eRNA-independent group at the nuclear periphery.

We indeed found that in control cells *MYC* exon RNA-FISH signals were much stronger at the interior of the nucleus, whereas when *CCAT1* expression was attenuated, the remained *MYC* exon signals were mainly located at the nuclear periphery (Figure 6B). Of interest, *CCAT1* and *MYC* RNA FISH signals were often not in close proximity, indicating an indirect role of *CCAT1* eRNA in *MYC* transcription, potentially by sponging transcriptional repressor miRNAs.

The *MYC* exon signals, moreover, appear to form agglomerates (Figure 6B), the number and size of which are *CCAT1* expression dependent (Figure 6C), indicating the possibility of this ncRNA of being a seed for the formation of transcriptional condensates. Interestingly, the *MYC* transcript clusters devoid of intron signal often protruded towards the nuclear periphery, suggesting that a part of them could be a stock for a following gene-gating-independent nuclear export. It was also possible to observe cells with all three *MYC* alleles active, two distant and one proximal to the periphery - a scenario that could be explained by the co-presence in the same cells of both gated and not gated nuclear export mechanisms of *MYC* mRNAs (Figure 6B, left panel).

Summarizing these unexpected results, we concluded the *CCAT1* eRNA functioned as a switch between gating-dependent and independent export mechanisms: in the nuclear interior, its expression promoted *MYC* transcription and the formation of transcriptional condensates that are not-subjected to the gating mechanism, while its absence allowed the OSE-*MYC* complex to reach the nuclear periphery and participate in gene gating.

### 4.3.3 *CCAT1* eRNA prevents OSE and *MYC* from reaching the nuclear periphery by promoting transcriptional elongation

To check the role of the *CCAT1* eRNA in the movement of the OSE and *MYC* regions toward the nuclear periphery, we performed the 3D-DNA-FISH analysis in the si*CCAT1*-treated cells. We thus observed an increased accumulation of both OSE and *MYC* alleles at the nuclear periphery in the absence of *CCAT1* eRNA (Figure 7A). Moreover, we noticed an increased number of OSE and *MYC* alleles per cell within 0,6μm from the periphery in the treated cells, indicating that a subpopulation of cells was more responsive to *CCAT1* expression.

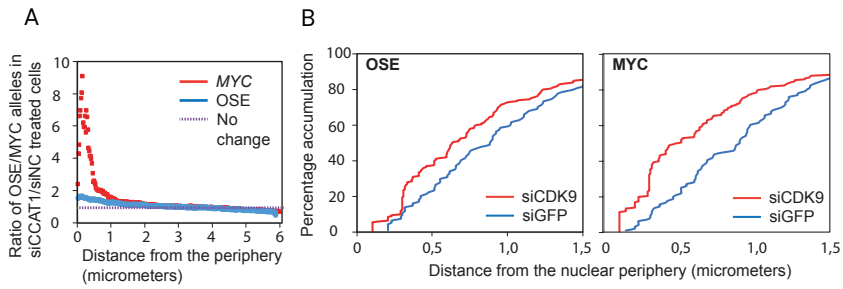


**Figure 7: Blocking transcription elongation increases the migration of the OSE and *MYC* alleles to the nuclear periphery**

*The knockdown of CCAT1 (A) and CDK9 (B) expression increases the accumulation of MYC and OSE alleles at the nuclear periphery. Talen together, these results indicate that the CCAT1 eRNA impedes gating of MYC by antagonizing the inhibition of transcriptional elongation.*

A previous report of our group showed that a pan-cyclin-dependent kinase (CDK) inhibitor, Flavopiridol, facilitated the movements of circadian genes to the nuclear periphery[43]. As CDK9 is a component of the positive transcription elongation factor complex, P-TEFb, which is responsible for inducing transcriptional elongation by phosphorylating and inhibiting the negative transcription elongation factors and by phosphorylating the RNA Pol II CTD on serin-2. Knockdown of CDK9 by siRNA treatment showed, similarly to what was observed in si*CCAT1*-treated cells, a shift towards the nuclear periphery of both *MYC* and OSE regions (Figure 7B). Although we don't have yet data showing an influence of the export rate in *siCDK9*-treated cells, this observation suggests that a block in transcription elongation promotes the migration of the *MYC* and OSE loci to the periphery. Hence, the *CCAT1* eRNA might work by promoting transcriptional elongation in the nuclear interior and, in doing that, inhibiting the gating process.

In summary, we have observed that the *CCAT1* eRNA works as a switch between gating-dependent and –independent nuclear export mechanisms. By increasing the transcriptional burst rate, with transcriptional condensate formation, and promoting transcriptional elongation in the nuclear interior, the *CCAT1* eRNA thus likely prevents the migration of *MYC* and OSE loci to the nuclear periphery, an essential event for the gating-mechanism, and, in this way, counterbalances the gating-dependent nuclear export rate of *MYC* transcripts.

# 5 Discussion

This thesis describes the discovery that gene gating, previously identified only in fungus and lower animals, exists also in human cells and, by inference, in other mammals. Importantly, the results also document that this principle contributes to the pathological proliferative advantage of cancer cells *via* the gating of active *MYC* alleles to the nuclear pores in a manner mediated by a distal oncogenic super-enhancer in HCT116 colon cancer cells.

The *MYC* locus is surrounded by enhancer regions that are activated in lineage-specific manners. While these are normally under strict control to prevent expression overshoot, the emergence of oncogenic super-enhancers changes the balance. In this thesis, I have focused on a region >500kb distal to *MYC* that is not active in normal cell counterparts, but has acquired oncogenic super-enhancer properties in colon cancer cells.

## 5.1 Stochastic versus directed movement

The heterogeneity of the number of *MYC* alleles at the nuclear periphery shown in Paper I suggests that the gating is a dynamic feature with a constant recycling of gated alleles to and from the interior of the nucleus. The results shown in Papers I and II suggest that the movement to the nuclear pore likely is at least partially directed. Although this is against current dogmas, there are precedents for such a principle[196,306]. In addition, our lab has previously shown that the synchronization of circadian gene expression involves rhythmic movements of circadian genes to and from the nuclear periphery[43], although the contribution of directed *vs* stochastic elements to these movements have not been directly visualized. However, this report was the fundament for exploring the possibility of circadian gene gating - that is, circadian genes reach the nuclear pore in active states to enhance the amplitude of cytoplasmic mRNA oscillations templated by circadian genes. This idea was supported by the observations that *MYC* expression can be under circadian control[285–287] and the physical interaction between OSE and *MYC* is under circadian control (Vestlund et al, unpublished). Moreover, it has been shown that nuclear actin and myosin play a role in the directed movement of chromatin regions away from the nuclear periphery in response to transcriptional activation[307]. There was therefore support in the literature that the OSE would reach the nuclear pores, at least partially, independent of diffusion.

The most surprising result of Paper I was that the OSE appears to be dragging the active *MYC* gene towards the nuclear pore, and not the other way around, and that the entire process was dependent on the β-catenin function. This information put the canonical WNT signaling pathway in control of the gating of *MYC*. Subsequent analyses revealed part of the underlying mechanisms, which involve the recruitment of β-catenin and the nucleoporin AHCTF1 to the OSE, but not to *MYC*. Very importantly, it could be

documented that the movement of the OSE towards the final 0,7μm stretch to the nuclear pore required the AHCTF1 function (Paper II). It thus appears that at least a portion of the chromatin movement of the OSE from an interior position to the nuclear periphery/pores might be directed and under control of WNT-dependent deposition of AHCTF1 to the OSE.

## 5.2 The non-canonical role of CTCF in the gating process: the AHCTF1-β-catenin connection

Given the dogma that CTCF binding sites are at the bases of chromatin loops, it was surprising that the mutant OSE allele was able to maintain strong physical interactions with *MYC* despite its inability to interact with CTCF (Paper II). This is in line, however, with more recent work documenting that enhancer-promoter interactions can be unaffected by knockdown of CTCF expression[37,38], thus questioning the generality of the CTCF-chromatin looping link.

Another surprise was that when the WNT signaling pathway converging on the OSE has been abrogated, it has resulted in loss of several features including facilitated nuclear export of *MYC* mRNAs and loss of *CCAT1* eRNA expression. This observation prompted an analysis of the link between β-catenin and CTCF. Paper I shows that a drug, BC21, that inhibits the interaction between β-catenin and TCF4, results in reduction of nuclear export of *MYC* mRNAs, hinting at the possibility that β-catenin and CTCF complemented each other. A complicating factor was that the interaction between CTCF and β-catenin was very weak, suggesting the involvement of yet another factor. We screened a number of nucleoporins for their ability to interact with CTCF and found that AHCTF1 was the most promising candidate. Indeed, knockdown of its expression resulted not only in reduction of nuclear export rates of *MYC* mRNAs, but also inhibited the ability of the OSE to reach the nuclear periphery/pore. The interpretation that AHCTF1 interacts with the CTCF binding site has been confirmed by knocking down CTCF expression and examining the mutant HCT116 cells by ChIP (Paper II). These observations have not, however, explained the link to β-catenin until we have shown that the binding of AHCTF1 to the OSE requires the β-catenin-TCF4 interaction (Paper II).

The role of AHCTF1 in gene gating highlights another fundamental contribution of this study: nucleoporins as chromatin organizers. Considering the original gene gating theory from Blobel[194], in which NPCs were essential for defining 3D nuclear coordinates for specific chromatin loci in yeast[194], and the hypothesis by which NPCs are one of the most ancient forms of nuclear envelope[1], this alternative function of nucleoporins could have been essential during evolution to establish and/or maintain nuclear compartments before the appearance of other structures, such as LADs.

NPCs could also play a role in maintaining specialized compartments, different from LADs, at the nuclear periphery: Smith and collegues[308] have observed the presence in

mice of what they called KODs, H3K9me2-Only Domains, rich in tissue-specific H3K9me2-enriched enhancers with very low lamin B interactions. These enhancers possess inactive or poised markers, and they are located in A-type of open chromatin compartments. According to the authors, the KODs could facilitate enhancer-promoter interactions and regulate spatiotemporal transcriptional programs. AHCTF1 and other nucleoporins have been shown to promote chromatin decompaction in Drosophila[309], indicating a possible role in the formation of the KODs domains. That could also explain why the proximity between *MYC* and the OSE is directly proportional to the distance to the periphery, as shown in Paper I.

Of note, the model cell line, HCT116, harbors an autocrine WNT signaling loop, and its mutated β-catenin cannot be degraded[288,289], to cause a constant activation of the pathway. New evidence has shown that β-catenin targets, other than being cell-specific, are also time-dependent[310], opening the question if the OSE is one of the few targets which β-catenin binds continuously to, or if it is an early or late target, and consequently if the gene-gating mechanism would be always activated or not. Another possibility is that CTCF would be first needed to bind the OSE to keep its immediate environment free from repressive marks[311]. This in turn would promote the formation of the β-catenin/TCF4 complex, which in turn would collaborate with CTCF to recruit AHCTF1 to the OSE. Since the WNT pathway is frequently altered in several types of cancer[228], the emergence of gene gating could be a consequence of prolongated exposure to nuclear β-catenin.

The overall view of the gene gating principle is based on extraction procedures to represent merely snapshots of key events and are therefore likely to be restricted to cell subpopulations. Such dynamics are poorly understood, although our lab has observed that the processing of lamin A might have a role to play. Thus, by inhibiting Farnesylation of prelamin using Lonafarnib and hence rendering it inaccessible for its further processing within the nuclear membrane, the binding of AHCTF1 to the OSE was increased several-fold. Since both AHCTF1 and prelamin A bind to the OSE (Lim et al, preliminary results) we envisage that prelamin A is involved in the formation of the AHCTF1-containing complex at the OSE-specific CTCF binding site. According to this reasoning, its processing might disrupt the AHCTF1-OSE interaction, thus enabling the gated *MYC* alleles to return to the nuclear interior.

## 5.3 The *CCAT1* eRNA paradox

The role of *CCAT1* in gating was initially considered straightforward due to the results obtained using HCT116 cells lacking a functional CTCF binding site within the *CCAT1* intron. Thus, the mutant cells displayed both a loss of gating and loss of WNT activated *CCAT1* eRNA transcription, as shown in Paper II, suggesting a clear connection. However, since *CCAT1* eRNA RNA FISH signal drops around 1μm from the nuclear periphery, any

effect on the gating process would occur at a distance from the nuclear periphery and thus be indirect. While knocking down *CCAT1* eRNA expression/transcription produced the expected correlation with a concomitant loss of *MYC* transcription, the results of the nuclear export assays in Paper III complicated this interpretation. Thus, the several-fold higher rate of nuclear export of *MYC* mRNAs in cells with attenuated *CCAT1* eRNA expression was hardly consistent with promoting the gating process.

