

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

REGULATION OF *MYC* TRANSCRIPTION IN 3D: IMPLICATIONS FOR TUMOR DEVELOPMENT

Ilias Tzelepis



**Karolinska
Institutet**

Stockholm 2022

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by Universitetsservice US-AB, 2022

© Ilias Tzelepis, 2022

ISBN 978-91-8016-523-5

Cover illustration: A graphical summary of the gene gating principle, as uncovered by the studies of this thesis, created by Mireia Cruz De los Santos.

Regulation of *MYC* transcription in 3D: Implications for tumor development

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Ilias Tzelepis

The thesis will be defended in public at Lars Klareskog NBU1:02, Karolinska
Universitetssjukhuset Solna, Stockholm, on 25 February 2022 at 09.30 AM.

Principal Supervisor:

Anita Göndör
Karolinska Institutet
Department of Oncology-Pathology

Co-supervisors:

Barbara Scholz
Karolinska Institutet
Department of Oncology-Pathology

Galina Selivanova
Karolinska Institutet
Department of Microbiology, Tumor
and Cell biology

Opponent:

Associate Professor Ragnhild Eskeland
Oslo University
Institute of Basic Medical Sciences

Examination Board:

Professor Richard Rosenquist Brandell
Karolinska Institutet
Department of Molecular Medicine and Surgery

Associate Professor Anita Öst
Linköping University
Department of Biomedicine

Professor Ann-Kristin Östlund Farrants
Stockholm University
Department of Molecular Biosciences

To my father, for without him I would have never made it
Στον πατέρα μου, διότι χωρίς εκείνον δεν θα τα είχα καταφέρει ποτέ

“Reach what you cannot”
-Nikos Kazantzakis, Report to Greco

POPULAR SCIENCE SUMMARY OF THE THESIS

English

DNA is the molecule containing all the genetic information an individual carries. It is found in the inner space of our cells, the so-called nucleus. A question arising, though, is how such a 2 meter-long molecule can fit into the cell nucleus that has the size of a few micrometers (μm , 10^{-6} m). That happens because DNA is organized in an extremely compacted way, under the form of a structure named as chromatin. However, the organization of chromatin in the three-dimensional space of the nucleus is not random, but rather highly specific. In fact, decades of research have shown that the spatial organization of chromatin (or, the “nuclear architecture” as we could call it) plays a crucial role in determining which part of our genes is active or not; in other words, what part of our total DNA will be active at a certain time or stage of our life, or under certain conditions, such as a disease. This broad field of biology is called epigenetics and studies how the genetic information of a person collaborates with external, environmental stimuli in order to regulate gene expression, i.e. the activation of specific genes they carry. In simple terms, we could say it’s the study of how our environment and way of life can affect our genes and potentially cause a disease.

In the research covered by this thesis, we have uncovered a new molecular mechanism, previously unknown in humans, by which the nuclear architecture conspires with an external signal in order to regulate the activation of an important gene, strongly involved in cancer, and named as “MYC”. This mechanism seems to be present in cancer cells, but not in healthy cells, and is widely known as “gene gating”. More specifically, cancer cells manage to increase the expression of MYC gene and survive, by recruiting this gene to specific structures of the nuclear membrane called nuclear pores. In this way, the mRNA molecules produced from the transcription of MYC are exported quickly to the cytoplasm and thus, are protected by the rapid degradation taking place in the nucleus of the cells. That ultimately leads to high levels of MYC protein, compared to healthy cells.

Moreover, we were able to identify the molecular factors (or, “key players”) involved in this process, as well as provide the first genetic evidence of its existence in humans. The discovery of such a mechanism playing a key role in cancer, or even in the initiation of it, can help us to better understand how this complex disease works and therefore, come up with new potential therapeutic targets for cancer treatment. Finally, the new knowledge provided by this research could set the ground for the application of new strategies to fight against cancer.

Ελληνικά

Το DNA είναι το μόριο που εμπεριέχει όλη τη γενετική πληροφορία που φέρει ένα άτομο. Βρίσκεται στον εσωτερικό χώρο των κυττάρων μας, στον αποκαλούμενο πυρήνα. Ένα ερώτημα που προκύπτει, ωστόσο, είναι το πώς ένα τέτοιο μόριο μήκους περίπου 2 μέτρων μπορεί να χωρέσει στον πυρήνα του κυττάρου που έχει μέγεθος μερικών μόνο μικρομέτρων (μm , 10^{-6} m). Αυτό συμβαίνει διότι το DNA είναι οργανωμένο με έναν ακραία συμπαγή τρόπο, υπό τη μορφή μίας δομής που ονομάζεται χρωματίνη. Ωστόσο, η οργάνωση της χρωματίνης στον τρισδιάστατο χώρο του πυρήνα δεν είναι τυχαία, αλλά ιδιαίτερα συγκεκριμένη. Για την ακρίβεια, δεκαετίες ερευνών έχουν δείξει ότι η χωρική διάταξη της χρωματίνης (ή, η «αρχιτεκτονική του πυρήνα», όπως θα την αποκαλούσαμε) παίζει πολύ σημαντικό ρόλο στον καθορισμό του ποιο μέρος των γονιδίων μας είναι ενεργό ή όχι -με άλλα λόγια, ποιο κομμάτι του συνολικού μας DNA θα είναι ενεργό μία δεδομένη χρονική στιγμή ή στάδιο της ζωής μας, ή υπό συγκεκριμένες συνθήκες, όπως μία ασθένεια. Αυτό το ευρύ πεδίο της βιολογίας ονομάζεται επιγενετική και μελετά το πώς η γενετική πληροφορία ενός ατόμου συνεργάζεται με εξωγενή, ερεθίσματα του περιβάλλοντος με σκοπό να ρυθμίσει τη γονιδιακή έκφραση, δηλ. την ενεργοποίηση συγκεκριμένων γονιδίων που αυτό φέρει. Με απλά λόγια, θα μπορούσαμε να πούμε πως είναι η μελέτη του πώς το περιβάλλον και ο τρόπος ζωής μας μπορούν να επηρεάσουν τα γονιδιά μας και ενδεχομένως να προκαλέσουν μία ασθένεια.

Στην έρευνα που καλύπτει η παρούσα διατριβή έχουμε ανακαλύψει έναν νέο μοριακό μηχανισμό, μέχρι πρότινος άγνωστο στους ανθρώπους, με τον οποίο η αρχιτεκτονική του πυρήνα συνεργάζεται με ένα εξωγενές ερέθισμα με σκοπό να ρυθμίσει την ενεργοποίηση ενός σημαντικού γονιδίου, έντονα εμπλεκόμενου στον καρκίνο, και το οποίο ονομάζεται «MYC». Αυτός ο μηχανισμός φέρεται να είναι παρών στα καρκινικά κύτταρα, αλλά όχι στα υγιή κύτταρα, και είναι ευρέως γνωστός ως «γονιδιακή πύλη (gene gating)». Πιο συγκεκριμένα, τα καρκινικά κύτταρα καταφέρνουν να αυξάνουν την έκφραση του γονιδίου MYC και να επιβιώνουν, μέσω της μετακίνησης αυτού του γονιδίου σε ειδικές δομές της πυρηνικής μεμβράνης που ονομάζονται πυρηνικοί πόροι. Κατ' αυτό τον τρόπο, το μόριο του αγγελιαφόρου RNA (mRNA) που παράγονται από την μεταγραφή του MYC εξάγονται ταχεία στο κυτταρόπλασμα και επομένως, προστατεύονται από την γρήγορη αποικοδόμηση που λαμβάνει χώρα στον πυρήνα των κυττάρων. Αυτό τελικά οδηγεί στην παρουσία υψηλών επιπέδων πρωτεΐνης MYC, σε σχέση με τα υγιή κύτταρα.

Ακόμη, μπορέσαμε να ταυτοποιήσουμε τους μοριακούς παράγοντες (ή, «παράγοντες-κλειδιά») που εμπλέκονται σε αυτή τη διαδικασία, καθώς και να παράσχουμε την πρώτη γενετική ένδειξη της ύπαρξής της στον άνθρωπο. Η ανακάλυψη ενός τέτοιου μηχανισμού που παίζει σημαντικό ρόλο στον καρκίνο, ή ακόμη και στην έναρξη αυτού, μπορεί να μας βοηθήσει να κατανοήσουμε καλύτερα το πώς λειτουργεί αυτή η πολύπλοκη ασθένεια και επομένως, να βρούμε νέους πιθανούς θεραπευτικούς στόχους για αντι-καρκινική αγωγή. Τέλος, η νέα γνώση που παρέχεται από την παρούσα έρευνα θα μπορούσε να θέσει τα θεμέλια για την εφαρμογή νέων στρατηγικών καταπολέμησης κατά του καρκίνου.

Svenska

DNA är molekylen som innehåller all genetisk information som en individ bär på. Den hittas innerst i våra celler, i den så kallade kärnan. En fråga som man kan ställa sig är dock, hur kan en 2 meter lång molekyl få plats i cellens kärna som har storleken bara några få micro meter (μm , 10^{-6} m). Detta kan ske på grund av att DNA är organiserad på ett extremt kompakt sätt, under formen av en struktur som heter kromatin. Emellertid är inte organisationen av det tre dimensionella rummet av kärnan slumpartad, utan faktiskt väldigt specifik. Faktum är att årtionden av forskning har visat att organisation av kromatin (eller, "kärnans arkitektur" som man kan kalla det) spelar en viktig roll i att bestämma vilken del av våra gener som är aktiva eller inte, med andra ord; vilken del av hela vårt DNA kommer vara aktiv under en specifik del av våra liv, eller under speciella förhållanden så som vid sjukdom. Det här stora fältet av biologi kallas för epigenetik och här studerar man hur den genetiska informationen från en person samarbetar med externa miljöer. Med stimulans kan man reglera och aktivera för att se aktivitet i gener man bär. I enklare termer, kan man säga att det är en forskning där man kan se hur vår miljö och hur vårt sätt att leva, kan påverka våra gener och potentiellt orsaka en sjukdom.

I forskningen med utgångspunkt av denna tes, har vi upptäckt en ny molekyllär mekanism som hittills varit okänd hos människan. Kärnans uppbyggnad konspirerar med en extern signal för att reglera aktiveringen av en viktig gen som är starkt förknippad med cancer. Den heter "MYC". Den här mekanismen verkar finnas i cancer celler men inte i hälsosamma celler. De är mest kända som "gene gating". Mer specifikt kan man säga att cancerceller klarar av att öka uttrycket av MYC genen och överleva genom att rekrytera denna gen till specifika strukturer av kärnans membran som kallas för porer. På detta sätt kommer mRNA molekyler producerades av transkriptionen av MYC, att snabbt exporteras till cytoplasman. Denna är skyddad av den snabba degradering som sker i cellens kärna. Detta leder till slut till högre nivåer av MYC protein, jämfört med hälsosamma celler.

Utöver detta klarade vi av att identifiera de molekyllära faktorerna (eller, "nyckel spelarna") som är involverade i denna process. Men även bidra till det första genetiska beviset på dess existens i människan. Upptäckten av en sån mekanism som har en nyckelroll inom cancer, även i ett tidigt stadie, kan hjälpa oss bättre förstå hur denna komplexa sjukdom fungerar och genom det upptäcka nya potentiella terapeutiska mål för cancer behandlingar. Till slut; Denna nya kunskap som upptäckts genom denna forskning, kan bana väg för nya strategier och tillämpningar i kampen mot cancer.

ABSTRACT

This thesis uncovers how chromatin organization conspires with nuclear architecture and environmental stimuli in order to regulate gene expression in disease and particular, in cancer. In Paper I, we have unraveled a mechanism of oncogenesis, previously unknown in humans, and widely known as gene gating. Specifically, we have shown that in human colon cancer cells (HCT116) the oncogenic super-enhancer (OSE) of *MYC* increases its expression levels post-transcriptionally, by tethering *MYC* to the nuclear pore complex (NPC). This process facilitates the export of *MYC* transcripts to the cytoplasm and enables them to escape the rapid decay taking place in the nucleus. This phenomenon does not seem to be present in the healthy counterparts of these cells, human colon epithelial cells (HCECs), indicating that this is a unique feature of cancer. Moreover, our findings show that this mechanism is mediated by AHCTF1 (also known as ELYS): a mobile nucleoporin, part of the NPC, that binds on chromatin. Finally, it is also regulated by the canonical WNT signaling pathway and the complex formation between TCF4 and β -catenin, as shown by the use of the inhibitor BC21.

In Paper II, we have further explored the molecular factors involved in the gating of *MYC*, as well as provided the first genetic evidence of this mechanism in humans. More precisely, by using CRISPR-Cas9 genetic engineering we generated two different clones with a mutated CTCF binding site (CTCFBS) within the OSE and showed that their inability to efficiently bind CTCF is associated with reduced *MYC* mRNA export. In addition, this process confers to the wild type cells a growth advantage over the mutant cells and requires the canonical WNT signaling pathway for the recruitment of the OSE from intra-nucleoplasmic positions. Our findings furthermore indicate that once the OSE has reached a peripheral position ($<0.7\mu\text{m}$), the CTCFBS-mediated *CCAT1* eRNA activation takes place and promotes the recruitment of AHCTF1 to the CTCFBS. That will ultimately lead to the efficient tethering of *MYC* to the nuclear pores and its subsequent gating, whilst pointing out the existence of a novel WNT/ β -catenin-AHCTF1-CTCF-eRNA circuit in the regulation of pathogenic *MYC* expression.

In summary, the findings covered by the present thesis provide new insights in the regulation of oncogenic *MYC* expression by the 3D nuclear architecture and widen our understanding on the processes underlying tumor development. Such knowledge can improve the diagnosis, as well as potentially contribute to the identification of new therapeutic targets in cancer therapy.

LIST OF SCIENTIFIC PAPERS

- I. *WNT signaling and AHCTF1 promote oncogenic MYC expression through super-enhancer-mediated gene gating*. Nature Genetics, 51, 1723–1731, 2019.

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

- II. *Canonical WNT signaling-dependent gating of MYC requires a non-canonical CTCF function at a distal binding site*. Nature Communications, 13, 204, 2022.