An important clue to this enigma was provided by the RNA-FISH analyses showing that *MYC* transcription was most prominent at a distance from the nuclear periphery in control cells. As opposed to these large clusters of *MYC* mRNAs, probably representing ongoing re-initiation of transcription, in the si*CCAT1*-treated cells (Paper III) we have observed considerably weaker *MYC* RNA-FISH signals that localized almost exclusively at the nuclear periphery. This observation gave rise to the idea that the increased rate of *MYC* transcription in the presence of *CCAT1* expression would counteract the migration of the *MYC* alleles to the nuclear pores, which are surrounded by repressive heterochromatic structures enriched in LMNA/B-binding chromatin. Since *MYC* transcription is indeed regulated at its elongation[312], we considered the possibility that the gating mechanism would benefit from a delayed release of this elongation block, i.e. taking place only when approaching the nuclear periphery. In line with this supposition, we have recollected that the inhibition of transcriptional elongation by Flavopiridol has speeded up the mobility of circadian genes to reach the nuclear periphery[43]. It is noteworthy therefore that inhibition of the expression of CDK9, which is a component of the P-TEFb complex that alleviates inhibition of transcriptional elongation by phosphorylating, among other factors, Serine 2 of Pol II, has significantly increased the migration of *MYC* to the nuclear periphery. Experiments are ongoing to explore whether reversible inhibitors of the CDK9 function increase the nuclear export rates of *MYC* mRNAs.

Irrespective of these considerations, it is far from clear how the *CCAT1* eRNA is able to counteract *MYC* gating by increasing its transcriptional rate. Candidate possibilities include the ability of the *CCAT1* eRNA to sponge a subset of miRNAs targeting the stability of mRNAs encoding functions antagonistic to transcriptional elongation, for example. Alternatively, the ability of the *CCAT1* eRNA to interact with a range of protein factors, such as CHTOP that is a key export factor as well as AHCTF1 (BioGrid), might compete out those factors from intended targets, such as the primary *MYC* transcript, and thus impede *MYC* gating and facilitated nuclear export rates of derived mRNAs.

Even these speculative scenarios might prove too simplistic. Another unexpected outcome of Paper III is that the *CCAT1* eRNA regulates the expression of distal (>170kb downstream) *CCAT2* gene. Similar results were obtained for *PCAT1* (not shown), another non-coding RNA gene which is >150kb upstream of *CCAT1*. Both of these non-coding genes have been connected to cancer development[261,313,314]. It thus appears that the

*CCAT1* eRNA regulates the expression of other genes over large distances perhaps by promoting phase separations by analogy to other non-coding RNAs[4,315–317]. Very importantly, this effect of *CCAT1* eRNA is lost in the mutant HCT116 cells lacking a functional CTCF binding site in the first intron of the *CCAT1* gene. Hence, either the long-range function of *CCAT1* eRNA depends on the integrity of this CTCF binding site, and/or the mutant cells experienced a crisis during clonal expansion to disconnect its functional relationship with the *CCAT1-CCAT2-PCAT1-MYC* gene cluster. In support of the former interpretation, our lab has observed that CTCF quantitatively binds *CCAT1* eRNA (Nikosjkov et al, preliminary result), suggesting that the *CCAT1*-specific CTCF binding site keeps newly produced *CCAT1* eRNA close to its template. Since CTCF can form oligomers and simultaneously bind several proteins[22], such a complex might facilitate the formation of OSE-specific condensates or microspeckles that govern the transcription over large regions including that of the *CCAT1-CCAT2-PCAT1-MYC* cluster.

## 5.4 Is the gating of *MYC* in HCT116 cells a unique or a widespread phenomenon?

Focusing on a model system was essential to identify the gating principle and the involved players and their interactions. However, this approach suffers from the possibility that it is a highly specialized and uncommon process in other cell lines or cell types. Against this backdrop, we have observed that the facilitated mRNA export pathway typical of gating applies not only to the *MYC* and *FAM49B* gene products, but also to mRNAs produced from the key regulators themselves including *CTCF, CTNNB* and *AHCTF1* to mention a few, in HCT116 cells (Sumida et al, preliminary result). Ongoing experiments attempt to decipher genome-wide patterns of nuclear export rates using the SLAM-seq (thiol(SH)-linked alkylation for the metabolic sequencing)[318] technique. Moreover, preliminary results show that *MYC* mRNA export is facilitated in the MCF10A cell line similar to HCT116 cells (Sumida et al, preliminary result). Finally, and perhaps most importantly, our lab has preliminary results documenting that the OSE region is in direct physical proximity to the nucleoporin NUP133 when at the nuclear periphery in both cancer cells and invading endothelial cells in thin sections of breast cancer (Pei et al, preliminary result). This does not necessarily mean that the *MYC* gene is gated in such cells, although it fulfills one of the requirements. It will thus be important to pursue this link using breast cancer explants and examine if the nuclear export rate of *MYC* mRNAs is facilitated in breast cancer cells *ex vivo*. However, when combining this data with the results from the MCF10A cells, the picture emerging is in support of *MYC* being gated also in breast cancer.

## 5.5 Summary

The results from this thesis support the existence of a novel version of gene gating in mammals. It involves the trafficking of gated alleles from the interior of the nucleus to the nuclear pores to facilitate the export of derived mRNAs and their escape from the rapid nuclear decay kinetics. This principle entails several steps and we are just starting the scratch the surface. Although the results generated so far are limited in scope in that they are primarily based on a colon cancer cell line, HCT116, they nonetheless suggest that gene gating is a feature of other cancers, in particular breast cancer, as well. Important questions for the future include whether gene gating plays a role in cancer evolution or primarily functions to manifest a proliferative phenotype, and whether mammalian gene gating is only a pathological phenomenon or a normal principle hijacked by cancer cells. Perhaps the best indication that gene gating might occur also in normal cells is provided by the *IGH* gene, which has the task to produce enormous amounts of derived cytoplasmic mRNAs in plasma B cells. Indeed, the *IGH* locus is positioned at the nuclear periphery of such cells in the mouse, mediated by a 5'-flanked enhancer region[319].

# 6 Conclusions

The work of this thesis describes the discovery of the gene gating principle underlying pathological levels of *MYC* expression in human colon cancer cells, HCT116. This is the first time a gene gating mechanism, originally proposed in 1985[194], has been described in mammals.

The gating mechanism includes the following events (**Papers I** and **II**):

- CTCF and the β-catenin/TCF4 complex bind to sites that are close neighbors within the Oncogenic Super Enhancer region;
- By stochastic movements, the OSE reaches from an intra-nuclear position the perinuclear position, where AHCTF1 is recruited by the combined efforts of both CTCF and β-catenin/TCF4;
- Concomitantly, the proximity between *MYC* and the OSE is reduced directly proportional to their distances from the nuclear periphery;
- AHCTF1 promotes the recruitment of the OSE and *MYC* to the Nuclear Pore Complex;
- At the nuclear pore, the nuclear export of *MYC* transcripts is facilitated to enable their escape from the faster nuclear degradation rate compared to that in the cytoplasm
- The OSE is critical to this process, as its functional absence – such as in normal colon epithelial cells, or upon the attenuation of the binding of CTCF (siCTCF-treated cells, or D3 and E4 CTCFBS-mutated clones), β-catenin (BC21-treated cells) or AHCTF1 (siAHCTF1-treated cells) to the OSE - impedes the gating mechanism.

In **Paper III**, the role of the *CCAT1* eRNA in the gating process is elucidated further. While the *CCAT1* eRNA appears to promote the transcriptional elongation of *MYC*, possibly by forming condensates together with OSE-bound CTCF, it also impedes the migration of *MYC* to the nuclear pores and hence the nuclear export rate of *MYC* mRNA.

In summary, the work of this thesis has generated data that profoundly compound our understanding of how pathological levels of *MYC* expression are attained. It raises the possibility that two opposing mechanisms - one that increases *MYC* expression despite reduced transcription by gene gating and another that antagonizes gene gating but increases *MYC* transcription - balance the cancer cell's response to environmental cues.

# 7 Points of perspective

The gene gating model described in this thesis opens a new perspective on how a gene can be post-transcriptionally regulated in the context of the 3D nuclear architecture.

Although the mechanisms underlying the gating principle have been uncovered to a degree, several loose ends remain. For example, it will be important to explore what principle(s) drive the directed migration of the OSE and *MYC* to the nuclear periphery. Another fundamental question addresses whether or not the features regulating the inhibition of *MYC* transcription are directly linked to the migration process and whether such stalled RNA polymerase complexes are particularly sensitive to DNA damage. A possibility to be explored is whether the MYC protein is part of a feed-forward principle driving the gating of its own gene. Finally, it will be essential to uncover how the *CCAT1* eRNA is able to control the nuclear export pathways of the *MYC* mRNA and whether there is a link to DNA repair. Real-time live-cell imaging of chromatin movements in single cells, in response to rapidly acting drugs targeting actin polymerization, the elongation phase of transcription and aspects of the DNA repair machinery could help answer such questions.

A fundamental question addresses to what extent the gating process that we have uncovered in a model system applies generally. For example, can it occur in normal cells under the pressure of producing vast amounts of proteins, such for the *IgG* gene? How general will the gating principle be in different stages of cancer development and in different types of cancer? Preliminary evidence suggests the existence of *MYC* gating in breast cancer cells. This observation is in line with that the WNT pathway and MYC expression are overactive in many types of cancer[228,232], in particular in cancer stem-like cells[320]. More broadly and touched upon above, how many other genes are also gated? Again, preliminary results show that several of the key players in the gating process, such as *CTCF*, *AHCTF1* and *CTNNB* are themselves gated to suggest a network of interdependent functions. It is thus most pertinent to identify genome-wide patterns of gene gating and how these can be recruited to promote cancer evolution and metastasis. These considerations impinge on the question to what degree environmental cues regulate processes in the nuclear architecture, such as the circadian cycle[43]. Will crosstalk between signalling pathways promote or antagonize gene gating? Finally, would mechanic stimuli[199] be a factor to consider in the regulation of gene-specific gating mechanisms?

The MYC function has been historically very hard to target pharmaceutically due to the absence of pocket for high-affinity binding of inhibitors, even though new strategies have recently been developed to design drugs against the function of this common oncogene[321]. Nonetheless, the gene gating principle offers new perspectives in targeting pathological expression of MYC. Olaparib, a PARP inhibitor that indirectly inhibits

PARP1- CTCF interaction[43] and interferes with the insulator function of CTCF[42] by blocking its parylation, has been shown to be beneficial in breast and prostate cancer already[322].

A good amount of evidence accumulated by our group indicates that the side effect of Roscovitine[323], which otherwise has been shown to be a potentially powerful cancer drug[323], could be due to alterations in *MYC* expression due to the efficient inhibition of the gene gating mechanism by the drug. The availability of more potent derivates of Roscovitine and their use in targeting gene gating could help to revive the use of this drug in combination with other drugs targeting the WNT pathway, for example. The benefit of such a combinatorial approach might be to reduce the side effects while increasing their potency.

In summary, the work of this thesis opens many doors for designing new drugs or re-evaluating old ones!

# 8 Acknowledgements

This PhD has been a true Odyssey, but as for any epic adventure, it could not be completed without the help and the indispensable guidance and advice of precious and unforgettable people, some of whom have been with me during the entire journey and some have joined at different moments.

First and foremost, I have to thank my main supervisor, Professor **Anita Göndör**, who let the journey begin and that has guided me through all the stages. Her scientific acuteness and kindness have been precious to completing the work in this thesis and to survive to the hard times of a researcher's life. Your advice has been really important to develop not only my skills as a researcher but also to grow up as a person. I'm really glad to have been part of your group and have worked in such a scientifically rich environment. Although my personality has not always been very calm, you have always treated me with comprehension and respect. I will never forget our, sometimes, intense scientific discussion and the passion you have for this work and the field of 3D genome organization. Thanks for having always encouraged me to go on and not stop in front of problems and, at the same time teaching me that on some occasions it is good and needed to take other paths. Thanks for having trusted me with many responsibilities all over these years and having given me time and resources to explore my ideas and curiosities. The success of this journey is also thanks to your supervision and I hope that we will be able to collaborate also in the future. I wish you a very successful career and all the best for the future.

Another essential guide during this journey has been Professor **Rolf Ohlsson,** one of the most knowledgeable people I have ever met. Your passion and love for science are immense and your knowledge covers so many fields of science and more. Thanks for having me inspired every day since I started this adventure and for the great scientific discussion we have had. You have taught me two essential skills for a scientist, how to criticize your own and others' work and, perhaps more important, how to accept criticism. You have also shown me how hard work many times will later recompense you and how small achievements can, not only be fascinating but also be the base for many bigger ones. Thanks for your advice and prods and for helping me in these years to mature as a scientist. I hope our careers will meet again and I wish you a wonderful and happy future.

An immense thanks also to my co-supervisor Assistant Professor **Barbara Scholz**, a precious colleague, and a dear friend. Thanks for having helped me at any possible time, to have dealt with my sometimes stupid technical questions and irritable behavior, and to have guided me from the beginning in the research and lab life. You are one of the most calm and patient people I have ever met and your scrupulosity at work has been a precious lesson for me. You have been a precious friend in the needed times and I will

never forget the kindness and nice words you treated me with. I wish you a wonderful career and I hope we will remain friends for long.

A thanks also to my co-supervisor Professor **Galina Selivanova**, your small gesture and few words have always shown me your kindness and your support. Your scientific knowledge and the courses and seminar you have organized have been extremely important to develop my work and skills as a researcher. Thanks for all your support over these years, I wish you a brilliant future.

Another thank is to my co-supervisor Associate Professor **Andreas Lennartsson**, thanks for the small gestures in these years that have been enough for me to understand your interest in my progress and your support. Your advice has been useful and I hope in the future we could have more occasion to discuss science. I wish you an excellent career and the best in life.

Precious during these years has been also my mentor Associate Professor **Vicente Pelechano Garcia**: your scientific and career support has been really precious to take important decisions during my journey. You are one of those people who is enough to talk a few times to get inspired and having enough advice for a long time. Your kindness and acumen have guided me in many phases of this journey. Thanks for all your support and I hope you could continue mentoring me also in the future. I wish you a really brilliant career and life.

During these years many people have alternated in the group and all of them have been precious colleagues and dear friends. The success of this work is also thanks to their great work and help.

I will start to thank my current colleagues that have been really precious to help me in finalizing the work needed for this thesis.

Thanks **Jia Pei,** in these last few years you have been such an important friend and colleague, always ready to help and talk doesn't matter what thousands of things you will have to do at the same time. You are such a kind, generous, and resourceful person and a skilled and smart researcher. You have always nice words, and precious advice for work and life and you always dedicate your time to others. Your dedication to family and work was inspiring. Always believe in your skills and I'm sure you will have a successful career. **Felipe,** thanks for your precious help with experiments and for always being available to debate scientific topics or scrutinize texts and ideas. Your determination and clearness in your work and life have been inspiring. You are probably the "tidiest" researcher I have ever met; a skill I hope I can slowly inherit from you. Don't stop to believe in yourself and I'm sure you will have an excellent career. **Andrej,** you are a special person, it's impossible to make you angry or sad. You are always happy and that helped me many times during the hardest times at work. I'm sorry if sometimes I have

been not the best of me with you, but I have to really thank you for your patience. Don't stop to make your jokes, they have always helped to make intense moments more relaxed. Your broad knowledge and information about the world have always amazed me. Thanks for all your help in these years. **Narsis**, thanks to having been always available to discuss statistics and mathematical modeling, you are so knowledgeable in a such complicate field as mathematics. Your passion for research and the field is visible when talking with you. I wish you a really successful career. **Wei**, **Ayushi** and **Chen-Ying**, you are all very new in the lab but I could already appreciate your kindness during these intense months I was writing this thesis. Thanks for all the help and I wish you the best of luck in your time in the group.

A special thank is also to previous colleagues that have anyhow left an important mark on the work presented in this thesis and helped in my development as a scientist.

Thanks to **Ilias**, my brother in science, we started the journey together and lived happy and difficult moments at work and in life next to each other. You have been always there for help and I consider myself lucky to have been your colleague in these years. I miss our scientific debate and sometimes our divergence that, anyway, helped me to grow as a scientist. You are a strong person and I'm sure you will have a successful career. A special thanks to **Carolina**, I have always been impressed by your calmness and your smile also during hard times. You have been a really precious friend and important colleague. Your smile is very contagious and thanks for your kindness with words and small gestures. Thanks for have been always there with advice and for any type of help, at work and in life. Thanks for having guided me in the labs and for your support in these years anytime I was in need. You are a very strong and resourceful person as your new career achievements have already shown you. Your love for the family is inspiring, don't stop to reach your dreams and I wish you a lovely and wonderful life. Thanks, **Marta**, your rigorousness has been precious at the begging of my journey to set up a good researcher mindset. You have been always so kind and available to help at any moment also after your retirement. Your joy for life and your passion for the work are very admirable. Your help in managing the lab and your advice have been very precious. I wish you all the best. **Johanna**, your calmness and kindness in any situation are impressive. You are a very smart and resourceful person, a very skilled scientistic and knowledgeable clinician. Your determination is strong and will guide you in any career you would like to embark on. I was so amazed by you to persuade, very successfully, with your medical background, such a technical and molecular biology PhD, showing your interest in science as all. You will become an excellent clinician and also an amazing researcher if you would like. The best of luck for your future. **Deeksha,** you have always been so kind and patient trying your best to involve me as much as possible in your bioinformatics work. You have always been available for advice and help and I really thank you a lot for that, you work has been very precious. You are a strong person and

I'm, sure you will have a brilliant career. **Anna**, you are such a multitasking and resourceful person, I always have admired you for having so many skills in different fields: research, writing, networking and having such a happy family. It is impressive how you could combine so successfully your PhD with your family life and other hobbies. Your joy in life has always been contagious. Thanks for all your advice, help at work and nice barbecue in your beautiful home. **Rashid** and **Shuangyang,** thanks for your bioinformatics work behind the scene. A special thanks to Rashid for having done your best to introduce me to the programming world and for your help any time it was needed. **Honglei,** thanks for all the help in the years we have worked together. You are a skilled scientist and a very lovable person. I have always been impressed by how you manage to combine family and work life. You deserve the best in life and I wish that for you. **Feri**, you have been the first person to introduce me in the lab and in many methodologies used in this work. I still remember your "microscope exam" that you let me doing it twice before to be able to use it on my own. If I have decided later on to become much more skilled in microscopy is also probably thanks to that. You are a very strong and resourceful person and life is showing you what was already clear for me: you can achieve whatever you want. Good luck and all the best. **Manos**, thanks to having always been available to answer my bioinformatics doubts and helping me many times in solving bioinformatics problems. You are a very skilled bioinformatician, an adorable and joyful person, and a really kind friend. **Ilyas**, thanks for all the scientific debate we had. Although we haven't always been in agreement, those exchanges of views have been precious in my growth as a scientist. Thanks for your help in setting up and organizing experiments and your career advice. I wish you good luck with your new journey in Turkey. **Nori,** you are probably the most skilled scientist I have ever met, other than having a broad range of knowledge, you have shown me very impressive abilities in the bench work. Thanks for having introduced me to the world of chromatin interactions and sequencing methodologies. I wish you a happy and lucky life. **Lluis,** thanks for the smile and joy you had every time you were at work and for your help and advice at the begging of my PhD. You are a very kind and resourceful person and I wish you good luck with your career. **Hao** and **Agniezka**, thanks for the time passing together in and outside the lab, for the help with the experiment and for all your advice. I wish both of you the best of luck in your life.

My daily life at work would not be so memorable without all the colleagues working at the 6$^{th}$ floor of Bioclinicum.

**Konstantina**, you are a special person, very kind and emphatic, your small gesture like a word or a hug has changed my days so many times. You're strong and smart, just believe in yourself. Your friendship is very precious and I hope will last also when our life will get on separate paths. **Wen,** you are for sure the cutest! You are very smart and resourceful: a biologist, a statistician, a cook and a baker. Your smile is contagious and your kindness

a precious gift. I'm very glad about our friendship and I'm sure you will have a successful PhD. **Amineh**, your resourcefulness is amazing. You are a very strong person with a very fighting spirit. Thanks for all the kindness and friendship you have shown to me, you are a wonderful person. Good luck with your career. **Apple**, your kindness is impressive and your joy so contagious. Thanks to having cheering up many of my days and for your friendship. Good luck with your PhD and your future career. **Stephan**, you are a very likeable man, your Instagram reels and joke have made me laugh so many times and helped me in these intense writing days. Thanks for your friendship. **Pedro,** thanks for the many happy conversations we had in the cell culture that has made the work much easier and for always helping when needed. You are a really kind and generous person. Good luck with your PhD and future career. **Vicky**, you are a very kind person, thanks for all the nice time passed together discussing PhD students' welfare, science and social activities, your enthusiasm is contagious. Be strong and good luck with your Phd. **Paula**, you are really essential to the entire department and your help in those years has been so essential. You are very kind, patient and very resourceful. Thanks to be available any time I was in need of help and for all the work you do as the department coordinator. Associate Professor **Ekaterina Pokrovskaja (Katja)**, thanks for all the advice that you gave to me and the kindness you have shown me over these years. Your work for the entire department is so important and I have been always amazed by your patience and generosity. Your scientific questions during the seminars have always shown you scientific acumen as well. Your passion for the environment and sustainable development is really impressive. Thanks a lot for all your support and work.
A sincere thanks to all the other people with whom I have passed very joyful times during lunch and on many more occasions at work and that have always been available for help when needed: **Minkai**, **Chris**, **Mireia P.**, **Mireia C.**, **Aaran**, **Le Tong**, **Yuan**, **Yumen**, **Mark**, **Domenika**, **Sylvia**, **Wai Tao**, **Angelos**, **Alvaro.**

I would like also to thank all department leadership and administration: the head of the department Professor **Lars Holgren,** the previous and current directors of doctoral education Professor **Andreas Lundqvist** and Associate Professor **Per Hydbring**, the chairperson of the work environment committee Associate Professor **Charlotte Rolny,** the head of administration **Maria von Witting**, the HR manager **Anne Jensen**, the doctoral education administrator **Erika Rindsjö** and the HR administrator **Hanna Sillén**. As the department doctoral representative, I'm glad I could strictly collaborate with all of you in those years. With your work, you continuously contribute to the welfare, a nice work environment and good organization of the department, all essential features to be able to perform good and advance science. Thanks also to have been always available to answer my requests, help me when it was needed and listen to my suggestions.

For the development of my scientific skills and my knowledge of 3D nuclear architecture and genome organization, I'm also very grateful to all members of the EU-funded Marie

Curie ITN Action "**Chromatin3D**" consortium of which I'm really glad to have been a member: **Ourania**, **Nefeli**, **Ioannis-Marios**, **Tomas**, **Thodoris**, **Alessandro**, **Florian**, **Iwona**, **Nastia**, **Nina**, **David**, **Nikolas** and **Himanshu**, the coordinator Professor **Charalampos Spilianakis** and all other group leaders. All the conferences and events passed together have been very scientifically rich and dynamic and have contributed to developing new ideas and perspectives for my works. Moreover, the social time together has left many nice memories. I hope I can meet all of you again one day.

An important part of my life as a doctoral student at KI has been as a board member of the **Doctoral Students' Association (DSA)** of which I have been honored to be the chairperson. I'm really glad to all the people I have met in these years and with whom I collaborated to improve doctoral education and the welfare of the doctoral students at KI. I would like to thank all the people that have been part of the board or have helped as student representatives.

A special thanks to **Sebastian,** you are a great scientist, an amazing person a dear friend. I'm glad for the help in all the work that we have done together during my years as chairperson, I couldn't wish for a better vice. **Jan Philipp**, you are a really resourceful and skilled person and very knowledgeable in many fields. Thanks for your friendship and for your determination in these years to assure good mental health among the doctoral students at KI. **Leonie,** I have always admired your empathy for other people and your way to approach life and people. Thanks for all your work as a student representative and for your support as a friend. **Dina,** you are so generous, strong, and kind. I have always been amazed by your positivity in any situation and your smile is so contagious. You deserve all the best in your career. Thanks for all your moral support and nice words in these years, I really appreciated it. **Manouk,** you have a very fighting spirit and you know what you want, these features will bring you far away in your career and life. You are a really kind and generous person. Thanks for all your work in guiding the new generation of doctoral student representatives at KI.
Thanks to **Valeria**, **Leif**, **Fredrika**, **Jessica**, **Alicia**, **Lea**, **Jo**, **Nestor**, **Qiaoli**, **Iulia**, **Benedek, Petra** and all the other people I have met in DSA in those years, you are all special and you deserve to be thanked for all your work as student representative. Thanks also to the several presidia of **Medicinska Föreningen** with whom I have had the honor to collaborate in these years: **Jessica**, **Puck**, **Alexander**, **Andreas**, **Lovisa** and **Astrid**. Thanks for your work in keeping the reins of the KI student association and for having supported the DSA work.