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 and Anita Göndör*

shared first authors

CONTENTS

1	INTRODUCTION	9
1.1.	Introduction to Epigenetics	9
1.1.1.	Canalization and Plasticity in development	9
1.1.2.	Primary chromatin fibre and epigenetic modifications	9
1.1.3.	Stochasticity in epigenetic states	10
1.2.	Nuclear architecture and chromatin organisation in 3D	11
1.2.1.	The role of the nuclear architecture in the regulation of gene expression	11
1.2.2.	Nuclear pores, transcriptional memory and 3D genome organisation	12
1.2.3.	The gene gating principle	14
1.2.4.	Transcription, mRNA export and degradation machinery	15
1.3.	Chromatin crosstalk and 3D genome organizers	18
1.3.1.	Methodological advances in the analysis of 3D genome organization	20
1.3.2.	CTCF: the master weaver of the genome	23
1.3.3.	Circadian regulation of chromatin fibre interactions	24
1.4.	Chromatin organization and cancer	25
1.4.1.	<i>MYC</i> : a key player in cancer development	25
1.4.2.	WNT signaling pathway	27
1.4.3.	Non-coding RNAs in tumor development	30
2	RESEARCH AIMS	32
3	MATERIALS AND METHODS	33
3.1.	Cell culture and treatments	33
3.2.	Editing of the <i>CCAT1</i> -specific CTCFBS by CRISPR-Cas9	34
3.3.	siRNA transfection	34
3.4.	ChIP & Co-immunoprecipitation experiments (Chromatin/protein interactions)	34
3.4.1.	Chromatin immunoprecipitation (ChIP-qPCR and ChIP-seq)	34
3.4.2.	Co-immunoprecipitation and Western blotting	35
3.5.	Nodewalk	35
3.6.	<i>In Situ</i> methods	36
3.6.1.	Microscopy	36
3.6.2.	ChrISP	36
3.6.3.	3D DNA FISH	37
3.6.4.	RNA FISH	38
3.6.5.	In Situ Proximity Ligation Assay (ISPLA)	38
3.7.	RNA assays	39
3.7.1.	Pulse labeling of RNA & mRNA export assay	39
3.7.2.	Decay analyses	40
3.7.3.	Analysis of transcription & RNA-seq	40

3.8. Bioinformatic analyses.....	40
3.8.1. Computer simulation modeling of <i>MYC</i> mRNA cytoplasmic levels over time	41
3.8.2. Off-target whole genome sequencing	41
3.8.3. ChIP-seq & RNA-seq analysis.....	41
4 RESULTS	42
4.1. Members of the Nuclear Pore Complex interact with a chromatin network organized by <i>MYC</i> enhancers	42
4.2. The proximities between NUP133 and the OSE are dynamic, while <i>MYC</i> - OSE proximities correlate with the nuclear architecture	42
4.3. <i>MYC</i> expression is regulated post-transcriptionally by a gene gating mechanism.....	43
4.4. The super-enhancer-mediated gating of <i>MYC</i> is regulated by AHCTF1	44
4.5. WNT signaling mediates the recruitment of <i>MYC</i> to the nuclear pores and the export rate of <i>MYC</i> transcripts through β -catenin	45
4.6. Colon cancer cells gain an excessive growth advantage due to the binding of CTCF to the OSE of <i>MYC</i>	46
4.7. The increased nuclear export rate of mRNAs produced from <i>MYC</i> and <i>FAM49B</i> genes is controlled by the <i>CCAT1</i> -specific CTCFBS within the OSE... ..	47
4.8. β -catenin and CTCF recruit AHCTF1 to the OSE to facilitate its ability to reach the nuclear pore	48
4.9. The <i>CCAT1</i> -specific CTCFBS affects the <i>MYC</i> -OSE proximity at the nuclear periphery, but does not influence their overall interaction frequencies.	49
4.10. <i>CCAT1</i> eRNA expression is activated by WNT through the CTCFBS to induce the juxtaposition of the OSE to the nuclear periphery	50
5 DISCUSSION	51
5.1. Gene gating mechanism in humans	51
5.2. The role of the NPC in the regulation of gene expression	51
5.3. A novel role for CTCF.....	52
5.4. Involvement of a lncRNA, <i>CCAT1</i> , in the post-transcriptional regulation of gene expression	53
5.5. A new example of how chromatin responds to external signals	53
5.6. Potential involvement of other factors in gene gating	54
5.7. Reproducibility and stochastic chromatin networks driven by <i>MYC</i>	55
5.8. Advances in the understanding of and the fight against cancer.....	55
5.9. Promises for new cancer treatment.....	56
6 CONCLUSIONS	58
7 POINTS OF PERSPECTIVE	59
8 ACKNOWLEDGEMENTS	61
9 REFERENCES	66

LIST OF ABBREVIATIONS

3C	Chromatin Conformation Capture
3D	Three-dimensional
4C	Circular Chromatin Conformation Capture
5C	Chromosome Conformation Capture Carbon Copy
AHCTF1	AT-hook containing transcription factor 1
APC	Adenomatous polyposis coli
ARM	Armadillo repeats
ATAC-seq	Assay for Transposase-Accessible Chromatin using sequencing
ATP	Adenosine 5'-triphosphate
BAC	Bacterial Artificial Chromosome
BENC	Blood Enhancer Cluster
BMAL1	Brain and muscle Arnt-like protein-1
Bp/Kbp/Mbp	Base pairs/ Kilo-base pairs/ Million base pairs
BRD4	Bromodomain-containing protein 4
BRG1	Brahma-related gene 1
CAPTURE	CRISPR Affinity Purification in situ of Regulatory Elements
CARLo-5	Cancer-Associated Region lncRNA 5
CBC	Cap-Binding Complex
CBF β -SMMHC	Core Binding Factor β and Smooth-Muscle Myosin Heavy Chain
CBP	CREB binding protein
CCAT1/2	Colon cancer associated transcript 1/2
CDK8/9	Cyclin-Dependent Kinase 8/9
cDNA	Complementary DNA
ChIA-PET	Chromatin Interaction Analysis with Paired-End-Tag sequencing
ChIP	Chromatin immunoprecipitation
ChIP-chip	Chromatin immunoprecipitation on chip
ChIP-seq	Chromatin immunoprecipitation-sequencing
ChrISP	Chromatin In Situ Proximity
CK1	Casein Kinase 1
cLADs	Constitutive Lamina-Associated Domains
CLOCK	Circadian Locomotor Output Cycles Kaput
Co-IP	Co-immunoprecipitation
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats

CRM1	Chromosomal Region Maintenance 1
CRY	Cryptochrome
CT	Chromosomal territories
CTBP	C-terminal Binding Protein
CTCF	CCCTC-binding factor
CTCFBS	CTCF Binding site
CTD	Carboxy-Terminal Domain
DamID	DNA adenine methylation Identification
dCTP	2'-Deoxycytidine-5'-triphosphate
DHS	DNase Hypersensitivity
DIG	Digoxygenin
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DNA FISH	DNA Fluorescence In Situ Hybridization
DNMT	DNA Methyltransferase
dUTP	2'-Deoxyuridine, 5'-Triphosphate
DVL	Disheveled protein
EJC	Exon Junction Complex
EnhD	Enhancer D
eRNA	Enhancer RNA
EU	5'-ethynyl uridine
EZH2	Enhancer of Zeste Homolog 2
FAM49B	Family with sequence similarity 49 member B
FBXW7	F-box and tryptophan-aspartic acid (WD) repeat domain containing 7
FG NUPs	Phenylalanine-Glycine-rich Nucleoporins
fLADs	Facultative Lamina-Associated Domains
FPS	False-Positive Signal
G4	G-quadruplex
GANP	Germinal center-associated nuclear protein
GEO	Gene Expression Omnibus
GFP	Green Fluorescent Protein
GPSeq	Genomic loci Positioning by Sequencing
GSK3 β	Glycogen synthase kinase 3 β
GWAS	Genome-Wide Association Studies
H19 ICR	H19/IGF2 Imprinting Control Region

H2A/B, H3, H4	Histone 2A/B, Histone 3, Histone 4
H3K27ac	Histone 3 Lysine 27 acetylation
H3K27me3	Histone 3 Lysine 27 tri-methylation
H3K36me3	Histone 3 Lysine 36 tri-methylation
H3K4me1/3	Histone 3 Lysine 4 mono/tri-methylation
H3K9me2/3	Histone 3 Lysine 9 di/tri-methylation
HAT	Histone Acetyltransferase
HCECs	Human Colon Epithelial Cells
HCT116	Human Colon Tumor cells
HDAC4	Histone Deacetyltransferase 4
HeLa	Henrietta Lacks cell line
Hi-C	High-throughput Chromosome Conformation Capture
Hi-PLA	High-throughput imaging proximity ligation assay
hnRNPK	Heterogeneous Nuclear Ribonucleoprotein K
HOTAIR	HOX antisense intergenic RNA
HOX	Homeobox
HuR	Human RNA-binding protein
HYX	Hyrax
IFN- γ	Interferon gamma
Igh	Immunoglobulin heavy locus
IGV	Integrative Genomics Viewer
ISPLA	In Situ Proximity Ligation Assay
ISWI	Imitation SWI
LADs	Lamina-Associated Domains
LBR	Lamin B Receptor
LMNA	Lamin A/C
lncRNA	Long non-coding RNA
LOCKS	Large Organized Chromatin Lysine Modifications
LRP5/6	Low-density lipoprotein receptor-related protein 5/6
MALAT1	Metastasis associated lung adenocarcinoma transcript 1
MAX	MYC-Associated Factor X
MBD-R2	Methyl-CpG binding domain-R2
MED12	Mediator of RNA polymerase II transcription subunit 12
MESCs	Mouse Embryonic Stem Cells
miRNA	Micro RNA

MIZ1	MYC interacting zinc finger protein 1
mRNA	Messenger Ribonucleic Acid
mRNP	mRNA-protein complex
MYC (or c-MYC)	Myelocytomatosis proto-oncogene
NADs	Nucleolus-associated domains
ncRNA	Non-coding RNA
NES	Nuclear Export Signal
Neg	Negative
NGS	Next Generation Sequencing
NHE III ₁	Sodium–hydrogen exchanger 3
NMD	Nonsense-Mediated (mRNA) Decay
NPC	Nuclear Pore Complex
NPM1	Nucleophosmin 1
NSL	Non Specific Lethal
nt	Nucleotides
NUP	Nucleoporin
NXF1	Nuclear RNA Export Factor 1
NXT1	NTF2-related export protein 1
OSE	Oncogenic Super-Enhancer
PABPN1	Polyadenylate-Binding Nuclear Protein 1
PAF1	RNA Polymerase II Associated Factor 1
PAM	Protospacer Adjacent Motif
PARIS	Protein And RNA Isolation System
PARP1	Poly(ADP) Ribose Polymerase 1
PARylation	Poly(ADP)ribosylation
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
PER	Period
PKC	Protein Kinase C
POMs	Pore Membrane proteins
PP2A	Protein Phosphatase 2A
PRC1	Protein Regulator of Cytokinesis 1
PRC2	Polycomb Repressive Complex 2
PVT1	Plasmacytoma Variant Translocation 1
qPCR	Quantitative Polymerase Chain Reaction

Ran-GTP	GTP-binding RAs-related Nuclear protein
RCA	Rolling-Circle Amplification
RIC-seq	RNA In situ Conformation sequencing
rISH-PLA	RNA Whole-Mount In Situ Hybridization Proximity Ligation Assay
RNA	Ribonucleic Acid
RNA FISH	RNA Fluorescence In Situ Hybridization
RNA Pol II	RNA Polymerase 2
RNA-seq	Ribonucleic Acid-Sequencing
rRNPs	tRNA-protein complexes
RT	Room temperature
RT-qPCR (qRT-PCR)	Reverse transcription qPCR
SCF	SKP1–Cullin-1–F-box protein
SEN1/2	Sentrin-specific protease 1/2
Ser33/37/45	Serine 33/37/45
sgRNA	Single guide RNA
siRNA	Small interfering RNA
SNHG15	Small Nucleolar RNA Host Gene 15
snRNPs	Small nuclear Ribonucleoproteins
SNV	Single-nucleotide variant
sPom121	Soluble Pore Membrane 121
SR proteins	Serine and arginine-rich proteins
SRp20	Serine and arginine-rich protein 20
SUMOylation	Small Ubiquitin-like Modifier modification
SWI/SNF	Switching (SWI) and sucrose non-fermenting (SNF) factors
T58	Threonine 58
TAD	Topologically Associated Domain
TBP	TATA-Binding Protein
TCF/LEF	T-cell factor/Lymphoid enhancer factor
TCF4	Transcription Factor 4
TCF7L2	Transcription factor 7-like 2
TF	Transcription Factor
TFIIH	Transcription factor II Human
TGF- β (beta)	Tumor Growth Factor β
Thr41	Threonine 41
TIP60	Tat-interactive protein, 60 kDa

TLE1	Transducin-like enhancer protein 1
Tn5	Transposase 5
TOP1/2B	Topoisomerase 1/2B
TPR	Translocated promoter region
TREX, TREX-2	TRanscription and Export protein, 2
tRNAs	Transfer RNAs
Trx/MLL	Trithorax/Mixed Lineage Leukemia
U2OS	Human Bone Osteosarcoma Epithelial cells
UTR	Untranslated Region
WNT	Wingless/Integrated
WRE	Wnt-Responsive Element
WT	Wild Type
Xrn2	5'-3' Exoribonuclease 2

1 INTRODUCTION

1.1. Introduction to Epigenetics

1.1.1. Canalization and plasticity in development

Life in nature is not static. Organisms tend to adapt to environmental changes by sensing and responding to external cues. The same genetic background can thus give rise to distinct phenotypes based on the internal and external stimuli they are exposed to^{1,2}. The short-term impact of the environment on the phenotype of an individual is mainly regulated by certain mechanisms that represent a regulatory layer on top of the genetic material. Such mechanisms are called epigenetic and the broader field of biology studying these processes is termed epigenetics. Phenotypic differences between monozygotic twins³ and the different castes in the societies of individual honeybee colonies⁴ are typical examples of epigenetic regulation in nature.

Conrad Waddington⁵ was the first to describe the effect of epigenetics on the phenotype in his famous epigenetic landscape model. According to this model, an embryonic stem cell can initially differentiate to any cell type through a canalization process. However, once it has entered into a certain canal, the cell fate choices it can make become more and more restricted during the process of differentiation, reaching finally an ultimate cell state representing a mature cell type. In other words, regulators of differentiation during development act like a gravitational force driving a rolling ball to the bottom level of a sloping surface. Waddington thus defined canalization as a process ensuring that a cell would robustly follow a path to a distinct mature cell fate despite being exposed to variable conditions in its environment. He has also introduced the term developmental plasticity that refers to the ability of the genotype to give rise to different cellular phenotypes and support transitions between different cellular differentiation stages and adaptation. Waddington thus recognized the necessity of a regulatory layer that enables communication between the genotype and the environment and termed it epigenetics. He suggested that epigenetic principles underlie developmental plasticity and contribute to canalization. Research during the recent decades has explored the molecular basis of epigenetic mechanisms and uncovered that alterations in the regulation of the epigenome contribute to divergence from normal development towards diseases, such as cancer.

1.1.2. Primary chromatin fibre and epigenetic modifications

Although they may seem as two opposing concepts, canalization and plasticity during development have been suggested to represent two sides of the same coin⁶ and being regulated by chromatin, constituting complexes between DNA and proteins. Briefly, a primary chromatin fibre consists by a number of nucleosomes, whose main core is formed by an octamer of four different histone proteins (H2A, H2B, H3 and H4) wrapped within 147 bp of DNA. Linker

DNA sequences connect nucleosome cores to each other, giving an average nucleosome length of approximately 200 bp⁷.

Mitotically heritable chromatin modifications that regulate gene expression patterns but do not alter the DNA sequence itself are termed epigenetic marks and include DNA methylation, post-translational histone modifications, binding of chromatin architectural proteins and non-coding RNAs^{8,9}. Activating histone modifications, such as acetylation of H3 and H4 and methylation of H3K4 (ex. H3K4me3)¹⁰ increase the probability of gene activation by making the underlying DNA accessible to the transcriptional machinery. Conversely, the accumulation of repressive histone marks, such as H3K9me2/3 and H3K27me3¹⁰, promote gene silencing by creating a compact and inaccessible chromatin structure. Decades of studies have shown that the combination of different histone modifications could potentially expand strongly the information provided by the genetic code⁸. However, apart from defining different states of gene expression, the specificity of histone modification patterns is often used to identify DNA regulatory elements as well. For instance, active gene promoters are generally characterized by H3K4me3¹¹, while H3K36me3 can usually be found in actively transcribed gene bodies¹².