A special thanks to the central study director of doctoral education **Ingeborg Van Der Ploeg**, I'm really glad to have the possibility to collaborate with you on many occasions. Your passion for education is immense and your hard work to assure good quality doctoral education at KI is really unparalleled Thanks for your kindness and support in

these years. Thanks for the ethical discussion we had and for all your amazing work. I hope we can continue to collaborate also in the future.

My life in Stockholm would not have been the same without special people with whom I have shared unforgettable moments and memories and that have formed my family in this country. I'm talking about my "cazzoni" friends.

**Marco,** I admire you for your firmness and for being direct and honest. Thanks for having always been there in case of need, for moral support, for simply doing something together, and for telling me what I really need to hear but I didn't want to. Thanks for all our amazing debates that could last forever, they have often been so deep and left me thinking about them also after our conversations. **Joep,** you are one of the most sincere and authentic people I have ever met. I'm glad of your friendship. I admire your spontaneity and especially your way to face life. Thanks for all the moments together, for the help and support when needed and for showing me that sometimes is better just to live thoughtlessly. **Jemina**, I still have my doubt if you are really Finnish! You are such a kind, lovable and resourceful person. You are always available for help and advice. I admire your resolution and intrepidness. Thanks for all the amazing time passed together, for many times following my crazy ideas and for your friendship. **Gabriele,** you are so skilled in so many fields and I'm amazed by how you can find time to follow so many hobbies. You are really great in whatever you do. I admire you for your so broad and wide knowledge. Thanks for have been a good friend also when sometimes I didn't deserve it. **Alice,** thanks for your tenderness, I admire your simplicity, it's a great virtue. Thanks for having good words when needed and to find time for our friendship. I have so many funny moments with you, you really bring joy wherever you are. **Shane,** you can bring craziness and wisdom at the same time, that's so unique! You are a fabulous person, thanks for all your career and life advice, for always believing in me and for your guidance in these years. You are the precious friend that everyone should have. **Maria,** your sincerity and being direct is what makes you special. You know what to say and at which moments. Thanks for your sometimes hard, but needed and sincere, advice. **Stefania**, your blitheness is contagious. You bring joy and fun wherever you are, don't ever change that. Thanks to be you and have lighted up all moments we had together. Thanks, **Kostas** for your sincerity and good friendship. **Alek**, I admire your passion and dedication to work and your broad range of knowledge. Your taking life as it comes makes you quite a unique friend. **Nuria,** you are so special in so many ways. I admire your being so spontaneous, kind, wise, and resourceful. You have many talents and passions. Thanks for having been a good friend and bringing charm and fun to our moments together. With the resourceful person that is **Carlos** you make together a really beautiful couple. **Anna,** your energy is always so contagious. You are a sincere and trustworthy friend. You have made special many days here in Stockholm. I admire your free and rebellious spirit. I wish you a wonderful future with **Uri**.

Another special group of friends with whom everything has started is the Larsberg family: you, guys, have been so essential for starting my life here in Sweden. If after 8 years I'm still living here and I could complete my PhD is because my first years living in the same building as you helped me to not miss home. **Luisa**, I don't have words to describe your kindness and how much lovable person you are. I admire your firmness and your elegance in every moment. You are a really precious friend. **Lorenzo,** you are a very resourceful and smart person, your kindness is admirable. **Leonardo**, your free spirit and your so friendly personality have always amazed me. You never give up! **Gabriele,** I miss the Sunday lunches you were making for all of us.

Thanks to **Elisabeth** and **Thomas**, you guys are the best host and trip organizers I have ever met. Every time we organize something together is pure fun and relaxation. Both of you are so resourceful and skilled, I admire your being able to do almost everything and adapt to any situation. I'm amazed by your so adventurous spirit. Thanks for the unforgettable moments that we had together.

**Daniel,** thanks for your simplicity and for being a sincere and precious friend with whom it's always easy to talk and confide. And thanks to be often available for our last-minute Sunday lunch and for our theological discussions. I wish you all the best for your new adventure as husband and father.

I would like to thank also all the other friends I have met in these years in Stockholm, **Luca**, **Sara**, **Chiara**, **Zula**, **Louise** the scout group, **Jonas** and **Linnar**, and all the amazing people I have met and that have made my life here so enjoyable.

Se sono quel che sono oggi ed ho potuto intraprendere questa strada ed inseguire i miei sogni, molto è anche dovuto alle persone che sono state affianco a me per tutta o buona parte della mia vita e che continuamente mi hanno risollevato da momenti di sconforto, tipici del lavoro di un ricercatore, hanno creduto in me e mi hanno aiutato e sostenuto quando necessari: la mia famiglia e i miei amici di sempre.

**Ilaria**, **Giovanni**, **Mario**, **Valeria** e **Antonio** grazie a voi e alla vostra amicizia le prime fasi della mia educazione scientifica, durante le lauree triennale e magistrale sono scorse senza intoppi. Nonostante tutti questi anni e la lontananza mi lega ad ognuno di voi una profonda amicizia. Ognuno di voi mi ha lasciato un qualcosa di speciale: Ilaria la tua gioia di vita, Giovanni la tua determinazione, Mario la tua schiettezza, Valeria la tua genuinità, Antonio la tua euforia.

**Andrea**, **Clara**, **Cristina**, **Luca**, **Giorgia** e **Rosangela**, voi siete la mia ancora alle origini. Non importa quanto distanti e sparsi nel mondo viviamo e quanto orami poco spesso riusciamo ad incontrarci, soprattutto tutti insieme, ma ogni volta che accade è come se non ci fossimo mai separati e ritorniamo ad essere quei spensierati compagni di liceo di sempre. So che posso sempre contare su di voi e siete i primi che contatto per un

consiglio o per sentire una voce familiare che mi tiri su. Grazie! Andrea i tuoi consigli sono sempre preziosi, il tuo affrontare la vita con determinazione e il tuo saperti rialzare dopo ogni caduta ti rendono l'uomo forte e risoluto che sei. Clara, tu sei la dolcezza e la spensieratezza allo stesso tempo, sai farti valere e sai come ottenere quello che vuoi. Cristina, lo sai, la descrizione che più ti si addice è essere la mamma. Ti preoccupi sempre e sai sempre come consolare e tirare su. Sei forte e niente ti abbatte. Luca, tu sei l'equilibrio e la ponderatezza che ti permettono di affrontare tutti i momenti senza cadere. Vivi la vita a testa alta. Gorgia tu sei quella tenerezza di cui non si può fare a meno. Sai sempre come e quando dire la verità anche se non si vorrebbe sentire. Rosangela tu sei un fuoco, sempre schietta e diretta. Si può sempre contare su di te e le tue doti organizzative sono le migliori.

**Pierluca**, grazie per essere al mio fianco da sempre. So che qualsiasi cosa accade tu sarai li per me. Sei una delle persone più intrepide che io conosca, anche se non lo lasci a vedere. Affronti sempre il futuro con le migliori intenzioni e ti rialzi più forte ogni volta che cadi. Ti sarò per sempre grato per la bellissima infanzia trascorsa insieme e per ogni momento vissuto con te.

Un grazie speciale ai miei zii che mi hanno sempre dimostrato l'importanza della forza famiglia nonostante le distanze. Zio **Lello**, sei una delle persone più spiritualmente forti che io conosca, mi hai insegnato come la determinazione sia la chiave per il successo. Zio **Mario**, nessuno come te riesce a trovare la forza della gioia e della speranza anche nei momenti più tristi. Mi hai insegnato che spesso i piccoli gesti, come una semplice battuta o un abbraccio, possono avere un enorme valore e che essere sé stessi è sempre la scelta migliore. Un grazie anche a tutti le mie zie, **Christine**, **Elodie**, **Rosa** e cugini **Barbara**, **Lea**, **Matteo**, **Luca**, **Mattia** e **Chiara** con il quale sono cresciuto e con quali il ritrovarsi è sempre un momento speciale e di gioia. Siete sempre nel mio cuore. Ed ognuno di voi è speciale in qualche modo.

Un grazie a tutti i miei **nonni**, a chi c'è ancora e a chi mi guarda da lassù, e a zio **Paolo**: grazie per aver sempre creduto in me and avermi cresciuto sommerso di amore. Non c'è altro che un nipote possa desiderare.

Grazie ai miei fratelli, **Alessio** e **Valerio**, per essere tali sempre ed in ogni occasione, nonostante tutto. Siete speciali e avete contribuito in maniera decisiva alla mia crescita e maturazione. Alessio, sei semplicemente forte, resisti a tutto e non molli mai. La tua grinta ti porterà lontano, hai carisma, carattere, generosità, simpatia, tutto ciò che occorre per avere successo. Non smettere mai di credere in te stesso. Valerio, sei puro, fai e dici quel che pensi, non ti spaventa il giudizio altrui e sai adattarti in ogni situazione. Riesci sempre ad ottenere quel che vuoi con perseveranza e furbizia. Sei una forza della natura. Non arrenderti mai.

E per finire il grazie più importante di tutti, grazie **mamma** e **papà**, miei fari della notte e luce del mattino. Avete sempre creduto in me, forse anche più di quanto io, a volte, credessi in me stesso. Mi ha avete sempre sostenuto e dato la possibilità di inseguire qualsiasi strada io volessi. Questo traguardo non sarebbe stato raggiunto senza il vostro incondizionato amore e supporto. Mamma, sei semplicemente unica, sempre pronta a sacrificarti per me e la nostra famiglia. La tua presenza è sempre una gioia per il cuore e un energizzante per lo spirito. La tua dolcezza è immensa, il tuo spirito cosi vigoroso. Sei la mia bussola. Mi hai insegnato a dare sempre il meglio di me e a non arrendermi mai. Papà sei pura energia, sei sempre un modello da cui prendere inspirazione. Non molli mai e non ti tiri indietro di fronte alle avversità. La tua calma nei momenti più difficili e i tuoi consigli sono sempre stati una guida per la mia vita. Hai sempre creduto in me e mi hai insegnato che c'è sempre una soluzione a qualsiasi problema. Non so che parole usare per ringraziarvi entrambi dal più profondo del mio cuore.

# 9 References


1. Field, M. C. & Rout, M. P. Pore timing: the evolutionary origins of the nucleus and nuclear pore complex. *F1000research* 8, F1000 Faculty Rev-369 (2019).

2. Devos, D. P., Gräf, R. & Field, M. C. Evolution of the nucleus. *Curr Opin Cell Biol* 28, 8–15 (2014).

3. Wilson, K. L. & Dawson, S. C. Functional evolution of nuclear structure. *J Cell Biol* 195, 171–181 (2011).

4. Bhat, P., Honson, D. & Guttman, M. Nuclear compartmentalization as a mechanism of quantitative control of gene expression. *Nat Rev Mol Cell Bio* 22, 653–670 (2021).

5. Sabari, B. R., Dall'Agnese, A. & Young, R. A. Biomolecular Condensates in the Nucleus. *Trends Biochem Sci* 45, 961–977 (2020).

6. Ulianov, S. V., Gavrilov, A. A. & Razin, S. V. Chapter Five Nuclear Compartments, Genome Folding, and Enhancer-Promoter Communication. *Int Rev Cel Mol Bio* 315, 183–244 (2015).

7. Cremer, M. & Cremer, T. Nuclear compartmentalization, dynamics, and function of regulatory DNA sequences. *Genes Chromosom. Cancer* 58, 427–436 (2019).

8. Brangwynne, C. P. *et al.* Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science* 324, 1729–1732 (2009).

9. Padeken, J. & Heun, P. Nucleolus and nuclear periphery: Velcro for heterochromatin. *Curr Opin Cell Biol* 28, 54–60 (2014).

10. Tzur, Y. B., Wilson, K. L. & Gruenbaum, Y. SUN-domain proteins: "Velcro" that links the nucleoskeleton to the cytoskeleton. *Nat Rev Mol Cell Bio* 7, 782–788 (2006).

11. Brodsky, S. *et al.* Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. *Mol Cell* 79, 459-471.e4 (2020).

12. Misteli, T. The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell* 183, 28–45 (2020).

13. Stack, S. M., Brown, D. B. & Dewey, W. C. Visualization of interphase chromosomes. *J Cell Sci* 26, 281–299 (1977).

14. Lemaître, C. & Bickmore, W. A. Chromatin at the nuclear periphery and the regulation of genome functions. *Histochemistry and Cell Biology* 144, 111–122 (2015).

15. Branco, M. R. & Pombo, A. Intermingling of Chromosome Territories in Interphase Suggests Role in Translocations and Transcription-Dependent Associations. *Plos Biol* 4, e138 (2006).

16. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* 326, 289–293 (2009).

17. Mahmood, S. R. *et al.* β-actin dependent chromatin remodeling mediates compartment level changes in 3D genome architecture. *Nat Commun* 12, 5240 (2021).

18. Xie, X. *et al.* β-Actin-dependent global chromatin organization and gene expression programs control cellular identity. *Faseb J* 32, 1296–1314 (2018).

19. Mahmood, S. R., Said, N. H. E., Gunsalus, K. C. & Percipalle, P. β-actin mediated H3K27ac changes demonstrate the link between compartment switching and enhancer-dependent transcriptional regulation. *Genome Biol* 24, 18 (2023).

20. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380 (2012).

21. Hnisz, D., Day, D. S. & Young, R. A. Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell* 167, 1188–1200 (2016).

22. Dehingia, B., Milewska, M., Janowski, M. & Pękowska, A. CTCF shapes chromatin structure and gene expression in health and disease. *Embo Rep* 23, e55146 (2022).

23. Phillips, J. E. & Corces, V. G. CTCF: Master Weaver of the Genome. *Cell* 137, 1194–1211 (2009).

24. Filippova, G. N. *et al.* An Exceptionally Conserved Transcriptional Repressor, CTCF, Employs Different Combinations of Zinc Fingers To Bind Diverged Promoter Sequences of Avian and Mammalian c-myc Oncogenes. *Mol Cell Biol* 16, 2802–2813 (1996).

25. Ohlsson, R., Lobanenkov, V. & Klenova, E. Does CTCF mediate between nuclear organization and gene expression? *Bioessays* 32, 37–50 (2010).

26. Wang, H. *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res* 22, 1680–1688 (2012).

27. Lobanenkov, V. V. *et al.* A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene* 5, 1743–53 (1990).

28. Bell, A. C., West, A. G. & Felsenfeld, G. The Protein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators. *Cell* 98, 387–396 (1999).

29. Hanssen, L. L. P. *et al.* Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nat Cell Biol* 19, 952–961 (2017).

30. Guibert, S. *et al.* CTCF-binding sites within the H19 ICR differentially regulate local chromatin structures and cis-acting functions. *Epigenetics* 7, 361–369 (2012).

31. Lutz, M. *et al.* Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Res* 28, 1707–1713 (2000).

32. Chernukhin, I. *et al.* CTCF Interacts with and Recruits the Largest Subunit of RNA Polymerase II to CTCF Target Sites Genome-Wide. *Mol Cell Biol* 27, 1631–1648 (2007).

33. Lee, R. *et al.* CTCF-mediated chromatin looping provides a topological framework for the formation of phase-separated transcriptional condensates. *Nucleic Acids Res* 50, 207–226 (2021).

34. Gabriele, M. *et al.* Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* 376, 496–501 (2022).

35. de Wit, E. *et al.* CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell* 60, 676–684 (2015).

36. Davidson, I. F. & Peters, J.-M. Genome folding through loop extrusion by SMC complexes. *Nat Rev Mol Cell Bio* 22, 445–464 (2021).

37. Hsieh, T.-H. S. *et al.* Enhancer–promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nat Genet* 54, 1919–1932 (2022).

38. Aljahani, A. *et al.* Analysis of sub-kilobase chromatin topology reveals nano-scale regulatory interactions with variable dependence on cohesin and CTCF. *Nat Commun* 13, 2139 (2022).

39. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* 161, 1012–1025 (2015).

40. Wang, W., Ren, G., Hong, N. & Jin, W. Exploring the changing landscape of cell-to-cell variation after CTCF knockdown via single cell RNA-seq. *Bmc Genomics* 20, 1015 (2019).

41. Gong, Y. *et al.* Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nat Commun* 9, 542 (2018).

42. Yu, W. *et al.* Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. *Nature Genetics* 36, 1105–1110 (2004).

43. Zhao, H. *et al.* PARP1- and CTCF-Mediated Interactions between Active and Repressed Chromatin at the Lamina Promote Oscillating Transcription. *Molecular cell* 59, 984–97 (2015).

44. Callan, H. G. & Tomlin, S. G. Experimental studies on amphibian oocyte nuclei I. Investigation of the structure of the nuclear membrane by means of the electron microscope. *Proc Royal Soc Lond Ser B - Biological Sci* 137, 367–378 (1950).

45. Gerace, L. & Burke, B. Functional Organization of the Nuclear Envelope. *Annu Rev Cell Dev Bi* 4, 335–374 (1988).

46. Muchir, A. & Worman, H. J. The Nuclear Envelope and Human Disease. *Physiology* 19, 309–314 (2004).

47. Eriksson, M. *et al.* Recurrent de novo point mutations in lamin A cause Hutchinson–Gilford progeria syndrome. *Nature* 423, 293–298 (2003).

48. Rae, P. M. M. & Franke, W. W. The interphase distribution of satellite DNA-containing heterochromatin in mouse nuclei. *Chromosoma* 39, 443–456 (1972).

49. Schermelleh, L. *et al.* Subdiffraction Multicolor Imaging of the Nuclear Periphery with 3D Structured Illumination Microscopy. *Science* 320, 1332–1336 (2008).

50. Huisinga, K. L., Brower-Toland, B. & Elgin, S. C. R. The contradictory definitions of heterochromatin: transcription and silencing. *Chromosoma* 115, 110–122 (2006).

51. Zhang, T., Cooper, S. & Brockdorff, N. The interplay of histone modifications – writers that read. *Embo Rep* 16, 1467–1481 (2015).

52. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* 403, 41–45 (2000).

53. Schmid, C. D. & Bucher, P. ChIP-Seq Data Reveal Nucleosome Architecture of Human Promoters. *Cell* 131, 831–832 (2007).

54. Boyle, A. P. *et al.* High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell* 132, 311–322 (2008).

55. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biology* 109, 21.29.1-21.29.9 (2015).

56. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453, 948–951 (2008).

57. Wen, B., Wu, H., Shinkai, Y., Irizarry, R. A. & Feinberg, A. P. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat Genet* 41, 246–250 (2009).

58. Poleshko, A. *et al.* H3K9me2 orchestrates inheritance of spatial positioning of peripheral heterochromatin through mitosis. *Elife* 8, e49278 (2019).

59. Meuleman, W. *et al.* Constitutive nuclear lamina–genome interactions are highly conserved and associated with A/T-rich sequence. *Biotechfor* 23, 270–280 (2013).

60. Zullo, J. M. *et al.* DNA Sequence-Dependent Compartmentalization and Silencing of Chromatin at the Nuclear Lamina. *Cell* 149, 1474–1487 (2012).

61. Peric-Hupkes, D. *et al.* Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. *Mol Cell* 38, 603–613 (2010).

62. Tonekaboni, S. A. M., Haibe-Kains, B. & Lupien, M. Large organized chromatin lysine domains help distinguish primitive from differentiated cell populations. *Nat Commun* 12, 499 (2021).

63. Kumar, D., Cinghu, S., Oldfield, A. J., Yang, P. & Jothi, R. Decoding the function of bivalent chromatin in development and cancer. *Genome Res* 31, gr.275736.121 (2021).

64. See, K. *et al.* Histone methyltransferase activity programs nuclear peripheral genome positioning. *Dev Biol* 466, 90–98 (2020).

65. Towbin, B. D. *et al.* Step-Wise Methylation of Histone H3K9 Positions Heterochromatin at the Nuclear Periphery. *Cell* 150, 934–947 (2012).

66. Wu, R., Terry, A. V., Singh, P. B. & Gilbert, D. M. Differential Subnuclear Localization and Replication Timing of Histone H3 Lysine 9 Methylation States. *Mol Biol Cell* 16, 2872–2881 (2005).

67. Yokochi, T. *et al.* G9a selectively represses a class of late-replicating genes at the nuclear periphery. *Proc National Acad Sci* 106, 19363–19368 (2009).

68. Tammer, L. *et al.* Gene architecture directs splicing outcome in separate nuclear spatial regions. *Mol Cell* 82, 1021-1034.e8 (2022).

69. Caballero, L. M. *et al.* The inner nuclear membrane protein Lem2 coordinates RNA degradation at the nuclear periphery. *Nat Struct Mol Biol* 29, 910–921 (2022).

70. Lin, D. H. & Hoelz, A. The Structure of the Nuclear Pore Complex (An Update). *Annu Rev Biochem* 88, 1–59 (2019).

71. Watson, M. L. Further Observations on the Nuclear Envelope of the Animal Cell. *J Cell Biology* 6, 147–156 (1959).

72. Mosalaganti, S. *et al.* AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science* 376, eabm9506 (2022).

73. Hampoelz, B., Andres-Pons, A., Kastritis, P. & Beck, M. Structure and Assembly of the Nuclear Pore Complex. *Annu Rev Biophys* 48, 1–22 (2019).

74. Souquet, B. *et al.* Nup133 Is Required for Proper Nuclear Pore Basket Assembly and Dynamics in Embryonic Stem Cells. *Cell Reports* 23, 2443–2454 (2018).

75. Bilir, Ş. *et al.* Roles of Nup133, Nup153 and membrane fenestrations in assembly of the nuclear pore complex at the end of mitosis. *Genes Cells* 24, 338–353 (2019).

76. Kimura, N. *et al.* Identification of a novel transcription factor, ELYS, expressed predominantly in mouse foetal haematopoietic tissues. *Genes Cells* 7, 435–446 (2002).

77. Pascual-Garcia, P. *et al.* Metazoan Nuclear Pores Provide a Scaffold for Poised Genes and Mediate Induced Enhancer-Promoter Contacts. *Mol Cell* 66, 63-76.e6 (2017).

78. Liu, X. *et al.* In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell* 170, 1028-1043.e19 (2017).

79. Ibarra, A., Benner, C., Tyagi, S., Cool, J. & Hetzer, M. W. Nucleoporin-mediated regulation of cell identity genes. *Genes Dev* 30, 2253–2258 (2016).

80. Pascual-Garcia, P. & Capelson, M. Nuclear pores in genome architecture and enhancer function. *Curr Opin Cell Biol* 58, 126–133 (2019).

81. Kalverda, B., Pickersgill, H., Shloma, V. V. & Fornerod, M. Nucleoporins Directly Stimulate Expression of Developmental and Cell-Cycle Genes Inside the Nucleoplasm. *Cell* 140, 360–371 (2010).

82. Cubeñas-Potts, C. *et al.* Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic Acids Res* 45, 1714–1730 (2017).

83. Ramírez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun* 9, 189 (2018).

84. Gozalo, A. *et al.* Core Components of the Nuclear Pore Bind Distinct States of Chromatin and Contribute to Polycomb Repression. *Mol Cell* 77, 67-81.e7 (2020).

85. Schneider, M. *et al.* The Nuclear Pore-Associated TREX-2 Complex Employs Mediator to Regulate Gene Expression. *Cell* 162, 1016–1028 (2015).

86. D'Angelo, M. A., Gomez-Cavazos, J. S., Mei, A., Lackner, D. H. & Hetzer, M. W. A Change in Nuclear Pore Complex Composition Regulates Cell Differentiation. *Dev Cell* 22, 446–458 (2012).

87. Raices, M. *et al.* Nuclear Pores Regulate Muscle Development and Maintenance by Assembling a Localized Mef2C Complex. *Dev Cell* 41, 540-554.e7 (2017).

88. Jacinto, F. V., Benner, C. & Hetzer, M. W. The nucleoporin Nup153 regulates embryonic stem cell pluripotency through gene silencing. *Genes Dev* 29, 1224–1238 (2015).

89. Inoue, A. & Zhang, Y. Nucleosome assembly is required for nuclear pore complex assembly in mouse zygotes. *Nat Struct Mol Biol* 21, 609–616 (2014).

90. Shevelyov, Y. Y. The Role of Nucleoporin Elys in Nuclear Pore Complex Assembly and Regulation of Genome Architecture. *Int J Mol Sci* 21, 9475 (2020).