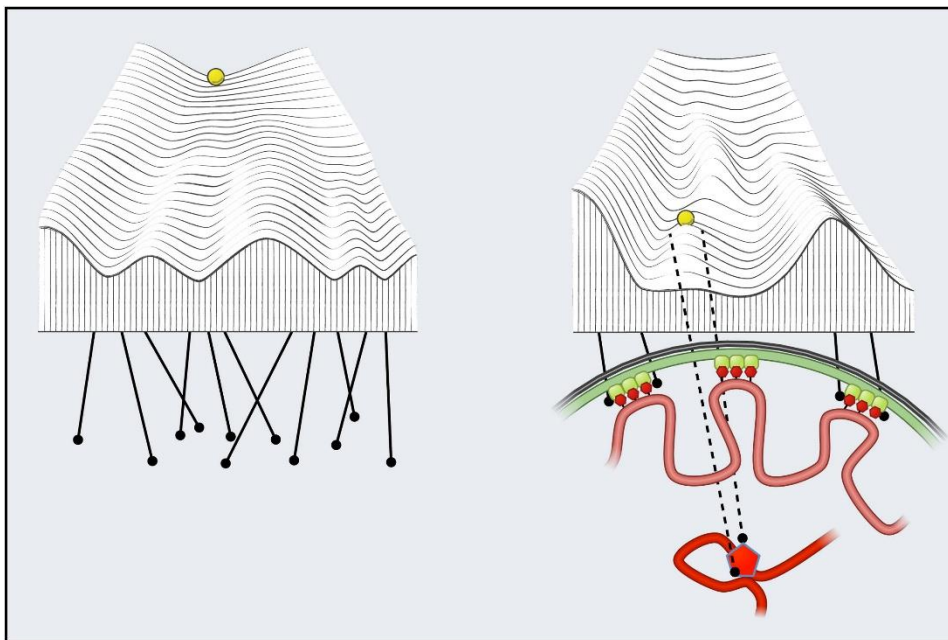


Figure 1: Regulated Noise in a Dynamic Epigenetic Landscape. Left panel: the classical Waddington representation of canalization. Right panel: the model suggested by Pujadas and Feinberg, in which the effects of transcriptional noise are modulated during development and in response to environmental signals. Red circles: chromatin modifications; Green: lamin proteins; Red pentagon: chromosome interactome mediators. Reprinted under the Creative Commons Attribution 4.0 International License, <https://creativecommons.org/licenses/by/4.0/>, by Cell, (2012), 1123-1131, 148(6). Elisabet Pujadas and Andrew Feinberg. Regulated Noise in the Epigenetic Landscape of Development and Disease.

1.1.3. Stochasticity in epigenetic states

Even though Waddington hypothesized the existence of epigenetic regulation, he also postulated that the surface of the landscape that directs cell differentiation is determined mainly

by genes⁶. Building on his metaphor, in 2012 Pujadas and Feinberg suggested a new model, in which epigenetic mechanisms would shape the surface of Waddington's landscape in part by regulating the level of stochastic fluctuations in gene expression, in a differentiation stage-specific manner⁶. According to this model, the epigenome thus does not only influence the mean level of gene expression, but, at the same time, also fine-tunes the level of gene expression variability to thereby facilitate or counteract noise-induced transitions between different cell states and thus, the control of differentiation potential. Alterations in the depth of the hills and valleys and the formation of canals are thus suggested here to be partially dictated by changes in chromatin conformations and to be under developmental control (**Figure 1**).

1.2. Nuclear architecture and chromatin organisation in 3D

1.2.1. The role of the nuclear architecture in the regulation of gene expression

The mechanisms by which chromatin regulates the level of transcriptional noise are not completely understood. However, it has been proposed that the architecture of the nucleus intertwined with the 3-dimensional (3D) folding of the genome plays important roles in this process^{6,13}. The organization of chromatin in the 3D space of the nucleus is thus neither random nor static, but rather highly dynamic with functional consequences on nuclear functions. In the interphase nucleus, chromosomes fold and occupy distinct territories, known as chromosomal territories (CT)^{14,15}. Although chromosomes can extensively intermingle with each other, chromosomal neighborhoods have been observed to have tissue- and cell-type specific preferences¹⁶. Furthermore, chromosomes display a radial organization in the nucleus¹⁷. In differentiated cells, silenced chromatin and gene-poor regions tend to occupy the space near the nuclear periphery or the nucleolus, which are generally characterized as repressive environments^{18,19}. Conversely, gene-rich regions and transcriptionally active genes tend to be positioned in the interior of the nucleus where the majority of transcription factories and nuclear speckles are localized, providing an environment rich in factors organizing the transcriptional and splicing machineries²⁰. The regions of the genome that localize to the lamina or the nucleolus are known as lamina-associated domains (LADs) or nucleolus-associated domains (NADs), respectively. LADs contain not only the constitutively silenced and AT-rich regions of the genome, which are called cell type-independent or constitutive LADs (cLADs), but also developmentally repressed genes, termed as facultative LADs (fLADs). In mammals, LADs are covering almost 40% of the genome and their size can vary from 10 Kbp to 10 Mbp²¹. The LADs in differentiated cells overlap to a large extent with long regions enriched in the repressive H3K9me2 marks, termed as H3K9me2 Large Organized Chromatin K9 modifications (H3K9me2 LOCKs)²². Although the spatial separation between transcriptionally permissive and repressive chromatin environments is quite apparent in differentiated cells, the mechanisms underlying this process and its consequences on gene expression are poorly understood. However, given that chromatin marks are reversible, separation between active

and inactive modifications has been suggested to decrease the level of stochastic fluctuations in chromatin modifications and thereby expression noise^{6,13}.

1.2.2. Nuclear pores, transcriptional memory and 3D genome organization

Localization of genomic loci to the lamina has not only been linked to the formation of repressive transcriptional memory during cell differentiation, but has also been implicated in the rapid activation of inducible genes in response to environmental signals^{23–27}. This process takes place in certain structures embedded in the nuclear envelope and known as nuclear pores. Nuclear pores are large (50 to 125 MDa, depending on the species) multi-protein complexes, known as nuclear pore complexes (NPCs), which consist of about 30 different proteins termed nucleoporins or Nups²⁸ (**Figure 2**). A subset of these proteins is anchored permanently at the nuclear membrane and called “Poms”, while the majority of them are soluble peripheral proteins²⁸. The NPC is composed by two main structures: the cytoplasmic filaments forming the cytoplasmic side of nuclear pores and the basket-like structure formed by the proteins residing only at the nuclear side of the NPC. Its shape is characterized by an 8-fold rotational symmetry and contains at least eight copies of each Nup (500-1000 individual proteins in total)^{28–34}. Moreover, instead of being continuously embedded at the NPC, some nucleoporins like AHCTF1/ELYS have been found to demonstrate extensive mobility within the nucleus; a characteristic that could attribute new interest regarding their role³⁵. In addition to its mobility, AHCTF1 has also been shown to both interact with chromatin and function as a key organizer of the pre-nucleopore complex³⁶.

Beyond the regulation of nucleo-cytoplasmic transport, nuclear pores have been reported to involve more functions, including RNA processing and control, DNA repair, as well as regulation of 3D genome organization and gene expression³⁷. In addition, nucleoporins have been participating in the regulation of phenotypic plasticity in response to environmental stimuli, and the regulation of pluripotency and differentiation^{24,27}.

Such features have been attributed to the ability of nucleoporins to bind a variety of genes and regulatory elements in *Drosophila* and mammalian cells, leading to either activation or repression events^{21,37,38}. A nucleoporin widely studied in the regulation of gene expression is represented by NUP98²³. ChIP-seq analysis thus revealed that NUP98 binds to active developmentally regulated genes during the differentiation of human embryonic stem cells²⁵. This observation is in line with previous studies in *Drosophila* showing that Nup98 associates with several active genes, including Hox genes³⁹. The recruitment of Nup98 to those genes is facilitated by the interaction between Nup98 and the histone-modifying complexes MBD-R2/NSL and Trx/MLL, highlighting the involvement of this nucleoporin in epigenetic modulation^{37,39}. Other nucleoporins in *Drosophila*, notably Nup50 and Nup62, interacted with transcriptionally active genes mainly involved in developmental regulation and cell cycle²³. Of considerable interest, NUP50 seems to play a similar role in humans as well, as this highly mobile nucleoporin associates with sites of active transcription marked by RNA Polymerase II

in myoblasts to contribute to myotube differentiation⁴⁰. However, the mechanism(s) by which NUPs and NPCs might contribute to gene activation is not completely understood.

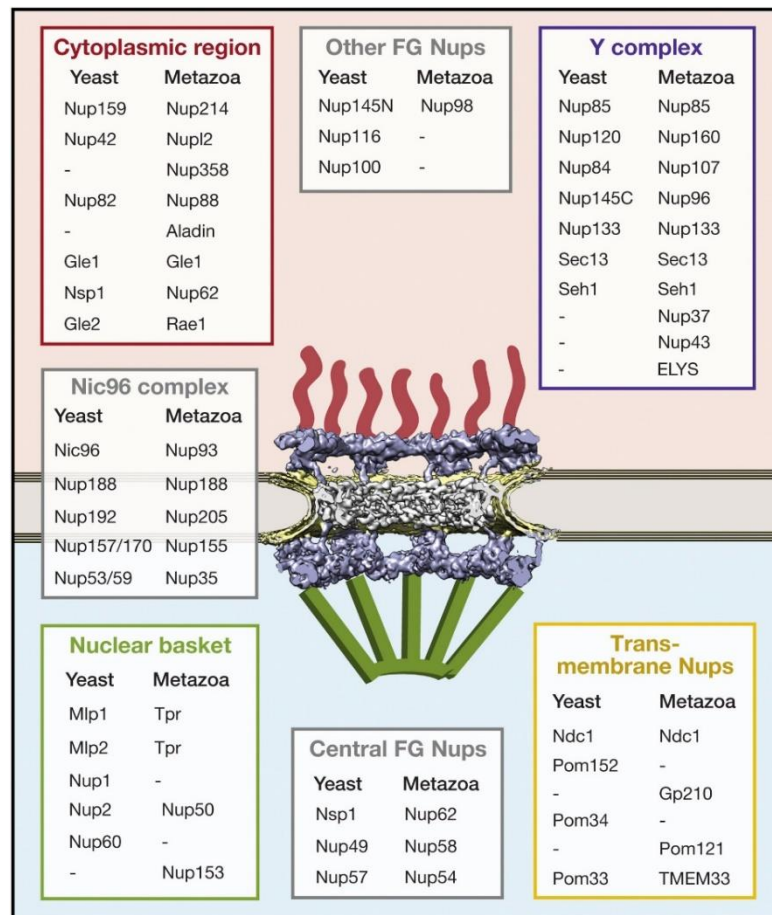


Figure 2: Structure and components of the Nuclear Pore Complex. Reprinted under the Creative Commons Attribution 4.0 International License, <https://creativecommons.org/licenses/by/4.0/>, by Cell (2016), 1162-1171, 164 (6). Kevin E.Knochenhauer, Thomas U.Schwartz. The Nuclear Pore Complex as a Flexible and Dynamic Gate.

Apart from gene activation, the role of NUPs in the formation of silent chromatin and transcriptional repression has been well-established²¹. For example, NUP153 has been recently reported to play a key role in the control of pluripotency in mouse embryonic stem cells (mESCs) by inducing the recruitment of the Polycomb-repressive complex 1 (PRC1) to a subset of developmental genes to maintain their repression²⁷. Similarly, NUP155 was found to physically interact with another repressive chromatin modifier, HDAC4, in cardiomyocytes⁴¹. In fact, it was proposed that HDAC4 regulates gene expression by altering the chromatin association of NUP155 at certain genes, implying that NUP155-chromatin interaction might be necessary for HDAC4-mediated repression. Moreover, ChIP-chip analysis of NUP93 chromatin binding in HeLa cells showed that the genome-wide binding sites of this nucleoporin are enriched in silent histone methylation marks⁴². Finally, an example demonstrating that certain nucleoporins are involved in both transcriptional activation and repression is provided by NUP98 in *Drosophila*²³. DNA adenine methylation Identification (DamID) analysis thus

classified the genomic regions bound by Nup98 into off-pore and on-pore groups. While the first group was mainly associated with actively transcribed genes, the latter group was enriched in transcriptionally inactive regions of the genome, including chromatin insulator-binding sites⁴³. In conclusion, these observations together seem to point towards the general hypothesis that nuclear pores might have dual roles in transcriptional regulation to potentially provide a platform for the transitions between different transcriptional states.

Recently, an increasing number of studies has implicated NUPs in the formation of transcriptional memories, not only in yeast, but also in mammalian cells. For example, Interferon gamma (IFN- γ)-inducible genes maintain their transcriptional memory via the persistent acquisition of H3K4me2 marks at their promoters, linked to the recruitment of nucleoplasmic NUP98 and a poised RNA Polymerase II²⁶. Similarly, a variant transmembrane nucleoporin Pom121 (sPom121) does not localize to the nuclear pores, but instead interacts with nucleoplasmic NUP98 at the promoters of its target genes and controls the initiation of their transcription in human cells⁴⁴.

Overlayed on these observations, chromatin organization within the 3D space of the nucleus can contribute to the formation of transcriptional memories. In yeast, this is exemplified by the formation of chromatin loops between the promoters and terminators of certain genes; a configuration that facilitates transcriptional re-initiations⁴⁵. Intriguingly, the formation of such gene loops involves NUPs while tethering one or both gene ends at the NPC⁴⁶. Moreover, analysis of the promoters in a group of induced yeast genes with transcriptional memory identified a DNA motif that drives gene-specific inter-chromosomal clustering at nuclear pores⁴⁷. Tethering of chromatin to the nuclear pores in *Drosophila* seems to be regulated by a negative loop formed by Nup155 and Nup93⁴⁸. These two nucleoporins recruit Nup62 at the NPC in order to suppress excessive chromatin tethering by Nup155, thereby controlling large-scale chromatin organization at the nuclear pores. Taken together, these studies implicate NPCs as direct mediators of the 3D folding of the genome to provide a platform promoting transcriptional activation even in the repressive environment of the nuclear periphery^{45,49}. However, far less is known about the role of NUPs and NPCs in the 3D organization of the mammalian genome.

1.2.3. The gene gating principle

In 1985 Guenter Blobel suggested that nuclear pores may participate in the regulation of gene expression at the post-transcriptional level, by coordinating mRNA expression, processing and nuclear export in yeast⁵⁰. This process, termed gene gating, has gained support through the years, not only in yeast, but also in *Drosophila* and *C. elegans*^{28,51,52}. However, it still remains unknown whether such a mechanism takes place in humans, even though multiple reports support a novel role of nucleoporins and NPCs in the regulation of a variety of nuclear processes also in mammalian cells⁵³. A fundamental difference between mammalian and yeast/*Drosophila* nuclei is represented by the nuclear volume, as well as a much higher

sequence complexity. Gene gating involves, not only the recruitment of inducible genes to the nuclear pores, but also the facilitated and rapid export of their transcripts to the cytoplasm. In that context, several questions arise, such as how this principle would negotiate the dramatically increased nuclear volume, what are the molecular factors that might coordinate or be part of this process and whether this phenomenon could be regulated by external stimuli, like signaling pathways and circadian rhythm. Finally, it is still an enigma whether this process involves a certain number of genes or it would be rather genome-wide.

1.2.4. Transcription, mRNA export and degradation machinery

A potential gene gating mechanism would likely need to be tightly inter-connected with the mRNA processing and export machineries. Another key question addresses how such a link with escape degradation processes.

During elongation of transcription by the RNA Polymerase II complex, a pre-mRNA molecule is produced. During this process, various proteins, such as SR proteins and snRNPs, rapidly associate with it, in a highly ordered assembly manner²⁹. This results in the generation of a pre mRNA-protein complex termed pre-mRNP. The structure of pre-mRNPs continuously changes while new proteins and additional mRNA from the ongoing transcription are associated to the complex. The associated proteins of an mRNP are crucial, since they can define many aspects of its fate, such as export and localization in the cytoplasm, coupling with the translation machinery, stability and degradation. The conventional consensus is that once the mRNP is released from its gene, it enters the interchromatin compartment and moves by diffusion. Subsequently, mRNPs which are export competent dock at the NPCs and are exported to the cytoplasm.

1.2.4.1. Regulation of transcription

The nuclear compartmentalization creates microenvironments that can profoundly influence the coordination, efficiency and regulation of transcription⁵⁴. It is generally believed that the RNA molecules play a role in the formation of these compartments, since they are able of attracting and maintaining freely diffusible components from the nucleoplasmic pool, resulting in reversible membrane-free structures, called droplets⁵⁵⁻⁵⁷. Therefore, such nuclear compartments generally form at active genes, which moreover may constitute transient microenvironments favourable for mRNP formation. Prior to transcriptional activation, the chromatin is extensively unfolded and modified, enhancing their abilities to screen their neighborhood to form loop structures. This may subsequently lead to transient physical proximity of the transcribed genes and their enhancers, resulting in statically assembled structures, known as transcription factories⁵⁸. However, it still remains unclear what is the exact role of chromatin movements in the regulation of transcription, as well as what is the functional significance of the establishment of such a microenvironment for transcription.

As the transcription occurs, multiple proteins bind the newly synthesized transcript, leading to the pre-mRNP assembly. It has been suggested that the RNA Pol II elongation complex, and the C-terminal domain (CTD) of its largest subunit in particular, plays a key role here by coordinating the interactions between the transcript and specific proteins⁵⁹. These features coordinate the relationship between the chromatin structure and the transcription rate to subsequently influence splice site choices^{60,61}. Overlayered on these principles, the deposition of histone variants and nucleosome density can regulate the movement of RNA Pol II during transcription⁶², while histone modifications can influence both transcriptional rate and the recruitment of splicing factors^{63,64}. These complex scenarios are further compounded by the observations that small and long non-coding RNAs might play a crucial role in local alterations of chromatin modifications that can affect the splicing pattern⁶⁵.

1.2.4.2. Pre-mRNA processing at the gene

Pre-mRNA molecules are submitted to several steps of processing during and after transcriptional termination. During transcription, they start folding in alternative ways, following energy rules^{66,67}, in manners that can affect their interactions with *trans*-acting factors, and the transcription rate^{68,69}. While the pre-mRNA is still associated with its template, capping enzymes recognize its 5' end and the nuclear cap-binding complex proteins bind to the cap forming the CBC⁷⁰, which will later be involved in the initial translation at the cytoplasm⁷¹.

After the capping and while the pre-mRNA is still transcribed, the spliceosome assembly takes place in order to perform intron excision⁷². During splicing, introns are removed in an overall 5' to 3' end order, but not all introns are excised while associated with their templates. Some pre-mRNAs lose their introns at the interchromatin instead^{73,74} and this variation could reflect differences in splicing kinetics, alternative splicing regulation and the length of the gene^{73,75}. Another crucial step is the binding of SR proteins, which is partially sequence-specific. The phosphorylation of these factors is important for their recruitment to the pre-mRNA and their presence is necessary for constitutive splicing, while they can also affect alternative splice site choices⁷⁶. SR proteins are important for efficient export too, since they serve as export adaptors for NXF1, which is the main mRNP export receptor. Finally, they can also regulate 3' processing of pre-mRNAs, translation initiation and mRNA stability. The last step of splicing involves the deposition of the EJC core at exon-exon junctions. The EJC core itself plays a central role in several post-transcriptional procedures^{77,78}, by controlling the recruitment of different proteins, such as export adaptors, translation initiation factors, as well as the constitution of a functional nonsense-mediated decay (NMD) complex.

Apart from capping and splicing, most mRNAs are cleaved and polyadenylated. This process is performed by the 3' cleavage and polyadenylation machinery. This step is essential, since proper 3' end formation is required for mRNA export⁷⁹. The polyadenylation machinery is recruited to the pre-mRNA via the RNA Pol II CTD and specifically, the TREX complex, which seems to be necessary for both 3' processing and release of the transcript from its template^{80,81}. The length of the poly(A) tail in mammals is controlled by PABPN1, which is deposited upstream the poly(A) signal of the mRNA, potentially also requiring NPM1⁸².

Finally, transcriptional elongation is terminated, based on the recognition of proper transcription termination signals⁸³. These can vary and must discriminate between properly completed, stalled or prematurely terminated transcriptional elongation. Transcriptional termination, which can be influenced by several conditions, such as cancer⁸⁴, can affect alternative poly(A) site usage. This is a crucial event as transcript isoforms with different 3' UTR lengths demonstrate differential stability, translation properties and localization in the cytoplasm⁸⁵.

1.2.4.3. Quality control and degradation

To handle defective transcripts generated during transcription or upon deficient post-transcriptional processing, the cells have evolved elaborate quality controls. Thus, transcripts with errors in splicing, capping, assembly and 3' end formation are all led to degradation⁸⁶. Degradation can occur at the gene, either by the exosome or by a 5'-3' exonuclease such as Xrn2⁸⁷, while transcripts with retained introns or faulty 3' ends are retained at the gene^{88,89}, which either subjects them to degradation⁹⁰ or provides more time to enable completion of processing⁹¹.

1.2.4.4. mRNPs in the interchromatin compartment

Although splicing can be initiated during transcription, it is also often completed post-transcriptionally following the release of the primary transcript from its template⁷³. Numerous clusters of granules called interchromatin granule clusters or speckles are located in the interchromatin and contain high concentrations of spliceosome components⁹². The juxtaposition of active genes at the surface of granule clusters may facilitate the recruitment of the spliceosome to the nascent transcripts⁹³. In addition, the synthesis of the poly(A) tail can also be completed in the space of the interchromatin compartment²⁹.

Importantly, mRNA molecules can be subjected to a variety of chemical modifications in the interchromatin, such as 5-methylcytosine, N⁶-methyl-adenosine, pseudouridine, 5-hydroxymethylcytosine and N¹-methyl-6adenosine⁹⁴⁻⁹⁶. These modifications are essential, since they can define the function and fate of the mRNAs, change their conformation and block or promote key RNA-protein interactions. The enzymes that mediate or remove these modifications are usually localized in the nucleus and frequently associated with the TREX complex responsible for the nuclear export of fully mature mRNAs²⁹.

The interchromatin compartment covers almost the 50% of the nuclear volume and is often connected to the nuclear pore complex through the formation of an irregular network of narrow channels²⁹. Differences in the density of chromatin structures can significantly influence the movement of mRNPs, which move inside the nuclei by diffusion⁹⁷⁻¹⁰⁰ to reach the nuclear pores, following a kinetics ranging from 6 to 50 min in diploid cells^{101,102}, with small mRNPs moving faster than large ones¹⁰¹.

1.2.4.5. mRNPs at the nuclear pore complex and export

The nuclear export of the mRNPs is finally manifested at the nuclear pore complex (NPC), which is responsible for almost all bidirectional transport of molecules between the cytoplasm and the nucleus. During their export through the NPC, the mRNPs are initially docked onto the nuclear basket, then translocated through the central channel and finally, released from the cytoplasmic fibrillar structures²⁹. The time an mRNP spends at the NPC until its export is completed can vary from 12 ms to several seconds^{101,103–107}.

The key nucleoporins contain domains rich in phenylalanine (F) and glycine (G) residues, which are flexible, and fill the central channel of the nuclear pores, forming a barrier preventing passive diffusion for molecules larger than 30–40 kDa. In addition, extensive unfolding might be required for large mRNPs to enable their passage through the NPC. Upon exit from the central channel, mRNPs first associate with the cytoplasmic part of the NPC, which may serve as a cytoplasmic platform¹⁰⁸, followed by the removal of export factors, which are eventually recycled back to the nucleus¹⁰⁹. At this point, the mRNPs are unable to make the journey back into the nucleus, due to their compositional and conformational change which promotes directionality. Once in the cytoplasm, mRNP molecules are again remodeled, in a process essential for their stability and translation¹¹⁰.

During ongoing transcription, the primary transcripts are being prepared for export via association with export receptors. These receptors bind to mRNPs through export adaptor proteins. Most mRNPs use the export receptor NXF1/NXT1 heterodimer^{111–113}, while the export of snRNPs, microRNAs, tRNAs and rRNPs require the receptor CRM1, in association with Ran-GTP¹¹⁴. The Human antigen R (HuR) protein, which includes a Nuclear Export Signal (NES), binds to these molecules and, in turn, will bind CRM1²⁹. For a subset of mRNAs, a 50 nucleotides-long sequence in their 3' UTR is sufficient to direct these transcripts into a CRM1-dependent export pathway¹¹⁵. NXF1 and CRM1 both interact with different sets of export adaptors. Some of these, such as ALY/REF, belong to the TREX complex with functions in transcription elongation, genome stability and mRNA export^{111,116}, and the TREX-2 protein complex, that is known to interact with the NPC basket^{117,118}. An essential subunit of the TREX-2 complex is GANP, which can interact directly with NXF1 to promote rapid changes in gene expression¹¹⁹. Lastly, also SR proteins, such as 9G8 and SRp20, can serve as export adaptors to compound the complexity of this process^{120,121}. In conclusion, the existence of alternative export pathways and their formation reflects on the various principles of recruiting different export adaptors and the subsequent binding of the export receptors, to likely facilitate the export of selected mRNA subpopulations^{122,123}.

1.3. Chromatin crosstalk and 3D genome organizers

Layered on top of the radial genome organization, dynamic chromatin fibre interactions can take place within the 3D context of the nuclear architecture¹²⁴. In agreement with the spatial

separation between active and inactive domains, maps of chromatin fibre contact probabilities documented that active regions tend to interact with other active regions, and inactive regions tend to be in close proximity to other inactive regions⁸ (**Figure 3**). Dynamic physical contacts between distant regions diversify gene expression patterns, respond to environmental signals and require chromatin mobility and protein complexes that transiently stabilize chromatin fibre interactions^{125,126}. Interactions between chromatin regions in *cis*, i.e. within the same chromosome, are more frequent and lead to the formation of local chromatin loops to participate in the formation of topologically associated domains (TADs)¹²⁵. Transient contacts between regions in different chromosomes, i.e. in *trans*, form chromatin bridges and require large-scale chromatin fibre movements²¹. Chromatin looping has been extensively studied for certain loci and has been linked to many genomic functions¹²⁷. Interestingly, the most important and well-known role of chromatin loop formation is to regulate gene expression by bringing in close proximity distal regulatory elements of the genome¹²⁸. The most frequent *cis*-regulatory elements involved in chromatin loops are enhancers, insulators and gene promoters¹²⁹.

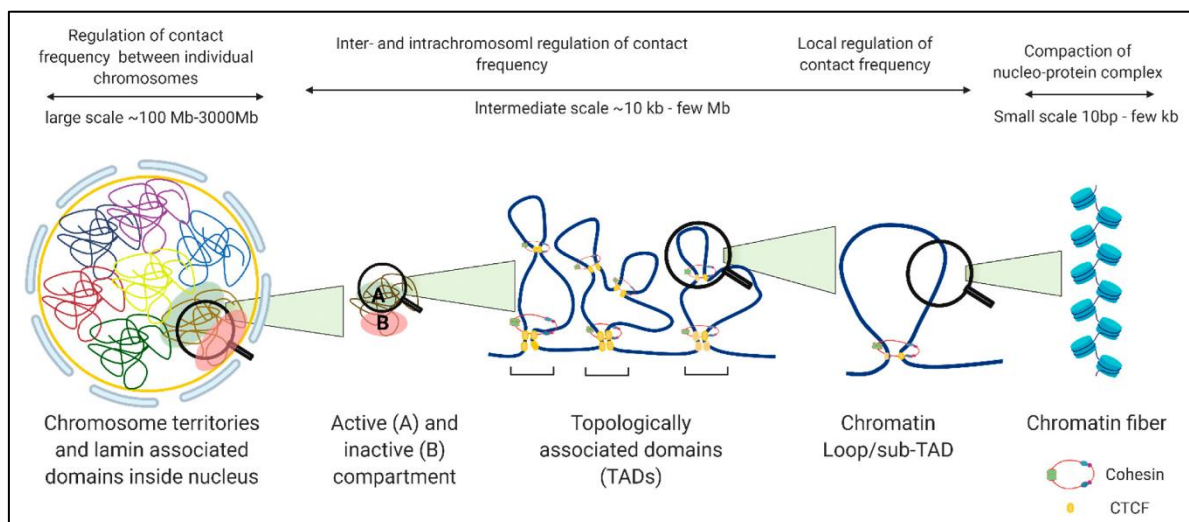


Figure 3: Schematic representation of 3D genome organization in the nucleus of interphase cells¹³⁰. Reprinted under the Creative Commons Attribution 4.0 International License, <https://creativecommons.org/licenses/by/4.0/>, by Biology 2021, 10(4), 272. Aktan Alpsoy,,Surbhi Sood, and Emily C. Dykhuizen. At the Crossroad of Gene Regulation and Genome Organization: Potential Roles for ATP-Dependent Chromatin Remodelers in the Regulation of CTCF-Mediated 3D Architecture.

Enhancers increase the probability and/or rate of gene expression and have been characterized as short DNA elements with a size of 50-1500 bps, enriched in transcription factor binding sites¹²⁹. Enhancer elements might be located both close to and within, as well as in a distance from the target genes they control¹²⁹. Enhancers can be general or cell-type-specific and signal-dependent, and are in this capacity the principal regulatory elements of the genome that coordinate developmental and oncogenic pathways^{21,131–133}. Based on their local chromatin structure and function, enhancer states have been classified as active, poised or inactive¹³⁴. Active enhancers are generally characterized by enrichment in H3K27ac and H3K4me1, whereas enrichment in H3K4me1 alone and/or the presence of H3K27me3 has been linked to

a poised state^{134–138}. In contrast to H3K4me1, H3K4me3 characterizes mainly active promoters¹³⁴. Another interesting aspect of these enhancer DNA sequences is that in many cases they are transcribed by RNA Pol II to produce short non-coding RNAs (eRNAs)¹³⁹. Even though eRNAs have been proposed to promote chromatin loop formation and thus, transcriptional activation¹⁴⁰, their exact role and mechanism of action remains to be fully elucidated. As enhancer action can regulate cell type- and differentiation stage- specific transcription, one of its main roles lies in defining cell states^{128,141,142}. In some instances, multiple enhancers are physically clustered together covering large regions in *cis* to coordinate their regulation of expression of their target genes in combinatorial manners. These elements, termed super-enhancers, ensure the robust expression of cell fate-determining genes, for example, or integrate signals from numerous cell fate-determining pathways to integrate the function of individual enhancers into a single signal output^{131,143}.

1.3.1. Methodological advances in the analysis of 3D genome organization

The study of the 3D genome organization and the crosstalk between different chromatin compartments would not have been possible without the immense advances in the available methodology through the years. Many of the methods applied in this field, such as 3C, 4C, 5C¹⁴⁴ and Hi-C, focus on the identification of physical chromatin-chromatin interactions.

Chromatin Conformation Capture or 3C was the first of the so-called “C” techniques and originally developed by Job Dekker in 2002¹⁴⁵. The 3C method was initially used to investigate the folding of chromosomes and the inter- and intrachromosomal communications between chromatin fibres predicted to interact with each other. Its methodology and the vast majority of its derivatives are based on formaldehyde cross-linking of chromatin in living cells. This stabilizes the dynamic contacts formed between distant loci interacting *in vivo*. Next, a restriction enzyme digestion step generates cohesive ends within chromatin fibre complexes, which can be ligated under very dilute conditions, provided they are physically proximal to each other, followed by PCR amplification using primers of regions suspected to interact with each other⁸.