91. Sunny, D. E. *et al.* Nup133 and ERα mediate the differential effects of hyperoxia-induced damage in male and female OPCs. *Mol Cell Pediatrics* 7, 10 (2020).

92. Iborra, F. J., Pombo, A., Jackson, D. A. & Cook, P. R. Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. *J Cell Sci* 109, 1427–1436 (1996).

93. Cisse, I. I. *et al.* Real-Time Dynamics of RNA Polymerase II Clustering in Live Human Cells. *Science* 341, 664–667 (2013).

94. Cho, W.-K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 361, 412–415 (2018).

95. Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* 175, 1842-1855.e16 (2018).

96. Zamudio, A. V. *et al.* Mediator Condensates Localize Signaling Factors to Key Cell Identity Genes. *Mol Cell* 76, 753-766.e6 (2019).

97. Wei, M. *et al.* Nuclear actin regulates inducible transcription by enhancing RNA polymerase II clustering. *Sci Adv* 6, eaay6515 (2020).

98. Zhou, Q., Li, T. & Price, D. H. RNA Polymerase II Elongation Control. *Biochemistry-us* 81, 119–143 (2012).

99. Guo, Y. E. *et al.* Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* 572, 543–548 (2019).

100. Lamond, A. I. & Spector, D. L. Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Bio* 4, 605–612 (2003).

101. Misteli, T., Cáceres, J. F. & Spector, D. L. The dynamics of a pre-mRNA splicing factor in living cells. *Nature* 387, 523–527 (1997).

102. Ding, F. & Elowitz, M. B. Constitutive splicing and economies of scale in gene expression. *Nat Struct Mol Biol* 26, 424–432 (2019).

103. Pederiva, C. & Farnebo, M. RNF8 - The Achilles heel of DNA repair when splicing rules. *Cell Cycle* 17, 137–138 (2018).

104. Galganski, L., Urbanek, M. O. & Krzyzosiak, W. J. Nuclear speckles: molecular organization, biological function and role in disease. *Nucleic Acids Res* 45, gkx759- (2017).

105. Fox, A. H. *et al.* Paraspeckles A Novel Nuclear Domain. *Curr Biol* 12, 13–25 (2002).

106. McCluggage, F. & Fox, A. H. Paraspeckle nuclear condensates: Global sensors of cell stress? *Bioessays* 43, 2000245 (2021).

107. Feng, P. *et al.* NONO and tumorigenesis: More than splicing. *J Cell Mol Med* 24, 4368–4376 (2020).

108. Lafontaine, D. L. J., Riback, J. A., Bascetin, R. & Brangwynne, C. P. The nucleolus as a multiphase liquid condensate. *Nat Rev Mol Cell Bio* 22, 165–182 (2021).

109. Lyons, S. M. *et al.* A subset of replication-dependent histone mRNAs are expressed as polyadenylated RNAs in terminally differentiated tissues. *Nucleic Acids Res* 44, 9190–9205 (2016).

110. Bergstrand, S. *et al.* Small Cajal body-associated RNA 2 (scaRNA2) regulates DNA repair pathway choice by inhibiting DNA-PK. *Nat Commun* 13, 1015 (2022).

111. Caudron-Herger, M. & Rippe, K. Nuclear architecture by RNA. *Curr Opin Genet Dev* 22, 179–187 (2012).

112. Bai, L., Santangelo, T. J. & Wang, M. D. SINGLE-MOLECULE ANALYSIS OF RNA POLYMERASE TRANSCRIPTION. *Annu Rev Bioph Biom* 35, 343–360 (2006).

113. Esteller, M. Non-coding RNAs in human disease. *Nat Rev Genet* 12, 861–874 (2011).

114. Girbig, M., Misiaszek, A. D. & Müller, C. W. Structural insights into nuclear transcription by eukaryotic DNA-dependent RNA polymerases. *Nat Rev Mol Cell Bio* 23, 603–622 (2022).

115. Watt, K. E., Macintosh, J., Bernard, G. & Trainor, P. A. RNA Polymerases I and III in development and disease. *Semin Cell Dev Biol* 136, 49–63 (2023).

116. Raff, B. A. • R. H. • A. J. • D. M. • M., Heald, R., Johnson, A., Morgan, D. & Raff, M. *Molecular Biology of the Cell*. (WW Norton & Co, 2022).

117. Pope, S. D. & Medzhitov, R. Emerging Principles of Gene Expression Programs and Their Regulation. *Mol Cell* 71, 389–397 (2018).

118. Jana, T., Brodsky, S. & Barkai, N. Speed–Specificity Trade-Offs in the Transcription Factors Search for Their Genomic Binding Sites. *Trends Genet* 37, 421–432 (2021).

119. Kamieniarz-Gdula, K. & Proudfoot, N. J. Transcriptional Control by Premature Termination: A Forgotten Mechanism. *Trends Genet* 35, 553–564 (2019).

120. Eisenberg, E. & Levanon, E. Y. Human housekeeping genes, revisited. *Trends Genet* 29, 569–574 (2013).

121. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Gene Dev* 25, 1010–1022 (2011).

122. Fowler, T., Sen, R. & Roy, A. L. Regulation of Primary Response Genes. *Mol Cell* 44, 348–360 (2011).

123. Nie, Z. *et al.* c-Myc Is a Universal Amplifier of Expressed Genes in Lymphocytes and Embryonic Stem Cells. *Cell* 151, 68–79 (2012).

124. Thurlings, I. & Bruin, A. de. Cell Cycle Oscillators, Methods and Protocols. *Methods Mol Biology* 1342, 71–88 (2015).

125. Ramirez-Carrozzi, V. R. *et al.* A Unifying Model for the Selective Regulation of Inducible Transcription by CpG Islands and Nucleosome Remodeling. *Cell* 138, 114–128 (2009).

126. Natoli, G., Ghisletti, S. & Barozzi, I. The genomic landscapes of inflammation. *Gene Dev* 25, 101–106 (2011).

127. Mirny, L. *et al.* How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J Phys Math Theor* 42, 434013 (2009).

128. Blum, R., Vethantham, V., Bowman, C., Rudnicki, M. & Dynlacht, B. D. Genome-wide identification of enhancers in skeletal muscle: the role of MyoD1. *Gene Dev* 26, 2763–2779 (2012).

129. Field, A. & Adelman, K. Evaluating Enhancer Function and Transcription. *Annu Rev Biochem* 89, 1–22 (2020).

130. Pachano, T., Haro, E. & Rada-Iglesias, A. Enhancer-gene specificity in development and disease. *Development* 149, (2022).

131. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27, 299–308 (1981).

132. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc National Acad Sci* 107, 21931–21936 (2010).

133. Pekowska, A. *et al.* H3K4 tri-methylation provides an epigenetic signature of active enhancers. *Embo J* 30, 4198–4210 (2011).

134. Han, Z. & Li, W. Enhancer RNA: What we know and what we can achieve. *Cell Proliferat* 55, e13202 (2022).

135. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39, 311–318 (2007).

136. Hamamoto, K. & Fukaya, T. Molecular architecture of enhancer–promoter interaction. *Curr Opin Cell Biol* 74, 62–70 (2022).

137. Thiecke, M. J. *et al.* Cohesin-Dependent and -Independent Mechanisms Mediate Chromosomal Contacts between Promoters and Enhancers. *Cell Reports* 32, 107929 (2020).

138. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* 169, 930-944.e22 (2017).

139. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* 171, 305-320.e24 (2017).

140. Gu, B. *et al.* Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science* eaao3136 (2018) doi:10.1126/science.aao3136.

141. Benabdallah, N. S. *et al.* Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Mol Cell* 76, 473-484.e7 (2019).

142. Kumar, V., Kumar, A., Mir, K. U. I., Yadav, V. & Chauhan, S. S. Pleiotropic role of PARP1: an overview. *3 Biotech* 12, 3 (2022).

143. Nalabothula, N. *et al.* Genome-Wide Profiling of PARP1 Reveals an Interplay with Gene Regulatory Regions and DNA Methylation. *Plos One* 10, e0135410 (2015).

144. Altmeyer, M. *et al.* Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nat Commun* 6, 8088 (2015).

145. Heist, T., Fukaya, T. & Levine, M. Large distances separate coregulated genes in living Drosophila embryos. *Proc National Acad Sci* 116, 15062–15067 (2019).

146. Whyte, W. A. *et al.* Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell* 153, 307–319 (2013).

147. Lovén, J. *et al.* Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers. *Cell* 153, 320–334 (2013).

148. Tang, F., Yang, Z., Tan, Y. & Li, Y. Super-enhancer function and its application in cancer targeted therapy. *Npj Precis Oncol* 4, 2 (2020).

149. Hnisz, D. *et al.* Super-Enhancers in the Control of Cell Identity and Disease. *Cell* 155, 934–947 (2013).

150. Pott, S. & Lieb, J. D. What are super-enhancers? *Nature Genetics* 47, ng.3167 (2014).

151. Di Micco, R. *et al.* Control of Embryonic Stem Cell Identity by BRD4-Dependent Transcriptional Elongation of Super-Enhancer-Associated Pluripotency Genes. *Cell Reports* 9, 234–247 (2014).

152. Willi, M. *et al.* Facultative CTCF sites moderate mammary super-enhancer activity and regulate juxtaposed gene in non-mammary cells. *Nat Commun* 8, 16069 (2017).

153. Huang, J. *et al.* Dissecting super-enhancer hierarchy based on chromatin interactions. *Nat Commun* 9, 943 (2018).

154. Consortium, T. F. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455–461 (2014).

155. Lam, M. T. Y., Li, W., Rosenfeld, M. G. & Glass, C. K. Enhancer RNAs and regulated transcriptional programs. *Trends Biochem Sci* 39, 170–182 (2014).

156. Kim, T.-K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–187 (2010).

157. Alvarez-Dominguez, J. R., Knoll, M., Gromatzky, A. A. & Lodish, H. F. The Super-Enhancer-Derived alncRNA-EC7/Bloodlinc Potentiates Red Blood Cell Development in trans. *Cell Reports* 19, 2503–2514 (2017).

158. Paralkar, V. R. *et al.* Unlinking an lncRNA from Its Associated cis Element. *Mol Cell* 62, 104–110 (2016).

159. Espinosa, J. M. Revisiting lncRNAs: How Do You Know Yours Is Not an eRNA? *Mol Cell* 62, 1–2 (2016).

160. Shii, L., Song, L., Maurer, K., Zhang, Z. & Sullivan, K. E. SERPINB2 is regulated by dynamic interactions with pause-release proteins and enhancer RNAs. *Mol Immunol* 88, 20–31 (2017).

161. Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* 494, 497–501 (2013).

162. Chen, Y. *et al.* Colon cancer associated transcripts in human cancers. *Biomed Pharmacother* 94, 531–540 (2017).

163. Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 498, 516–520 (2013).

164. Bose, D. A. *et al.* RNA Binding to CBP Stimulates Histone Acetylation and Transcription. *Cell* 168, 135-149.e22 (2017).

165. Beltran, M. *et al.* The interaction of PRC2 with RNA or chromatin is mutually antagonistic. *Genome Res* 26, 896–907 (2016).

166. Arnold, P. R., Wells, A. D. & Li, X. C. Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate. *Frontiers Cell Dev Biology* 7, 377 (2020).

167. Schaukowitch, K. *et al.* Enhancer RNA Facilitates NELF Release from Immediate Early Genes. *Mol Cell* 56, 29–42 (2014).

168. Dueva, R. *et al.* Neutralization of the Positive Charges on Histone Tails by RNA Promotes an Open Chromatin Structure. *Cell Chem Biol* 26, 1436-1449.e5 (2019).

169. Hnisz, D. *et al.* Convergence of Developmental and Oncogenic Signaling Pathways at Transcriptional Super-Enhancers. *Mol Cell* 58, 362–370 (2015).

170. McCleland, M. L. *et al.* CCAT1 is an enhancer-templated RNA that predicts BET sensitivity in colorectal cancer. *J Clin Invest* 126, 639–652 (2016).

171. Zhang, Z. *et al.* Transcriptional landscape and clinical utility of enhancer RNAs for eRNA-targeted therapy in cancer. *Nat Commun* 10, 4562 (2019).

172. Gonzalez, M. N., Blears, D. & Svejstrup, J. Q. Causes and consequences of RNA polymerase II stalling during transcript elongation. *Nat Rev Mol Cell Bio* 22, 3–21 (2021).