The 3C technique can determine the physical proximity of remote sequences with the resolution of a few kilobase pairs, depending on the restriction enzyme of choice, but these interactions are determined semiquantitative at best¹²⁴. Moreover, the need for a preconceived idea of the interacting partners prompted the development of an alternative method called Circular Chromosome Conformation Capture or 4C. The 4C approach includes a circularization step that permits the identification of interacting genomic sequences by using primers annealing on the bait, i.e. a known sequence of interest⁸⁸. Thus, innovation rendered it possible to pursue physical interactions between chromatin fibres without prior knowledge or expectation of the existence of the investigated interactions in a high throughput manner. However, since the 4C method is unable to examine all to all chromatin networks, the Hi-C technique was innovated by combining proximity-based ligation, like the aforementioned methods, with massively

parallel sequencing¹⁴⁷. The Hi-C technique and its many derivatives were used to construct spatial proximity maps of the human genome, confirming the presence of sub-chromosomal domains termed TADs^{8,147}.

Even though several Hi-C protocols have been adapted to map chromatin networks in single-cell level or small cell populations, their lack of ability to quantitatively assess specific enhancer-gene communications rendered them less useful. To overcome this limitation, our research group recently developed a novel, 3C-based, method for the investigation of chromatin-chromatin interactions termed Nodewalk^{148,149}. This technique is based on the idea that the limitation of several logarithmic PCR amplification steps in the conventional “C” techniques to hamper their usefulness can be overcome by introducing a linear amplification step, by generating an RNA intermediate. Moreover, each interactor, termed node, of a single locus identified in an initial screen could then be used as secondary bait. By adding node after node, a comprehensive and genome-wide chromatin network can be generated. The big advantage of Nodewalk over other similar methods in this field is its unrivalled sensitivity with parallel high resolution, that can be used to identify stochastic interactions and dynamic changes in chromatin structures in very small cell populations. Indeed, using the Nodewalk approach, it was possible to determine that enhancer-*MYC* interactions are not only stochastic but also mutually exclusive¹⁴⁸.

Apart from the analyses of physical chromatin fibre interactions, several *in situ* methods have been employed over the years for the visualization of the localization of genes and other DNA elements in the 3D space of the nucleus. One of these methods, extensively used, is represented by the 3D DNA FISH. Upon formaldehyde cross-linking of cells or tissues and subsequent denaturation, fluorescently labeled DNA probes anneal to target regions and allow the identification of gene positions at a certain time point or under any conditions. The useful feature of this method is providing information about where the targeted regions are within the nuclear architecture when they are proximal to each other. The disadvantage of this technique is its relatively poor resolution. A comprehensive understanding of the 3D chromatin networks requires a combination of “C” and DNA FISH techniques.

Apart from the *in situ* localization of nucleic acid molecules, it is also possible to map proximities between proteins, as well as protein-DNA complexes. The In Situ Proximity Ligation Assay (ISPLA), a rolling-circle amplification-based technique, was initially developed for identifying the proximity, and therefore, the potential for physical interaction, between two different proteins. ISPLA was first developed by Ulf Landegren and his group in 2007¹⁵⁰ and since then, many other novel methods have been developed based on it, such as the rISH-PLA for the detection of RNA-protein interactions in intact cells¹⁵¹ and the recent Hi-PLA by Tom Misteli¹⁵².

By combining features of 3D DNA FISH and ISPLA, it is also possible to screen for proximities between a DNA region and a particular epitope, using the ISPLA adaptation called Chromatin In Situ Proximity (ChrISP)^{153,154}. ChrISP is an assay that translates the proximity between two different chromatin fibers or between a protein and a genomic locus in 3D into a

fluorescent signal. It is based on proximity ligation assay, but omits the rolling-circle amplification (RCA) step. Instead, the proximities of the two targets are visualized by a fluorescent connector oligonucleotide, termed splinter, which forms a circular DNA molecule upon ligation with another circle-forming oligonucleotide, termed backbone. This modification allows the detection of proximities between two epitopes by overcoming the steric hindrances from any nuclear structure. Apart from its reproducibility, the biggest advantage of ChrISP in comparison with conventional DNA FISH is the high resolution of ≤ 168 Å in single cells, making it a very useful tool in the studies of this field.

Other technical innovations include the Assay for Transposase-Accessible Chromatin with high-throughput sequencing or ATAC-seq, which addresses accessible chromatin regions and nucleosome positioning across the genome¹⁵⁵. It follows a simple two-step protocol and can be performed with a relatively small amount of cells (about 50,000 cells). The method relies on the next-generation sequencing (NGS) library preparation using the hyperactive transposase Tn5. Moreover, an increasing number of studies have been also using an assay called DamID for the mapping of *in vivo* protein-genome interactions, as an alternative to ChIP-seq. This method relies on the fusion of a protein of interest to *Escherichia coli* DNA adenine methyltransferase (dam). The resulting fusion protein preferentially methylates the adenine residues in the DNA adjacent to the native binding sites of the dam fusion partner. Since this methylation is absent in most eukaryotes, it serves as a unique tag to identify protein interaction sites genome-wide. Finally, the adenine-methylated DNA fragments are isolated by selective PCR and identified by microarray hybridization or sequencing¹⁵⁶.

Finally, two additional methods frequently used in the investigation of protein-genome interactions are represented by the ChIP-Loop and ChIA-PET. ChIP-Loop was first described in 2005 for the study of Rho Syndrome and is a combination of the standard 3C and ChIP methods¹⁵⁷. In this technique, 3C chromatin is prepared and then immunoprecipitated with an antibody against a protein of interest. While the DNA is still coupled to the beads, ligation occurs and qPCR is finally used to detect interactions of interest. The advantage of this method is it decreases the background noise of 3C experiments, while increasing the specificity by aiming on a known protein mediating the chromatin fibres interaction. Following the same idea, ChIA-PET is a method used to identify the protein mediators of 3D interactions. The chromatin interaction analysis by paired-end tag sequencing captures the chromatin fibre interactions that occur in the presence of a particular protein, like RNA Pol II^{8,128}. Interestingly, the complexity of transcriptional regulation in 3D was recently highlighted by Pol II ChIA-PET analyses, which demonstrated that in large cell populations multiple distant enhancers across the genome can impinge on a shared group of target genes¹²⁸.

In conclusion, the innovation of several groundbreaking techniques in the chromatin field has significantly broadened our understanding on how the genome is functionally organized in the 3D space of the nucleus. A few years ago, Liu and colleagues followed the widely used CRISPR genome editing approach in order to unbiasedly detect long-range DNA interactions and locus-specific chromatin-regulating protein complexes, in a novel method called

CAPTURE (CRISPR affinity purification *in situ* of regulatory elements)¹⁵⁸. In this study, the researchers were able to demonstrate high-resolution and selective isolation of chromatin fibre contacts at a single-copy genomic locus by using sequence-specific guide RNAs and an *in vivo* biotinylated nuclease-deficient Cas9 protein. The use of this assay provided new insights into the spatial features of developmentally regulated super-enhancers and disease-associated *cis*-elements controlling gene transcription. Then, Cai et al benefit from the invention of a new technology termed RNA *in situ* conformation sequencing (RIC-seq), which combines proximity ligation mediated by RNA-binding proteins and deep sequencing, to identify the genome-wide profile of intra- and intermolecular RNA-RNA interactions¹⁵⁹. Interestingly, this study suggested that the long non-coding eRNA CCAT1-5L promotes *MYC* transcription by mediating chromatin looping through the interaction with RNA derived from the *MYC* promoter and enhancer, as well as the RNA-binding protein hnRNPK. Lastly, a novel method described as genomic loci positioning by sequencing (GPSeq) and invented by Girelli and colleagues was recently used to map the radial organization of the human genome at resolution of 100-kb¹⁷. This approach successfully reported genome-wide distances to the lamina all along the nuclear radius, revealing fundamental aspects of the genome architecture in humans.

1.3.2. CTCF: the master weaver of the genome

Transient contacts between regulatory elements require chromatin architectural proteins that fold the genome into 3D structures to thereby act as 3D genome organizers¹²⁵. Some of the most important and widely studied factors of this group are cohesin, condensins and lamins. However, the 3D genome organizer protein showing the most exceptional interest is CTCF.

CTCF, or CCCTC-binding factor, is a DNA-binding factor that was initially identified as a transcription factor regulating the transcription of the *c-MYC* gene^{160,161}. It contains 11 zinc fingers and is ubiquitously expressed suggesting that it contributes to housekeeping functions^{162,163}. For several years, CTCF has been considered to be an insulator protein, whose role was to interfere with enhancer-promoter interactions¹⁶³. The most well-characterized examples of regulatory elements that rely on CTCF functions are the *H19* imprinting control region (ICR) and the beta-globin locus control region. In the case of the *H19* ICR, CTCF mediates short- and long-range intra-chromosomal and inter-chromosomal interactions across the genome, which regulate gene expression *in cis* and *in trans*¹⁴⁶. Interestingly, CTCF-binding sites have been shown to be cell type-specific, and the DNA binding ability of CTCF is often regulated by DNA methylation and might be influenced by its post-translational modifications, such as SUMOylation and poly(ADP-ribosylation)^{163–165}.

Apart from enhancer blocker and chromatin barrier functions, CTCF can act also as a transcriptional activator or repressor, it can regulate the juxtaposition between distal enhancers and their promoters and facilitate transcriptional pausing and alternative mRNA splicing¹⁶³. A critical feature underlying most of its functions in gene regulation is its ability to mediate long-range chromatin interactions and to thereby define genome topology¹⁶³. Based on analyses of

chromatin fibre interactions by the Hi-C assay, the genome is organized into well-defined domains termed topologically-associated domains (TADs) showing high intra-domain chromatin contact probability^{163,166}. CTCF has important functions in defining TAD boundaries, thereby providing physical constraints on enhancer-promoter interactions to increase the specificity of long-range enhancer action^{166–169}. In conclusion, CTCF functions as a chromatin architectural protein that contributes to the establishment of 3D chromatin structures in the nucleus. Although TAD boundary locations seem stable in different cell types, TAD boundary strength is under developmental control that likely regulates transcriptional variability in response to environmental cues¹⁷⁰. However, very little is known about the regulation of spatial genome organization in response to internal and external signals and its role in regulating adaptive phenotypic plasticity.

1.3.3. Circadian regulation of chromatin fibre interactions

An important example of phenotypic plasticity in nature is represented by circadian rhythm. In order to adapt to cyclic environmental changes during the day, living organisms have evolved endogenous timers which are called circadian clocks. These internal timing systems consist of cell-autonomous clocks that are in harmony with the geophysical time and enable light-sensitive organisms to predict daily oscillations in the environment¹⁷¹. Interestingly, the phase of circadian rhythm can be entrained or synchronized by environmental time cues^{172–177}. Apart from light and food intake, the circadian phase can be reset by a variety of internal and external signals, such as hormones, cytokines, exercise, metabolism, and temperature. Importantly, disruption of the circadian rhythm has been linked to a variety of complex diseases and pathological conditions, such as cardiovascular diseases, diabetes, obesity, neuropsychiatric disorders, immunological diseases and cancer^{178–182}.

An increasing number of studies have deciphered a new aspect in the regulation of circadian transcription, which deals with the regulation of circadian chromatin transitions in the primary chromatin fibre and in the 3D space of the nucleus. The factors composing the circadian gene network thus do not function alone, but instead collaborate with chromatin modifiers in order to fine-tune the phase and amplitude of circadian gene expression^{183,184}. However, even though the interplay between 3D chromatin crosstalk and circadian rhythm has been well documented^{185–188}, the roles of the nuclear architecture and nuclear hallmarks have not been elucidated.

Our group provided new insights into the role of spatial genome organization in the entrainment of circadian rhythm¹⁸⁹. In this study, it was demonstrated that synchronization of circadian rhythm by serum shock involved the oscillating mobility of circadian loci between the transcriptionally permissive nuclear interior and the generally repressive nuclear periphery. Chromatin mobility was coordinated by the diurnal complex formation between the genome organizer CTCF and Poly(ADP-Ribose) Polymerase 1 (PARP1), a factor regulating DNA repair and gene expression^{190,191}. The recruitment of circadian loci to the periphery leads to the

gradual acquisition of repressive H3K9me2 marks and transcriptional repression concomitant with the formation of an inter-chromosomal interactome between circadian loci and LADs. Following repression and in parallel with the disruption of CTCF-PARP1 proximity, circadian gene loci are released back to the interior of the nucleus, where their transcription can be initiated again. Interestingly, depletion of CTCF, PARP1 or H3K9me2/3 marks from the nuclear periphery, or the disruption of the CTCF-PARP1 complex formation by Olaparib (a PARP1-4 inhibitor) in HCT116 cells inhibited the oscillating recruitment of circadian loci to the nuclear periphery and ultimately, the synchronization of circadian transcription by serum shock.

These observations highlight that active and inactive nuclear compartments communicate with each other to regulate oscillating gene expression. Moreover, the significance of the nuclear periphery in the regulation of circadian transcriptional attenuation suggests that repressed LADs might affect the circadian expression to indirectly influence cellular phenotypes. This idea is further supported by findings showing the transcriptional-translational feedback loops of the circadian clock are absent in mouse embryonic stem cells¹⁹² and emerge only during differentiation, concomitant with the establishment of repressed chromatin (LOCKs) at the lamina²².

Interestingly, circadian loci were recruited to the nuclear periphery in a transcriptionally active state. The overlap between the peak of their transcription and the peak of their recruitment to the periphery¹⁸⁹ suggests that they might land in an environment that is linked with and might even promote transcriptional activation. Such a permissive environment could be provided by the nuclear pores, setting up the stage to explore what is the potential role of the NPCs in this process and whether a mechanism of post-transcriptional regulation, such as gene gating, is part of the story.

1.4. Chromatin organization and cancer

1.4.1. *MYC*: a key player in cancer development

One of the genes involved in the aforementioned study, that was found to be under circadian control, is *MYC*. Myelocytomatosis proto-oncogene (*c-MYC*) is a protein-coding gene located in the gene desert of the human chromosomal band 8q24^{193,194}. Its protein is a key transcription factor involved in many cell-fate decisions, such as proliferation, cell cycle, metabolism and apoptosis^{195–197}. *MYC* has been the focus of many studies through the years due to its wide involvement in up to 70% of all types of human cancer¹⁹³, when it is mutated or transcriptionally dysregulated. Its role in cancer though has been found to be indirect as well, since it controls the recruitment of histone modifying enzymes, like HATs and ATP-dependent chromatin -remodeling complexes to promote or inhibit the transcription of its target genes^{198–}

200.

c-MYC belongs to the MYC super-transcription factor family, which also contains MYCL and MYCN. These proteins are characterized by a basic helix-loop-helix leucine zipper domain, that is necessary for the heterodimerization of c-MYC with the MYC-associated factor X (MAX)^{201,202}. The MYC:MAX protein complex represents the functional form of c-MYC and regulates the transcription of several cancer-related genes through DNA binding on specific conserved sequences of their promoters termed E-boxes^{203,204}. Apart from MAX, MYC heterodimerizes also with MIZ1 forming a protein complex that binds to non-E-box sequences of target gene promoters, such as those of the circadian core clock machinery, CLOCK and BMAL1, implicating MYC as a key regulator of the circadian and cell cycles²⁰⁵. Finally, cellular MYC levels are tightly controlled by regulated protein degradation. More specific, when phosphorylated on the residue T58, c-MYC is targeted by SCF^{FBXW7} for ubiquitylation and proteasomal degradation¹⁹³. This process is really important since the sequence around the T58 of MYC has been identified as a hotspot of mutations in human cancer¹⁹³ and studies have shown that disruption of the turnover of T58-phosphorylated MYC leads to MYC stabilization and tumorigenesis¹⁹³. Notably, a recent report uncovered a novel function for the cryptochrome repressor of the circadian clock, CRY2, in the control of MYC turnover, offering new insights into the link between cancer susceptibility and circadian disruption²⁰⁶.

Despite our vast knowledge about *MYC*, it still remains a great enigma in cancer treatment, since it cannot be targeted pharmacologically^{207,208}. Because of that, *MYC*-induced cancer therapy focuses more on the epigenetic machinery that regulates *MYC* expression and function in cancer cells. This makes our understanding of the genome organization around *MYC*, as well as of the *MYC*-induced epigenetic mechanisms mediating tumor progression critical. Four alternative promoters control *MYC* transcription (P0, P1, P2 and P3) and its expression is regulated by a number of enhancers located in the 2.8 Mb TAD where *MYC* is also located (chromosome 8q24)¹⁹³. Moreover, the region upstream of P1 is characterized by the presence of a nuclease hypersensitivity element III₁ (NHE III₁), which is responsible for almost 90% of *MYC* transcriptional control^{209,210}. The importance of this sequence located -142 to -145 bp upstream of promoter 1 is the formation of a G-quadruplex (G4) structure, due to five guanine nucleotides that attribute two intramolecular secondary structures in each strand²¹¹. Because of this G4 structure, NHE III₁ functions as a negative regulator of *MYC* transcription¹⁹³.

Several years of studies have uncovered the significant role enhancers play in the regulation of *MYC* expression. *MYC* is embedded inside a TAD, meaning that long clusters of enhancers (including an upstream large specific super-enhancer) are located at both sides of the gene and mediate tumor-specific chromosome looping with promoters. In particular, the presence of a CTCF binding site 2 Kb upstream of *MYC* is required for the formation of enhancer-promoter loops and it is known as enhancer-docking site²¹²⁻²¹⁴. In addition, the DNA sequence upstream of *MYC* promoter is also characterized by several other CTCF binding sites which serve in the formation of subTADs²¹². Interestingly, genome-wide association studies (GWAS) have shown that susceptibility to numerous diseases, and especially cancer, can increase upon abrogation of these enhancer-promoter interactions around *MYC*^{215,216}. A super-enhancer, recently termed as blood enhancer cluster (BENC) and conserved between mice and humans,

is also located 1.7 Mb downstream of *MYC*. This BENC is crucial for both normal hematopoiesis and leukemia stem cell development, since its disruption has been associated with an increase of differentiation-arrested multipotent progenitors, due to downregulation of *MYC* in hematopoietic stem cells^{217,218}. Evidence has shown that during hematopoiesis in mice, BENC regulates *MYC* transcription in a three-dimensional loop manner by increasing the chromatin accessibility and interaction with *MYC* promoter¹⁹³. Finally, the general gene desert located on chromosome 8q24 close to *MYC* is known to include several risk loci to multiple types of cancer, such as myeloma, pancreatic, colon, prostate, breast, thyroid, bladder and chronic lymphocytic leukemia^{219–222}.

An example of the tight interplay between the regulation of *MYC* expression and 3D genome organization was recently demonstrated. First, Jurian Schuijers and colleagues discovered that when the enhancer-docking site depicting on the CTCF binding site is genetically perturbed, this results in reduced CTCF binding, interaction with the *MYC* super-enhancer and therefore, decreased *MYC* expression and tumor cell proliferation²¹⁴. Using epigenetic editing with dCas9-DNMT they showed that this enhancer-docking site is hypomethylated in diverse cancers, making them vulnerable. In addition, other genes, including genes involved in cancer, were found to be controlled by similar enhancer-docking sites, indicating a common mechanism through which cancer cell oncogenes can generally hijack enhancers. Later, Pulikkan et al. showed that in leukemia cases, the fusion oncoprotein CBF β -SMMHC drives cancer cell viability by inhibiting the repression of *MYC* expression by the transcription factor RUNX1²²³. When the CBF β -SMMHC/RUNX1 interaction is pharmacologically prevented, RUNX1 mediates apoptosis through increased binding at three *MYC* distal enhancers, which represses *MYC* expression by replacing the BRG1 component of the SWI/SNF complex with the Polycomb-repressive complex component RING1B. Interestingly, these three enhancers were shown to be physically connected with *MYC* and functionally involved in the deregulation of its expression. Finally, using the novel ultra-sensitive Nodewalk method, our research group recently discovered that *MYC* comes in physical contact with its flanking enhancers in a mutually exclusive manner, pointing out that enhancer hubs impinging on *MYC* that are detected in large cell populations most likely do not exist in single cells¹⁴⁸. Furthermore, this study showed that *MYC* functions as the most central node of the chromatin network it participates in, meaning that most likely it drives the communications with its flanking enhancers. These encounters with pathologically activated enhancers are dynamic and were shown to involve less than 10% of active *MYC* alleles at any given time point in colon cancer cells. This clearly demonstrates that the wide range of environmental signals a cancer cell encounters during the neoplastic process have the ability to dynamically activate *MYC*, as a response to them.

1.4.2. WNT signaling pathway

Another characteristic of *MYC* with particular interest is its involvement and interaction with several signaling pathways^{224,225}. However, the pathway with the most well-known role in the

pathological regulation of *MYC* expression and function is WNT. A lot of attention has been drawn on WNT signaling pathway over the years due to its profound contributions in development, tissue homeostasis and human malignancies²²⁶.

WNT signaling consists by three distinct intracellular cascades. These are a) the canonical WNT pathway which regulates the transcription of its target genes by β -catenin, b) the non-canonical WNT, which is β -catenin-independent and participates in planar cell polarity signaling, and finally, c) the WNT-dependent calcium/PKC-dependent pathway^{227,228}. For the purpose of the present thesis, I am going to focus on the canonical WNT signaling pathway, which is also the most frequently involved in tumor development and cancer.

In the absence of a WNT ligand, when the signaling is inactive, β -catenin is lead to degradation, directly after its translation, upon phosphorylation by a cytoplasmic protein complex consisted of Axin, Glycogen synthase kinase 3 β (GSK3 β), protein phosphatase 2A (PP2A), casein kinase 1 (CK1) and the adenomatous polyposis coli (APC) tumor suppressor protein^{228,229}. Phosphorylation by this complex takes place at key Ser and Thr residues in the amino-terminal part of β -catenin (Ser33, Ser37, Thr41 and Ser45). When phosphorylation fails at these residues or the above protein complex is defective, β -catenin escapes degradation and that leads to pathological conditions, such as colon carcinoma^{226,229}.

In the presence of WNT ligands, like WNT3a, the transmembrane domain receptors Frizzled and co-receptors, such as LRP5 or LRP6, are activated and subsequently, the cytoplasmic protein DVL is recruited to the receptor (**Figure 4**). This will then lead to the sequestration of Axin from the degradation complex, causing β -catenin to escape degradation. Due to its hypophosphorylated state, β -catenin will enter the nucleus and physically interact with members of the TCF/LEF family proteins, such as TCF4, making a bipartite transcription factor^{230,231}. The T cell factor/lymphoid enhancer factor proteins are DNA-binding proteins, which permit to β -catenin access chromatin and regulate the transcription of its target genes. Recent studies have shown that the three-dimensional structure of β -catenin plays an important role in its interactions with nuclear factors. β -catenin belongs to the Armadillo (ARM) repeat protein superfamily and includes a central stretch of 12 ARM repeats (R1-R12)^{232,233}. This central region is necessary for the interaction of β -catenin with TCF, as well as for the activation of Wnt target genes. During activation, β -catenin recruits several factors, including the histone acetyltransferases p300, CBP and TIP60, the Mediator component MED12, the SWI/SNF factors BRG1 and ISWI, and finally, the PAF1 complex component parafibromin/HYX²²⁶. These factors open the chromatin conformation and set a platform for RNA Pol II to initiate transcription.

In the rest of the pathway, downstream to β -catenin, when WNT ligands are absent, TCF proteins bind to specific DNA sequences, that are part of their target genes and termed WNT response elements (WREs), inducing the repression of any WNT-independent expression²³⁴. This repression is mediated by TLE1, a long-range chromatin repressor protein, which collaborates with HDACs to compact local chromatin. The transcription of WNT target genes can also be inhibited by another chromatin repressor named C-terminal binding protein

(CTBP), which functions in a TCF-independent manner²³⁵. TLE1 is known to compete with β -catenin for TCF binding²³⁶, providing a molecular “switch” between repression and activation. However, in the presence of a WNT signal, the TCF recruitment to the enhancers controlling the expression of WNT target genes intensifies, suggesting the need for chromatin remodeling during activation^{237,238}.

As one of its many target genes, *MYC* demonstrates a strong correlation with WNT signaling. The nuclear translocation of β -catenin induces the upregulation of *MYC* and therefore, cellular proliferation. Upon interaction and complex formation with members of the TCF/LEF family, it facilitates the binding of the active complex to the WREs located downstream and upstream of *MYC*²³⁹. Subsequently, the WREs will regulate the expression of *MYC* through the formation of a large chromatin loop with the *MYC* promoter mediated by CTCF²⁴⁰. Moreover, Posternak and colleagues recently discovered that MYC protein can in turn enhance the mRNA cap methylation of genes involved in WNT/ β -catenin signaling²⁴¹. That could serve as a feedback loop mechanism, since increased mRNA cap methylation of Wnt signaling transcripts induced by MYC results in higher protein levels and enhanced WNT signaling activity. Of note, MYC induces the recruitment of RNA methyltransferase (RNMT) to WNT target gene promoters through the interaction between TFIID and the TIP60 acetyltransferase complex, which leads to enhanced phosphorylation of serine 5 on the carboxy-terminal domain of RNA polymerase II to promote transcriptional elongation. In conclusion, all the available evidence points out that *MYC* and WNT signaling are tightly interconnected and their crosstalk remains something to be further elucidated.

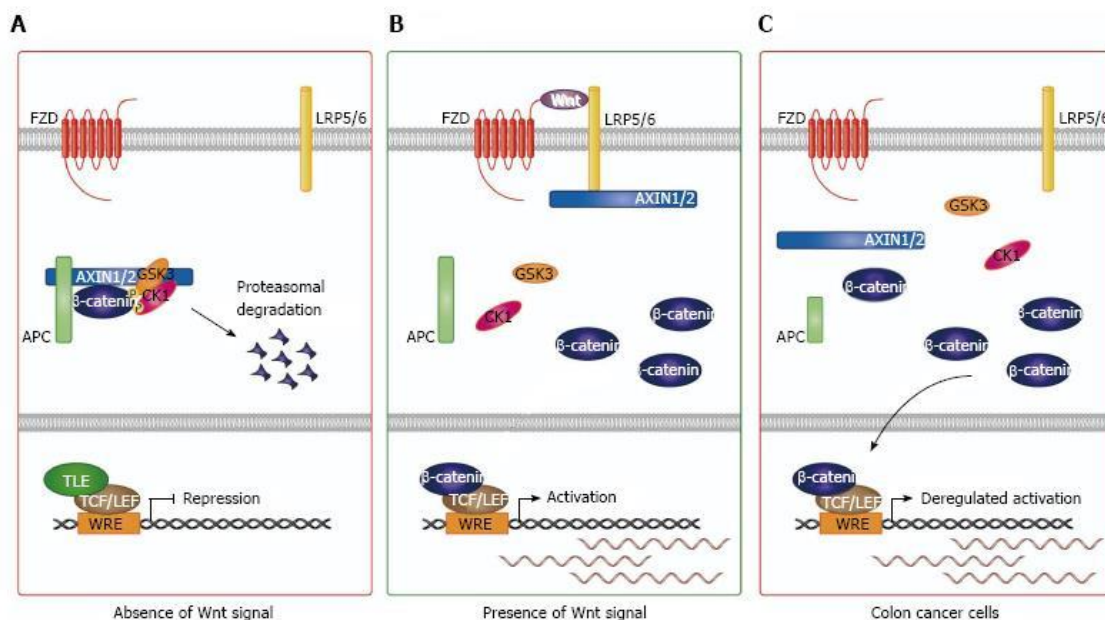


Figure 4: WNT/ β -catenin signaling pathway, in the absence (A) or presence (B) of a WNT ligand. In colon cancer cells (C), β -catenin is not targeted for proteasomal degradation, resulting in increased β -catenin levels in the nucleus and subsequent aberrant association with TCF/Lef at WREs to promote deregulated expression of WNT target genes. Reprinted under the Creative Commons Attribution 4.0 International License, <https://creativecommons.org/licenses/by/4.0/>, by World J Biol Chem. Nov 26, 2015; 6(4): 290-300. Rennoll S, Yochum G. Regulation of *MYC* gene expression by aberrant Wnt/ β -catenin signaling in colorectal cancer.

1.4.3. Non-coding RNAs in tumor development

Apart from WNT, transcriptome analyses of several tumor samples have shown that *MYC* expression, as well as transcriptional and post-transcriptional regulation, is extensively controlled by non-coding RNAs (ncRNAs), which induce tumor progression^{193,242–244}. ncRNAs can affect tumorigenic expression in multiple ways, such as directly promoting or inhibiting transcription, mediating protein binding on chromatin or protein-protein interactions and indirectly inducing changes in chromatin conformations. Some examples of ncRNAs involved in cancer are: a) *HOTAIR*, which is overexpressed in breast tumors and is known to promote cancer metastasis by recruiting chromatin repressor Polycomb proteins to specific genes²⁴⁵, b) *MALAT1*, which is involved in the regulation of tumor metastasis and cell growth^{246,247}, and finally, c) the lncRNA *LincRNA-p21*, which has been found to function as a transcriptional repressor in p53-dependent responses²⁴⁸ or collaborates with the RNA-binding protein HuR in order to suppress target mRNA translation²⁴⁹. However, what is the exact mechanism of function and the specific role of ncRNAs in tumor initiation and development still remains poorly understood and a topical research question deserving further attention.

A genomic region of particular interest, regarding *MYC* regulation, is the 8q24 gene desert. Several lncRNAs associated with cancer have been found to be transcribed in this region^{250–252}. Thus, the lncRNA cancer-associated region lncRNA 5 (*CARLo-5*) has been found to induce the development of colorectal cancer by inhibiting G1 arrest²⁵³. Research has shown that *MYC* enhancer regulates the expression of *CARLo-5* gene through direct long-range interaction with its promoter²⁵³. Moreover, *CCAT2* is another gene located in the 8q24 region with a well-characterized role in cancer. Specifically, this lncRNA has been found to promote tumor growth, chromosomal instability and metastasis in microsatellite-stable colorectal cancer by inducing the activation of TCF7L2-mediated WNT and its target genes, like *MYC*. In order to exert its metastatic effect, *CCAT2* induces the increase of two miRNAs located downstream of *MYC* gene and termed as *miR-17-5p* and *miR-20a*²⁵⁰. Similarly, the human colorectal cancer-specific *CCAT1-L* lncRNA is transcribed by a locus 515 kb upstream of *MYC* and overlapping with the *MYC* oncogenic super-enhancer, specific in colorectal cancer (**Figure 5**). This isoform is 5,200 nt long and has been demonstrated to control *MYC* transcriptional regulation. In particular, Jian-Feng Xiang et al. showed that *CCAT1-L* interacts with CTCF in HCT116 cells and promotes its binding to chromatin at the *MYC* locus, ultimately allowing it to mediate long-range chromatin interactions between *MYC* and its enhancers. Furthermore, overexpression of *CCAT1-L* in the same study enhanced *MYC* expression and promoted tumorigenesis, whereas knockdown of it reduced the expression of *MYC*²⁵⁴.