173. Landick, R. The regulatory roles and mechanism of transcriptional pausing. *Biochem Soc T* 34, 1062–1066 (2006).

174. Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat Genet* 49, 1045–1051 (2017).

175. Rahl, P. B. *et al.* c-Myc Regulates Transcriptional Pause Release. *Cell* 141, 432–445 (2010).

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

177. Proudfoot, N. J. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science* 352, aad9926 (2016).

178. Santos-Pereira, J. M. & Aguilera, A. R loops: new modulators of genome dynamics and function. *Nat Rev Genet* 16, 583–597 (2015).

179. Schmid, M. & Jensen, T. H. Controlling nuclear RNA levels. *Nat Rev Genet* 19, 518–529 (2018).

180. Berg, M. G. *et al.* U1 snRNP Determines mRNA Length and Regulates Isoform Expression. *Cell* 150, 53–64 (2012).

181. Brannan, K. *et al.* mRNA Decapping Factors and the Exonuclease Xrn2 Function in Widespread Premature Termination of RNA Polymerase II Transcription. *Mol Cell* 46, 311–324 (2012).

182. Kaida, D. *et al.* U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468, 664–668 (2010).

183. Chiu, A. C. *et al.* Transcriptional Pause Sites Delineate Stable Nucleosome-Associated Premature Polyadenylation Suppressed by U1 snRNP. *Mol Cell* 69, 648-663.e7 (2018).

184. Iasillo, C. *et al.* ARS2 is a general suppressor of pervasive transcription. *Nucleic Acids Res* 45, gkx647 (2017).

185. Davidson, L., Kerr, A. & West, S. Co-transcriptional degradation of aberrant pre-mRNA by Xrn2. *Embo J* 31, 2566–2578 (2012).

186. Meola, N. *et al.* Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts. *Mol Cell* 64, 520–533 (2016).

187. Magistris, P. D. The Great Escape: mRNA Export through the Nuclear Pore Complex. *Int J Mol Sci* 22, 11767 (2021).

188. Fan, J. *et al.* Exosome cofactor hMTR4 competes with export adaptor ALYREF to ensure balanced nuclear RNA pools for degradation and export. *Embo J* 36, 2870–2886 (2017).

189. Masuda, S. *et al.* Recruitment of the human TREX complex to mRNA during splicing. *Gene Dev* 19, 1512–1517 (2005).

190. Sträßer, K., Baßler, J. & Hurt, E. Binding of the Mex67p/Mtr2p Heterodimer to Fxfg, Glfg, and Fg Repeat Nucleoporins Is Essential for Nuclear mRNA Export. *J Cell Biology* 150, 695–706 (2000).

191. Ben-Yishay, R. *et al.* Imaging within single NPCs reveals NXF1's role in mRNA export on the cytoplasmic side of the pore. *J Cell Biol* 218, 2962–2981 (2019).

192. Khan, M., Hou, S., Chen, M. & Lei, H. Mechanisms of RNA export and nuclear retention. *Wiley Interdiscip Rev Rna* 14, e1755 (2023).

193. Cullen, B. R. Nuclear RNA export. *J Cell Sci* 116, 587–597 (2003).

194. Blobel, G. Gene gating: a hypothesis. *Proc National Acad Sci* 82, 8527–8529 (1985).

195. Ben-Yishay, R., Ashkenazy, A. J. & Shav-Tal, Y. Dynamic Encounters of Genes and Transcripts with the Nuclear Pore. *Trends Genet* 32, 419–431 (2016).

196. Burns, L. T. & Wente, S. R. From Hypothesis to Mechanism: Uncovering Nuclear Pore Complex Links to Gene Expression. *Mol Cell Biol* 34, 2114–2120 (2014).

197. Pascual-Garcia, P. & Capelson, M. Nuclear pores as versatile platforms for gene regulation. *Curr Opin Genet Dev* 25, 110–117 (2014).

198. Malerba, M. E. & Marshall, D. J. Larger cells have relatively smaller nuclei across the Tree of Life. *Evol Lett* 5, 306–314 (2021).

199. Stephens, A. D., Banigan, E. J. & Marko, J. F. Chromatin's physical properties shape the nucleus and its functions. *Curr Opin Cell Biol* 58, 76–84 (2019).

200. Cavalli, G. & Heard, E. Advances in epigenetics link genetics to the environment and disease. *Nature* 571, 489–499 (2019).

201. Badeaux, A. I. & Shi, Y. Emerging roles for chromatin as a signal integration and storage platform. *Nature reviews. Molecular cell biology* 14, 211–24 (2013).

202. Waddington, C. H. The Epigenotype. *Int J Epidemiol* 41, 10–13 (2012).

203. Waddington, C. H. The Strategy of the Genes. (2014) doi:10.4324/9781315765471.

204. Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: A Landscape Takes Shape. *Cell* 128, 635–638 (2007).

205. Pujadas, E. & Feinberg, A. P. Regulated Noise in the Epigenetic Landscape of Development and Disease. *Cell* 148, 1123–1131 (2012).

206. Ladewig, J., Koch, P. & Brüstle, O. Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies. *Nat Rev Mol Cell Biology* 14, 225–36 (2013).

207. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502, 59–64 (2013).

208. Rodriguez, J. *et al.* Intrinsic Dynamics of a Human Gene Reveal the Basis of Expression Heterogeneity. *Cell* 176, 213-226.e18 (2019).

209. Suter, D. M. *et al.* Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics. *Science* 332, 472–474 (2011).

210. Rodriguez, J. & Larson, D. R. Transcription in Living Cells: Molecular Mechanisms of Bursting. *Annu Rev Biochem* 89, 1–24 (2020).

211. Hager, G. L., McNally, J. G. & Misteli, T. Transcription Dynamics. *Mol Cell* 35, 741–753 (2009).

212. Symmons, O. & Raj, A. What's Luck Got to Do with It: Single Cells, Multiple Fates, and Biological Nondeterminism. *Mol Cell* 62, 788–802 (2016).

213. Finn, E. H. & Misteli, T. Molecular basis and biological function of variability in spatial genome organization. *Science* 365, (2019).

214. Sumida, N. *et al.* MYC as a driver of stochastic chromatin networks: implications for the fitness of cancer cells. *Nucleic Acids Res* 48, 10867–10876 (2020).

215. Saxena, K., Subbalakshmi, A. R., Kulkarni, P. & Jolly, M. K. Cancer: More than a geneticist's Pandora's box. *J Biosciences* 47, 21 (2022).

216. Finn, E. H. *et al.* Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* 176, 1502-1515.e10 (2019).

217. Feinberg, A. P. & Levchenko, A. Epigenetics as a mediator of plasticity in cancer. *Science* 379, eaaw3835 (2023).

218. Cheng, F., Liu, C., Shen, B. & Zhao, Z. Investigating cellular network heterogeneity and modularity in cancer: a network entropy and unbalanced motif approach. *Bmc Syst Biol* 10, 65 (2016).

219. Franco, I. & Eriksson, M. Reverting to old theories of ageing with new evidence for the role of somatic mutations. *Nat Rev Genet* 23, 645–646 (2022).

220. Yu, M., Hazelton, W. D., Luebeck, G. E. & Grady, W. M. Epigenetic Aging: More Than Just a Clock When It Comes to Cancer. *Cancer Res* 80, 367–374 (2020).

221. Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science* 357, (2017).

222. Berben, L., Floris, G., Wildiers, H. & Hatse, S. Cancer and Aging: Two Tightly Interconnected Biological Processes. *Cancers* 13, 1400 (2021).

223. McDonald, O. G. *et al.* Epigenomic reprogramming during pancreatic cancer progression links anabolic glucose metabolism to distant metastasis. *Nat Genet* 49, 367–376 (2017).

224. Hansen, K. D. *et al.* Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* 43, 768–775 (2011).

225. Timp, W. *et al.* Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. *Genome Med* 6, 61 (2014).

226. Feinberg, A. P., Koldobskiy, M. A. & Göndör, A. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat Rev Genet* 17, 284–299 (2016).

227. Dhanasekaran, R. *et al.* The MYC oncogene — the grand orchestrator of cancer growth and immune evasion. *Nat Rev Clin Oncol* 19, 23–36 (2022).

228. Zhan, T., Rindtorff, N. & Boutros, M. Wnt signaling in cancer. *Oncogene* 36, 1461–1473 (2017).

229. Yoshida, G. J. Emerging roles of Myc in stem cell biology and novel tumor therapies. *J Exp Clin Canc Res* 37, 173 (2018).

230. Sphyris, N., Hodder, M. C. & Sansom, O. J. Subversion of Niche-Signalling Pathways in Colorectal Cancer: What Makes and Breaks the Intestinal Stem Cell. *Cancers* 13, 1000 (2021).

231. Baluapuri, A., Wolf, E. & Eilers, M. Target gene-independent functions of MYC oncoproteins. *Nat Rev Mol Cell Bio* 21, 255–267 (2020).

232. Fatma, H., Maurya, S. K. & Siddique, H. R. Epigenetic modifications of c-MYC: Role in cancer cell reprogramming, progression and chemoresistance. *Semin Cancer Biol* 83, 166–176 (2022).

233. Soucek, L. *et al.* Modelling Myc inhibition as a cancer therapy. *Nature* 455, 679–683 (2008).

234. Wu, C.-H. *et al.* Cellular senescence is an important mechanism of tumor regression upon c-Myc inactivation. *Proc National Acad Sci* 104, 13028–13033 (2007).

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

236. Kortlever, R. M. *et al.* Myc Cooperates with Ras by Programming Inflammation and Immune Suppression. *Cell* 171, 1301-1315.e14 (2017).

237. Mathsyaraja, H. *et al.* Max deletion destabilizes MYC protein and abrogates Eμ-Myc lymphomagenesis. *Gene Dev* 33, 1252–1264 (2019).

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

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

240. Tesi, A. *et al.* An early Myc-dependent transcriptional program orchestrates cell growth during B-cell activation. *Embo Rep* 20, e47987 (2019).

241. Lin, C. Y. *et al.* Transcriptional Amplification in Tumor Cells with Elevated c-Myc. *Cell* 151, 56–67 (2012).

242. Patange, S. *et al.* MYC amplifies gene expression through global changes in transcription factor dynamics. *Cell Reports* 38, 110292 (2022).

243. Baluapuri, A. *et al.* MYC Recruits SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation. *Mol Cell* 74, 674-687.e11 (2019).

244. Lorenzin, F. *et al.* Different promoter affinities account for specificity in MYC-dependent gene regulation. *Elife* 5, e15161 (2016).

245. Daniel, C. J. *et al.* The Myc Gene, Methods and Protocols. *Methods Mol Biology* 2318, 69–85 (2021).

246. Su, Y. *et al.* Post-translational modification localizes MYC to the nuclear pore basket to regulate a subset of target genes involved in cellular responses to environmental signals. *Gene Dev* 32, 1398–1419 (2018).

247. Marzac, C. *et al.* ATP Binding Cassette transporters associated with chemoresistance: transcriptional profiling in extreme cohorts and their prognostic impact in a cohort of 281 acute myeloid leukemia patients. *Haematologica* 96, 1293–1301 (2011).

248. Matkar, S. *et al.* An Epigenetic Pathway Regulates Sensitivity of Breast Cancer Cells to HER2 Inhibition via FOXO/c-Myc Axis. *Cancer Cell* 28, 472–485 (2015).

249. Rennoll, S. & Yochum, G. Regulation of MYC gene expression by aberrant Wnt/β-catenin signaling in colorectal cancer. *World J Biological Chem* 6, 290–300 (2015).

250. Schuijers, J. *et al.* Transcriptional Dysregulation of MYC Reveals Common Enhancer-Docking Mechanism. *Cell Reports* 23, 349–360 (2018).

251. Thandapani, P. Super-enhancers in cancer. *Pharmacol Therapeut* 199, 129–138 (2019).

252. Lancho, O. & Herranz, D. The MYC Enhancer-ome: Long-Range Transcriptional Regulation of MYC in Cancer. *Trends Cancer* 4, 810–822 (2018).

253. Wasserman, N. F., Aneas, I. & Nobrega, M. A. An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. *Genome Res* 20, 1191–1197 (2010).

254. Huppi, K., Pitt, J. J., Wahlberg, B. M. & Caplen, N. J. The 8q24 Gene Desert: An Oasis of Non-Coding Transcriptional Activity. *Frontiers Genetics* 3, 69 (2012).