Interestingly, evidence has indicated that *MYC* can also be part of a positive feedback loop by activating the expression of the lncRNA *PVT1* upon binding to *PVT1* gene promoter. In that case, the upregulated *PVT1* is sponging the miRNA miR-486-3p, ultimately promoting proliferation and survival of cervical cancer cells²⁵⁵. Research has shown that *PVT1* gene can have a tumor-suppressor role as well, acting independently of the *PVT1* lncRNA, by competing for enhancers with *MYC* gene and therefore, abrogating *MYC* enhancer-promoter interactions,

finally leading to *MYC* downregulation²⁵⁶. Finally, apart from lncRNAs, miRNAs can play a crucial role in the regulation of *MYC* expression too. More specifically, *miR-129-2* can function as a tumor suppressor by downregulating *MYC* and blocking hepatocellular carcinoma proliferation and tumor growth²⁵⁶, while *miR-26a* has been found to inhibit cell growth in lymphoma cells by impeding the PRC2 subunit EZH2 and *MYC*²⁵⁷. Taken together, these findings emphasize on the important need for further investigation which will uncover the extended interplay between ncRNAs, *MYC* enhancer-promoter communications and WNT signaling in tumor development.

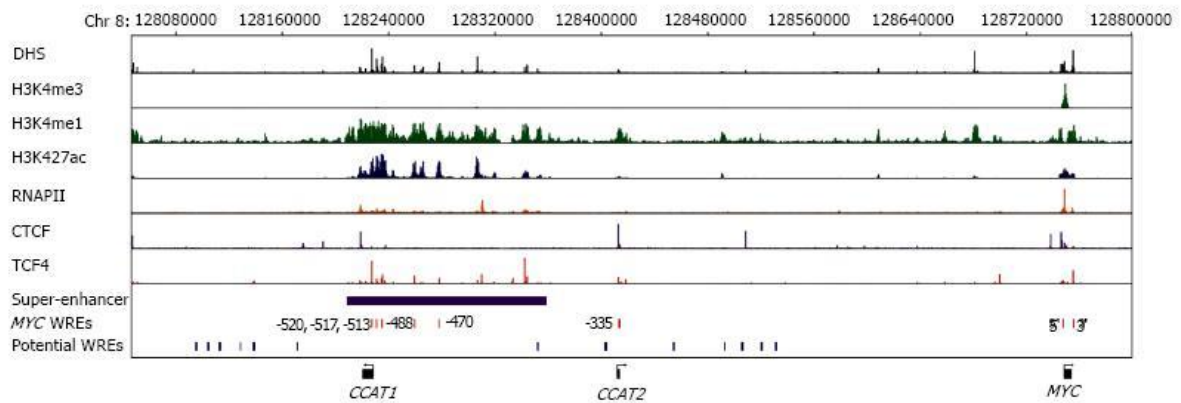


Figure 5: The *MYC* genomic locus in colorectal cancer. Histone marks and RNA Pol II, CTCF and TCF4 binding sites are shown as provided by ChIP-seq data in HCT116 (human colon cancer) cells. The DNase hypersensitivity (DHS) and ChIP-seq data depicted here were acquired from the WashU Epigenome Browser (<http://epigenomegateway.wustl.edu/>). Reprinted under the Creative Commons Attribution 4.0 International License, <https://creativecommons.org/licenses/by/4.0/>, by World J Biol Chem. Nov 26, 2015; 6(4): 290-300. Rennoll S, Yochum G. Regulation of *MYC* gene expression by aberrant Wnt/ β -catenin signaling in colorectal cancer

2 RESEARCH AIMS

The overall aim of the present thesis was to explore what is the role of the 3D nuclear architecture in the regulation of gene expression and how chromatin fiber movements and transitions respond to external stimuli during cancer development. To this end, two independent studies were implemented in order to answer the following research questions:

- 1) Is gene gating a mechanism taking place in human cancer cells to explain over-expression of *MYC*? What is the exact role of the NPC in the regulation of gene expression? How are chromatin movements coordinated by the WNT signaling pathway and how do they regulate *MYC* expression in cancer? Finally, what is the function, as well as the mechanism of action, of the colorectal oncogenic super-enhancer (OSE) controlling *MYC*?
- 2) What are the molecular factors and underlying mechanisms involved in the gating of *MYC*? How do these orchestrate enhancer-promoter interactions and localize them within the 3D nuclear architecture? More specifically, is CTCF, a master regulator of spatial genome organization and mediator of chromatin mobility, playing a key role in the WNT-regulated gene gating process of *MYC*?

3 MATERIALS AND METHODS

Methodology:

The general methodology followed in both studies covered by this thesis is mainly focused on the analysis of 3 different features: a) chromatin & protein physical interactions, b) localization within the 3D nuclear architecture and c) RNA assays. These methods are further described in detail below.

3.1. Cell culture and treatments

Both studies of the present thesis were performed using the human colon cancer cell line HCT116 and the normal human colon epithelial cells, HCEC. The HCT116 cells were kindly provided by B. Vogelstein (John Hopkins Medical School and Sidney Kimmel Comprehensive Cancer Center, USA), while HCECs were commercially provided by ScienCell. Both cell lines were cultured at 37 °C under 5% CO₂. HCT116 cells were chosen for this study, because a) of previous findings from our group related to this study in this cell line^{148,149,189}, b) the vast majority of *MYC* alleles are transcriptionally active any given time in these cells¹⁴⁹ and c) they are a Wnt-dependent cancer cell line, through a constantly active autocrine feedback loop²⁵⁸. Accordingly, HCECs were chosen as the healthy counterparts of HCT116, in order to see whether our findings in HCT116 are a feature unique to cancer cells and thus, to the disease. Interestingly, HCECs lack the oncogenic super-enhancer present in HCT116.

HCT116 cells were treated with 10 µM β-catenin and TCF Inhibitor V (BC21, Merck Millipore) or the equivalent amount of the solvent DMSO, as a control, for 16 h. These treatment conditions were followed since they demonstrated the optimal drug efficiency with low toxicity. BC21 was used in order to inhibit the WNT signaling pathway downstream, by disrupting the TCF4/β-catenin complex formation and β-catenin recruitment to chromatin. For Paper I, recombinant human WNT3a (R&R Systems) was administered to HCECs (added directly to the cell culture medium for the indicated periods of time), in an effort to examine whether and how this would affect *MYC* expression in these cells. Moreover, in the same study, leptomycin B was administered to HCT116 cells, since it inhibits the export of nuclear products²⁵⁹. That treatment would permit to assess whether the ongoing nuclear export of *MYC* transcripts in HCT116 cells affected the measured half-life of *MYC* mRNAs (see “Pulse labeling of RNA & mRNA export assay” below).

3.2. Editing of the *CCAT1*-specific CTCFBS by CRISPR-Cas9

For Paper II, we took advantage of the genome editing CRISPR-Cas9 method in order to generate HCT116 cell lines with mutated CTCFBS. This would prevent the binding of CTCF on this site of the OSE, and thus, would allow us to investigate the potential role of CTCF in gating. During this procedure, key sequences within the main CTCFBS of the OSE (chr8:128,219,114-128,219,767) were replaced using the CRISPR/Cas9 technology custom service of Synthego (CA, USA). The sequence within the CTCFBS was modified from CTCACCATTTGGAGGGCATTG to TTCATTATTTTATTTTCATTG. The edited cell pool was used to seed single cells for clonal expansion. Each well seeded was imaged every 2-3 days and rigorously tracked to ensure the population was truly clonal and only the progeny of a single cell. Two clones (D3 and E4) were selected and expanded. No selection agents were used to enrich for edited populations. The selection of two different clones was important in order to show that our observations were not the result of an artifact or clone-specific. As a control to the edited clones, a cell population which was not exposed to the reagents of the CRISPR process, from the initial HCT116 cell line, was used (termed as WT).

In addition, in order to perform co-culture experiments and distinguish the mutants from the wild-type cells, we designed and used qPCR primers annealing specifically on the mutated sequence of the CTCFBS. Accordingly, alternative primers, annealing only on the wild-type DNA sequence of the CTCFBS, were also used.

3.3. siRNA transfection

For both studies, HCT116 cells were transfected with siRNA against CTCF (Paper II) or AHCTF1 (Paper I & II) to downregulate CTCF or AHCTF1 expression, respectively. siGFP was used as a transfection control in both cases. The transfection was performed for either 48h (Paper I) or 72h (Paper II), using in both cases the Lipofectamine RNAiMax transfection reagent (Life Technologies), since upon protocol optimization these conditions were found to be optimal. This assay would allow us to observe the role of CTCF and/or AHCTF1. The downregulation efficiency was validated by RT-qPCR, as well as by ChIP-qPCR analysis, and was estimated at 85% on average, making our assay highly effective.

3.4. ChIP & Co-immunoprecipitation experiments (Chromatin/protein interactions)

3.4.1. Chromatin immunoprecipitation (ChIP-qPCR and ChIP-seq)

Extended chromatin immunoprecipitation experiments were performed in order to investigate the chromatin-protein interactions driving chromatin movements and a potential gene gating mechanism behind them. In particular, for Paper I, a NUP133 ChIP-qPCR assay was performed to analyze the recruitment of genomic loci to the nuclear pores, while an AHCTF1 ChIP was done to question the possible involvement of this nucleoporin to the process and understand its

precise mechanism of action. For Paper II, the binding of CTCF on the OSE in WT and mutant cells was analyzed by ChIP-qPCR, whereas a ChIP-seq approach was followed to validate whether CTCF binding was affected only on CTCFBS or other loci genome-wide in D3 and E4 mutant cells. Additionally, similarly to Paper I, the role of AHCTF1 in the gating of *MYC* was further investigated by ChIP-qPCR. Finally, TCF4 and β -catenin ChIPs were performed to score for potential changes in the WNT signaling pathway of WT and mutant cells. For these ChIP analyses, the fragmentation of chromatin was conducted with either sonication or enzymatic digestion using micrococcal nuclease (Cell Signaling), since both methods were tested and compared. For AHCTF1 binding to chromatin, which appears to be rather sensitive, ChIP was carried out under mild sonication conditions, in order to achieve the desired DNA fragments size without disrupting the protein/chromatin complexes. For the CTCF ChIP-seq experiments, the same protocol as for ChIP-qPCR was followed, but scaled-up to 3-fold, in order to increase the immunoprecipitated DNA material provided for sequencing.

3.4.2. Co-immunoprecipitation and Western blotting

An advantage of the methodology followed in our studies was the broad analysis of protein-protein interactions by co-immunoprecipitation and their additional validation by reverse co-IPs, in order to understand the specific role of the proteins involved in the gating process, as well as to uncover its steps by analyzing the protein complexes formed in it. Regarding this, in Paper I, TCF4/AHCTF1 and β -catenin/AHCTF1 co-immunoprecipitation experiments were performed in presence and absence of BC21 and DMSO control to decipher the role of WNT signaling and β -catenin in the gating of *MYC*. In Paper II, the interacting partners of CTCF were analyzed by co-immunoprecipitation of CTCF with AHCTF1, NUP133 and β -catenin. More precisely, the potential CTCF/AHCTF1 protein interaction was further examined under the presence of BC21, while co-immunoprecipitation of β -catenin with TCF4 was done as a control to validate the efficiency of the BC21 treatment. Importantly, the results of the co-IP assays were detected by a novel Western blotting technology called Simple Western using a WES/JESS system (ProteinSimple, Bio-Techne). This automated method permitted the better detection of the interacting partners due to its very high sensitivity and more precise quantitation and analysis of the data, thanks to the use of the Compass for SW software (ProteinSimple, San Jose, CA).

3.5. Nodewalk

A very important strength of our studies in comparison to others in this field was the use of the novel Nodewalk technique^{148,149} (see Introduction) for the analysis of chromatin fibre contacts, and specifically, the interactome of the *MYC* OSE. The identification of chromatin networks impinging on *MYC* and flanking enhancers was done as previously described in Sumida et al, 2018¹⁴⁹ and Sumida et al, 2020¹⁴⁸. Briefly, the Nodewalk technique converts ligated ('3C')

chromatin DNA fragments, reflecting physical proximities between distal regions in the living cell, into corresponding chimeric RNA sequences. Oligonucleotide sequences (baits) were annealed close to restriction enzyme sites flanking *MYC* and a selected set of enhancers, and this was followed by reverse transcription, with the resulting complementary DNAs subjected to high-throughput sequencing. Since the sequences generated by the different baits converged on common regions, the reads could be constructed into networks²⁶⁰ impinging on human *MYC*
149.

For Paper II, the analyses of physical interactions between *MYC* and the OSE were performed using an anchor representing the *MYC* gene. For this purpose, we took advantage of its ultra-sensitivity and highly quantitative features. For all integration analyses relevant to the Nodewalk networks, the 5 kb distance flanking each interactor was used as it approximates the average Hind III fragment size. The HCT-116 super-enhancers flanking *MYC* were identified from dbSuper, a super-enhancer database²⁰⁰. Finally, for Paper I, publicly available data for NUP153 binding were used (NUP153 DamID-seq peaks of U2OS cells were retrieved from Gene Expression Omnibus (GEO) with accession number GSE87831).

3.6. *In Situ* methods

3.6.1. Microscopy

For Paper I, cell imaging and the generation of optical sections in 3D were carried out on a Leica DMI 3000B fluorescent microscope (Leica) with an OptiGrid device (Grid Confocal) using Volocity software (Quorum Technologies). Stacks were taken at 0.3 μm intervals in the z axis. On average, the high number of 150-300 alleles were counted for distance measurements or ChrISP and ISPLA signal intensity in each case. For Paper II, the latest technology of a Leica DMI8 microscope, equipped with an HC PL APO 63x NA 1.4 oil subjective and DFC9000 camera, was used for the same purpose. The Instant Computational Clearing Method of the Thunder Imaging System (Leica Microsystems) was used to improve the signal-to-noise ratio. Stacks were taken using the software system optimized intervals in the z axis. The pictures were analyzed using the Imaris or Leica Application Suite X (LasX) software. Due to the limitations in the resolution of the fluorophores (with CY3 at 239,6 nanometer), the distance data were stratified using 240 nanometers as the first cut-off.

3.6.2. Chromatin In Situ Proximity (ChrISP)

In order to quantitatively analyze the proximities between members of the nuclear pore complex and *MYC* or OSE in Paper I, we took advantage of the importantly high resolution (167Å) provided by the Chromatin In Situ Proximity method developed by our group^{153,154} (see Introduction). In more detail, the ChrISP technique was used to investigate the proximity between the nucleoporin NUP133 and either the *MYC* promoter/gene body or the OSE. The

DNA FISH probes complementary to the *MYC* promoter or gene body, or the OSE were labeled with DIG-11-dUTP (Roche, 11573152910). For the proximity analysis between the DIG-labeled probes and NUP133, a tyramide signal amplification step (TSA kit with biotin-XX tyramide, Life Technologies, T-20921) was included to increase the concentration of biotin molecules in the vicinity of the NUP133 epitopes. The ChrISP assay was performed and quantitated as previously described^{149,153,154}. As a technical, negative control of the assay, a sample omitting the mouse secondary ChrISP antibody was included in parallel, called R plus control, in each experiment. False-positive signals (FPSs) of less than 0.05 were defined as the proportion of ChrISP signal of the technical negative control (experiment without one of the secondary antibodies) emerging above baseline. To distinguish the technical noise from experimental variation, we subtracted the intensity corresponding to FPS less than 0.05 from the intensity of each ChrISP signal in the samples. Signal intensities less than 0 were rounded to 0. The proximity between the *MYC* promoter or gene body and the OSE was analyzed as described by Sumida et al¹⁴⁹. Similarly, an additional ChrISP assay was performed with DNA FISH probes labeled with biotin or DIG aiming to determine the proximity between the OSE and *MYC* regions in relation to the nuclear periphery or pore. As *MYC* stochastically interacts with a range of enhancer elements equally distributed in both of its flanking TADs¹⁴⁹, we examined whether an enhancer more proximal to *MYC* would also show increased potential for interaction with *MYC* at the nuclear periphery or pore. For that, an enhancer region at the probe D position²⁶² (EnhD) was selected, as it does not bind either NUP153 or NUP133 (unpublished data). Finally, a negative site not interacting with *MYC* (Neg) was also used as a control for that assay.