255. Witte, J. S. Multiple prostate cancer risk variants on 8q24. *Nat Genet* 39, 579–580 (2007).

256. Sotelo, J. *et al.* Long-range enhancers on 8q24 regulate c-Myc. *Proc National Acad Sci* 107, 3001–3005 (2010).

257. Bahr, C. *et al.* A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature* 553, 515–520 (2018).

258. Dave, K. *et al.* Mice deficient of Myc super-enhancer region reveal differential control mechanism between normal and pathological growth. *Elife* 6, e23382 (2017).

259. Easton, D. F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447, 1087–1093 (2007).

260. Nissan, A. *et al.* Colon cancer associated transcript-1: A novel RNA expressed in malignant and pre-malignant human tissues. *Int. J. Cancer* 130, 1598–1606 (2012).

261. Ghafouri-Fard, S. & Taheri, M. Colon Cancer-Associated Transcripts 1 and 2: Roles and functions in human cancers. *J. Cell. Physiol.* 234, 14581–14600 (2019).

262. Xiang, J.-F. *et al.* Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res* 24, 513–531 (2014).

263. Ye, T., Zhang, J., Zeng, X., Xu, Y. & Li, J. LncRNA CCAT1 is overexpressed in tuberculosis patients and predicts their survival. *Immun Inflamm Dis* 10, 218–224 (2022).

264. Zhang, J. & Gao, Y. CCAT-1 promotes proliferation and inhibits apoptosis of cervical cancer cells via the Wnt signaling pathway. *Oncotarget* 8, 68059–68070 (2017).

265. Liu, Z., Chen, Q. & Hann, S. S. The functions and oncogenic roles of CCAT1 in human cancer. *Biomed Pharmacother* 115, 108943 (2019).

266. Jiang, Y. *et al.* Co-activation of super-enhancer-driven CCAT1 by TP63 and SOX2 promotes squamous cancer progression. *Nat Commun* 9, 3619 (2018).

267. Zhang, E. *et al.* H3K27 acetylation activated-long non-coding RNA CCAT1 affects cell proliferation and migration by regulating SPRY4 and HOXB13 expression in esophageal squamous cell carcinoma. *Nucleic Acids Res* 45, gkw1247 (2016).

268. Wang, D. *et al.* Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* 474, 390–394 (2011).

269. Söderholm, S. & Cantù, C. The WNT/β-catenin dependent transcription: A tissue-specific business. *Wires Mech Dis* 13, e1511 (2021).

270. Nusse, R. & Clevers, H. Wnt/β-Catenin Signaling, Disease, and Emerging Therapeutic Modalities. *Cell* 169, 985–999 (2017).

271. Anastas, J. N. & Moon, R. T. WNT signalling pathways as therapeutic targets in cancer. *Nat Rev Cancer* 13, 11–26 (2013).

272. Sekiya, T. & Zaret, K. S. Repression by Groucho/TLE/Grg Proteins: Genomic Site Recruitment Generates Compacted Chromatin In Vitro and Impairs Activator Binding In Vivo. *Mol Cell* 28, 291–303 (2007).

273. Tago, K. *et al.* Inhibition of Wnt signaling by ICAT, a novel β-catenin-interacting protein. *Gene Dev* 14, 1741–1749 (2000).

274. Takemaru, K.-I. *et al.* Chibby, a nuclear β-catenin-associated antagonist of the Wnt/Wingless pathway. *Nature* 422, 905–909 (2003).

275. Wetering, M. V. D., Castrop, J., Korinek, V. & Clevers, H. Extensive Alternative Splicing and Dual Promoter Usage Generate Tcf-1 Protein Isoforms with Differential Transcription Control Properties. *Mol Cell Biol* 16, 745–752 (1996).

276. Wu, B. *et al.* Ring Finger Protein 14 is a new regulator of TCF/β-catenin-mediated transcription and colon cancer cell survival. *Embo Rep* 14, 347–355 (2013).

277. Bauer, A., Huber, O. & Kemler, R. Pontin52, an interaction partner of β-catenin, binds to the TATA box binding protein. *Proc National Acad Sci* 95, 14787–14792 (1998).

278. Hecht, A., Vleminckx, K., Stemmler, M. P., Roy, F. van & Kemler, R. The p300/CBP acetyltransferases function as transcriptional coactivators of β-catenin in vertebrates. *Embo J* 19, 1839–1850 (2000).

279. Barker, N. *et al.* The chromatin remodelling factor Brg-1 interacts with β-catenin to promote target gene activation. *Embo J* 20, 4935–4943 (2001).

280. Kim, S., Xu, X., Hecht, A. & Boyer, T. G. Mediator Is a Transducer of Wnt/β-Catenin Signaling*. *J Biol Chem* 281, 14066–14075 (2006).

281. Doumpas, N. *et al.* TCF/LEF dependent and independent transcriptional regulation of Wnt/β-catenin target genes. *Embo J* 38, (2019).

282. Zambanini, G., Nordin, A., Jonasson, M., Pagella, P. & Cantù, C. A new cut&run low volume-urea (LoV-U) protocol optimized for transcriptional co-factors uncovers Wnt/b-catenin tissue-specific genomic targets. *Development* 149, (2022).

283. Myant, K. & Sansom, O. J. Wnt/Myc interactions in intestinal cancer: Partners in crime. *Exp Cell Res* 317, 2725–2731 (2011).

284. Easwaran, H. P. & Baylin, S. B. Role of Nuclear Architecture in Epigenetic Alterations in Cancer. *Cold Spring Harb Sym* 75, 507–515 (2010).

285. Liu, Z. *et al.* Circadian regulation of c-MYC in mice. *Proc National Acad Sci* 117, 21609–21617 (2020).

286. Repouskou, A., Sourlingas, T. G., Sekeri-Pataryas, K. E. & Prombona, A. THE CIRCADIAN EXPRESSION OF c-MYC IS MODULATED BY THE HISTONE DEACETYLASE INHIBITOR TRICHOSTATIN A IN SYNCHRONIZED MURINE NEUROBLASTOMA CELLS. *Chronobiol Int* 27, 722–741 (2010).

287. Repouskou, A. & Prombona, A. c-MYC targets the central oscillator gene Per1 and is regulated by the circadian clock at the post-transcriptional level. *Biochimica Et Biophysica Acta Bba - Gene Regul Mech* 1859, 541–552 (2016).

288. Bafico, A., Liu, G., Goldin, L., Harris, V. & Aaronson, S. A. An autocrine mechanism for constitutive Wnt pathway activation in human cancer cells. *Cancer Cell* 6, 497–506 (2004).

289. Morin, P. J. *et al.* Activation of β-Catenin-Tcf Signaling in Colon Cancer by Mutations in β-Catenin or APC. *Science* 275, 1787–1790 (1997).

290. Vestlund, J. *et al.* The Nodewalk assay to quantitate chromatin fiber interactomes in very small cell populations. *Nat Protoc* 18, 755–782 (2023).

291. Jensen, E. Technical Review: In Situ Hybridization. *Anat Rec* 297, 1349–1353 (2014).

292. Fredriksson, S. *et al.* Protein detection using proximity-dependent DNA ligation assays. *Nat Biotechnol* 20, 473–477 (2002).

293. Chen, X. *et al.* Chromatin in situ proximity (ChrISP): single-cell analysis of chromatin proximities at a high resolution. *BioTechniques* 56, 117–8, 120–4 (2014).

294. Sanderson, M. J., Smith, I., Parker, I. & Bootman, M. D. Fluorescence Microscopy. *Cold Spring Harb Protoc* 2014, pdb.top071795 (2014).

295. Felts, L., Kohli, V., Marr, J. M., Schumacher, J. & Schlicker, O. An Introduction to Computational Clearing - A new method to remover out-of-focus blur. Preprint at https://www.leica-microsystems.com/science-lab/life-science/an-introduction-to-computational-clearing/ (2020).

296. Scinto, S. L. *et al.* Bioorthogonal chemistry. *Nat Rev Methods Primers* 1, 30 (2021).

297. Lin, J.-S. & Lai, E.-M. Bacterial Protein Secretion Systems, Methods and Protocols. *Methods Mol Biology* 1615, 211–219 (2017).

298. Das, P. M., Ramachandran, K., vanWert, J. & Singal, R. Chromatin immunoprecipitation assay. *Biotechniques* 37, 961–969 (2004).

299. Harris, V. M. Western Blotting, Methods and Protocols. *Methods Mol Biology* 1312, 465–468 (2015).

300. Magnus, P. D. Background Theories and Total Science. *Philos Sci* 72, 1064–1075 (2005).

301. Krietenstein, N. & Rando, O. J. Mesoscale organization of the chromatin fiber. *Curr Opin Genet Dev* 61, 32–36 (2020).

302. Ho, D. *et al.* Enabling Technologies for Personalized and Precision Medicine. *Trends Biotechnol* 38, 497–518 (2020).

303. Lee, S. S.-J., Fullerton, S. M., Saperstein, A. & Shim, J. K. Ethics of inclusion: Cultivate trust in precision medicine. *Science* 364, 941–942 (2019).

304. Kahn, M. Can we safely target the WNT pathway? *Nat Rev Drug Discov* 13, 513–532 (2014).

305. Hyle, J. *et al.* Acute depletion of CTCF directly affects MYC regulation through loss of enhancer–promoter looping. *Nucleic Acids Res* 47, 6699–6713 (2019).

306. Wang, A. *et al.* Mechanism of Long-Range Chromosome Motion Triggered by Gene Activation. *Dev Cell* 52, 309-320.e5 (2020).

307. Chuang, C.-H. *et al.* Long-Range Directional Movement of an Interphase Chromosome Site. *Curr Biol* 16, 825–831 (2006).

308. Smith, C. L., Poleshko, A. & Epstein, J. A. The nuclear periphery is a scaffold for tissue-specific enhancers. *Nucleic Acids Res* 49, gkab392- (2021).

309. Kuhn, T. M., Pascual-Garcia, P., Gozalo, A., Little, S. C. & Capelson, M. Chromatin targeting of nuclear pore proteins induces chromatin decondensation. *J Cell Biol* 218, 2945–2961 (2019).

310. Pagella, P. *et al.* Time-resolved analysis of Wnt-signaling reveals β-catenin temporal genomic repositioning and cell type-specific plastic or elastic chromatin responses. *Biorxiv* 2022.08.05.502932 (2022) doi:10.1101/2022.08.05.502932.

311. Weth, O. *et al.* CTCF induces histone variant incorporation, erases the H3K27me3 histone mark and opens chromatin. *Nucleic Acids Res* 42, 11941–11951 (2014).

312. Keene, R. G., Mueller, A., Robert, L. & London, L. Transcriptional pause, arrest and termination sites for RNA polymerase II in mammalian N- and c- myc genes. *Nucleic Acids Res* 27, 3173–3182 (1999).

313. He, P. *et al.* Long non-coding RNA CCAT2 promotes prostate cancer cell proliferation and invasion by regulating the Wnt/β-catenin signaling pathway. *Oncol Lett* 20, 97 (2020).

314. Prensner, J. R. *et al.* The Long Non-Coding RNA PCAT-1 Promotes Prostate Cancer Cell Proliferation through cMyc. *Neoplasia* 16, 900–908 (2014).

315. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 361, (2018).

316. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* 169, 13–23 (2017).

317. Hirose, T., Yamazaki, T. & Nakagawa, S. Molecular anatomy of the architectural NEAT1 noncoding RNA: The domains, interactors, and biogenesis pathway required to build phase-separated nuclear paraspeckles. *Wiley Interdiscip Rev Rna* 10, e1545 (2019).

318. Herzog, V. A. *et al.* Thiol-linked alkylation of RNA to assess expression dynamics. *Nat Methods* 14, 1198–1204 (2017).

319. Yang, Q., Riblet, R. & Schildkraut, C. L. Sites That Direct Nuclear Compartmentalization Are near the 5′ End of the Mouse Immunoglobulin Heavy-Chain Locus. *Mol Cell Biol* 25, 6021–6030 (2005).

320. Yang, L. *et al.* Targeting cancer stem cell pathways for cancer therapy. *Signal Transduct Target Ther* 5, 8 (2020).

321. Duffy, M. J., O'Grady, S., Tang, M. & Crown, J. MYC as a target for cancer treatment. *Cancer Treat Rev* 94, 102154 (2021).

322. Bochum, S., Berger, S. & Martens, U. M. Small Molecules in Oncology. *Recent Results Canc* 211, 217–233 (2018).

323. Cicenas, J. *et al.* Roscovitine in cancer and other diseases. *Ann Transl Medicine* 3, 135 (2015).