3.6.3. 3D DNA FISH

For both studies, 3D DNA FISH was used in order to define the localization of *MYC* and the OSE in the 3D space of the nucleus, and quantitate the relative distance between them, as well as their distance to the nuclear periphery. For Paper I, this assay was performed in HCT116 and HCEC cells in comparison, whereas for Paper II, it was further examined in WT and mutant (D3, E4) HCT116 cells. The DNA FISH probes were prepared from a pool of PCR products spanning 8–10 kb regions of Hind III sites encompassing the *MYC* promoter and gene body (chr8:128,746,000–128,756,177), the OSE (positioned at chr8:128,216,526–128,225,855), or an in-between enhancer (EnhD, positioned at chr8:128413009–128414109)²⁶² (for Paper I), respectively. For Paper I, the probes were generated by PCR amplification using primers labeled with green 496-dUTP (Enzo, 42831) or Cy3-dCTP (GE Healthcare, PA53031). For Paper II, the probes were labeled with either biotin-16-dUTP (Roche, 11093070910) or digoxigenin-11-dUTP (Roche, 11573179910), using a Bioprime Array CGH kit (Life Technologies, 18095011). To ensure that the PCR probes identified the correct genomic region surrounding the OSE and *MYC* loci, the bacterial artificial chromosome (BAC) clone CTD-3066D1 was used. DNA FISH analyses were performed as previously described¹⁸⁹. Finally, the data analyses and distance measurements were done by comparing the distances between

the nuclear periphery or pores and the OSE ('a') or *MYC* ('b'): the subtraction of value a from value b generated a value 'c', which is positive when the OSE is closer to the periphery or nuclear pores and negative when *MYC* is closer to the periphery. That allowed us to assess the dynamics of the polarity between these loci.

3.6.4. RNA FISH

RNA FISH experiments were performed in both studies in order to visualize the localization of transcription in the 3D space of the nucleus and quantitate its levels. In Paper I, RNA FISH of *MYC* expression was done, using a probe covering the entire gene. The different RNA FISH probes were prepared as follows: the exon or intron probe was generated from a pool of four PCR products spanning the *MYC* promoter and its gene body (chr8:128,746,000–128,756,177) (GRCh37). The PCR products were sonicated and labeled with Biotin-16-dUTP (Roche, 11093070910) using a Bioprime Array CGH kit (Life Technologies, 18095-011). A mixture of equal amounts of each labeled PCR product was used as the FISH probe. The single-stranded intron 1 probe was prepared by generating double-stranded PCR fragments spanning *MYC* intron 1 (chr8:128,749,271–128,750,480), using a forward primer biotinylated at its 5' terminal. The double-stranded, biotinylated DNA was labeled with green 496 dUTP (Enzo Life Sciences, ENZ-42381) using a Bioprime Array CGH kit and then captured by Dynabeads M280 streptavidin (Thermo Fisher Scientific, 11205D). The beads were incubated at 98 °C for 5min to release and recover the non-biotinylated, labeled anti-sense strand. Finally, the single-stranded DNA was purified with Zymo Clean & Concentrator-5 (Zymo Research). RNA FISH analyses were performed as previously described¹⁸⁹. In Paper II, RNA FISH was performed to assess the potential ability of the *CCAT1*-L transcript and/or the transcriptional activity of *CCAT1* gene to facilitate OSE-*MYC* proximity. We chose this transcript, since we had earlier observed that WT HCT116 cells do not express the *CCAT1*-5L version, but instead, they prominently express the *CCAT1*-L version. For that, we first performed RNA FISH analyses to determine the distribution of active *CCAT1* alleles, followed by DNA FISH analyses using OSE- and *MYC*-specific probes to determine their proximities. The *CCAT1* RNA FISH probes were generated from two PCR products spanning a region within the OSE (positioned at chr8:128,216,526-128,225,855) and labeled with biotin-16-dUTP (Roche, 11093070910), as described above.

3.6.5. In Situ Proximity Ligation Assay (ISPLA)

In order to analyze how protein-protein proximities localize in the three-dimensional nucleus, ISPLA was performed. For Paper I, this assay was used to define the relationship of AHCTF1 with the WNT signaling pathway. For that, the following ISPLAs were conducted: β -catenin–AHCTF1, TCF4–AHCTF1, and β -catenin–TCF4, as an internal control for the efficiency of BC21 treatment. Samples that lacked the primary antibodies served as background controls.

For Paper II, the same method was used to define and quantitate the proximities between CTCF-NUP133 and CTCF-AHCTF1, as well as for investigating the potential effect of BC21 in the latter.

3.7. RNA assays

3.7.1. Pulse labeling of RNA & mRNA export assay

One important aspect of the methodology followed in both our studies was the clear separation of the nuclear from the cytoplasmic fraction, which allowed us to determine the export rate of *MYC* transcripts. For that purpose, we first conducted a pulse labeling of RNA by incubating HCEC and HCT116 cells with 0.5mM (final concentration) 5'-ethynyl uridine (Thermo Fisher Scientific, E10345) for 15 or 30min. 5'-ethynyl uridine (EU) would allow us to detect only the newly synthesized RNA. Subsequently, the EU was removed, separation of the nuclear and cytoplasmic fraction was performed 1h after the pulse chase, and the RNA was isolated using the Ambion PARIS system (Thermo Fisher Scientific, AM1921) according to the manufacturer's protocol. Upon many rounds of optimization experiments and testing of different separation methods, this appeared to be the most efficient and with the lowest contamination between the fractions. Then, labeled RNA or total RNA were purified with the RNeasy Mini kit (Qiagen, 74014) and 5'-ethynyl uridine–labeled RNAs were captured using a Click-iT nascent RNA capture kit (Thermo Fisher Scientific, C10365) following the manufacturer's instructions. Finally, they were converted into cDNA using a SuperScript VILO cDNA synthesis kit (Life Technologies, 11754050).

For the nuclear RNA export assay, in order to define the ratio between exported cytoplasmic and nascent nuclear RNA, the 5'-ethynyl uridine–labeled nuclear and cytoplasmic RNA that were fractionated and processed to cDNA above, were analyzed for the presence of intronic, as well as exonic regions of *MYC*, by qPCR analysis. To avoid an overestimation of cytoplasmic RNA due to its potential contamination by nuclear RNA, we determined and subtracted the quantity of nuclear RNA from the cytoplasmic fraction. This was implemented by first quantifying *MYC* intron 1 RNA sequences in both the cytoplasmic and nuclear fractions by qRT-PCR. Next, we used the ratio between cytosolic and nuclear intron RNA levels to estimate the level of contaminating nuclear *MYC* exon in the cytosolic fraction, followed by its subtraction from the cytosolic fractions and its addition to the nuclear fractions, respectively. We routinely detected on average 6% of contamination of nuclear *MYC* transcripts containing intron 1 sequences in the cytoplasm and on average 6% contamination of the mRNA for mitochondrial cytochrome B in the nuclear fractions.

For Paper II, the exact same methodology was followed, but instead of HCEC and HCT116 cells, the mRNA export assay was performed in WT and D3/E4 mutant HCT116 cells.

3.7.2. Decay analyses

In order to examine whether *MYC* transcripts were more stable in HCT116 cells than in HCECs and to estimate the mRNA stability in the nucleus and cytoplasm²⁶³, we performed a decay analysis by administering 5 µg/ml Actinomycin D (Sigma-Aldrich, A1410), which blocks transcription by inhibiting transcriptional elongation²⁰³. The incubation time of cells was 0, 0.5 and 1h, to evaluate the effect of the treatment, and was followed by monitoring the mRNA decay in total RNA preparations. Cytoplasmic and nuclear fractions were separated and total RNA purified as described above. An additional DNA digestion step was included to remove residual DNA (TURBO DNase, Ambion, AM1907). The samples were normalized to the total recovery of the nuclear and cytoplasmic fractions, respectively.

3.7.3. Analysis of transcription & RNA-seq

For the analysis of the transcription levels of *MYC* (Papers I & II), *FAM49B* and *CCAT1* (Paper II), qRT-PCR was performed. For Paper I, the transcriptional activity of both HCECs and HCT116 cells was examined, whereas for Paper II, the analysis was conducted in WT and D3/E4 mutant HCT116 cells. In case the levels of nascent transcripts were to be measured, the cells were labeled with EU as described above and harvested immediately, without prior removal of the EU. Subsequently, fractionation, RNA extraction and purification of the labeled RNA were followed as previously mentioned. The quality of purified RNA samples was assessed before cDNA synthesis (SuperScript VILO cDNA synthesis kit, Life Technologies, 11754050) using a Bioanalyzer 2100 (Agilent). For all qPCRs, the linear range of amplification was confirmed by serial dilution of sonicated genomic DNA from HCT116 cells. In Paper II, for the precise measurement of *MYC* and *FAM49B* mRNA levels, one million cells from WT HCT116 cells, and mutant D3 and E4 clones were counted. Before lysis, 1 µl of diluted (1/10) ERCC ExFold RNA Spike-In Mix (Thermo Fisher Scientific, catalogue number: 4456740) was added to each sample. Cell lysis and RNA purification was performed using RNeasy Mini kit (Qiagen, catalogue number: 74104) following the manufacturer's protocol.

For Paper II, RNA-seq was performed in order to investigate a potential involvement of the *CCAT1* eRNA in the gating of *MYC*. For that, we needed to first assess which of the known *CCAT1* RNA isoforms were expressed in our cell model, i.e. WT HCT116 cells. RNA-seq library constructions were generated with standard Illumina TruSeq Stranded mRNA kit with Poly-A selection and samples were sequenced on NovaSeq6000 (NovaSeq Control Software 1.7.0/RTA v3.4.4) with a 51nt(Read1)- 10nt(Index1)-10nt(Index2)-51nt(Read2) setup using “NovaSeqStandard” workflow in “SP” mode flowcell.

3.8. Bioinformatic analyses

3.8.1. Computer simulation modeling of *MYC* mRNA cytoplasmic levels over time

For the purpose of determining the expected difference over time in the levels of cytoplasmic *MYC* mRNA expression between HCT116 cells and HCECs, the data provided by the analyses of transcription (qRT-PCRs) and the mRNA export assays were used to create a computer simulation model. More precisely, to model the gradual increase of cytoplasmic levels of *MYC* mRNA over time from an initial time point, defined as α for the earliest measurable transcription rate, we followed a mathematical model which outlines the dynamics of nuclear (X) and cytoplasmic (Y) mRNAs²⁶⁵. For more details on the derived equation, please see the respective Methods part of Paper I.

3.8.2. Off-target whole genome sequencing

For Paper II, whole genome sequencing was performed in WT and D3/E4 mutant HCT116 cells, in order to score for any possible off-target effects of the CRISPR-Cas9 method used to specifically edit the *CCAT1*-specific CTCFBS. That would allow us to confirm whether any differences regarding the gating between WT and mutant cells were exclusively due to the different CTCFBS. Genome wide off-target sequencing was analyzed according to a modified method of GOTI²⁶⁶. Briefly, we first aligned clean reads to the genome and marked the duplicates. To reduce the false positive rate, we applied 3 methods to detect the SNV and Indel between D3/E4 and WT, mutect2²⁶⁷, Strelka2²⁶⁸ and LoFreq²⁶⁹. The overlapping SNVs and Indels were treated as true variants that were further annotated. The adjacent 22-bp sequences of the off-target variants were retrieved from the mapping files and blasted with the 22 bp sgRNA (19 bp sgRNA target sequence and 3 bp PAM). High sequence similarities indicated that the off-target variants were sgRNA-associated, while low sequence similarity meant sgRNA-independence²⁶⁶. To further screen for potential off-target sites that overlapped with the identified variant we used Cas-OFFinder (<http://www.rgenome.net/cas-offfinder/>).

3.8.3. ChIP-seq & RNA-seq analysis

Bioinformatic analysis was conducted on both the datasets generated by the ChIP-seq and RNA-seq assays described above and used in Paper II. For the ChIP-seq, the analysis was performed by nf-core/chipseq pipeline (version 1.2.2)²⁷⁰. Briefly, first the adapter was trimmed, followed by read alignment and mark of duplicates. Then, normalized bigWig files were created, strand cross-correlation peak and ChIP-seq quality measures were calculated, and the peaks were called. Finally, peak annotation was performed and all the read coverages were visualized by the IGV genome browser²⁷¹.

For the RNA-seq, data analysis was performed by nf-core RNA-seq pipeline (v.3.0)²⁷⁰ with default parameters. In brief, adapters and low-quality reads were first filtered. Then, clean reads were aligned to the GRCh37 human reference genome²⁷² and ERCC RNA spike-in fasta (ThermoFisher). Next, genome-wide coverage output in BEDGRAPH format were generated and bigWig coverage files were created. Finally, Integrative Genomics Viewer was again used to view the read coverage.

4 RESULTS

4.1. Members of the Nuclear Pore Complex interact with a chromatin network organized by *MYC* enhancers

The first observation that set the ground for further examination was that a chromatin network impinging on *MYC* gene and its enhancers was enriched in NUP153 binding. Specifically, using the Nodewalk technique and *MYC* as initial bait, we constructed a chromatin network and then defined the connectivity of its nodes, as well as their relationship to NUP153 occupancy, by comparing these data with previous, publicly available data from NUP153 ChIP-seq analyses²⁷³. Interestingly, this showed that in both human colon cancer cells and their healthy counterparts (HCT116 and HCECs, respectively), many of the most highly connected to *MYC* enhancers located within the two TADs flanking *MYC* are bound by this nucleoporin.

Based on this observation, we further examined the relationship between NUP153/NUP133 and the OSE by chromatin immunoprecipitation experiments, focusing on the oncogenic super-enhancer region that showed maximum interaction with *MYC* in HCT116 cells¹⁴⁹. We discovered that both the *MYC* gene and its colorectal oncogenic super-enhancer are in contact with these members of the nuclear pore complex (NPC). Using ChIP analyses and subsequent qPCR with primers covering 4 different regions of the super-enhancer, we found that the nucleoporin NUP133, which is part of the nuclear ring structure of the NPC, binds preferentially the OSE in HCT116 cells, while *MYC* promoter, preferentially interacts with NUP153, which forms the nuclear cage. However, the NUP133-OSE interaction was found to be absent in normal colon epithelial cells (HCECs), suggesting that this link is a cancer-cell-specific feature.

4.2. The proximities between NUP133 and the OSE are dynamic, while *MYC*-OSE proximities correlate with the nuclear architecture

To investigate to a greater extent the potential for interaction between NUP133 and the OSE in correlation with the nuclear architecture, we used Chromatin In Situ Proximity analysis in both HCECs and HCT116 cells. In agreement with the ChIP data, the results showed increased NUP133-OSE ChrISP signal in HCT116 cells, but not in human colon epithelial cells. Moreover, NUP133 was found to be in close proximity with the OSE, but not with *MYC*. Interestingly, the NUP133-OSE ChrISP intensity increased near the nuclear periphery of HCT116 cells, with more than 70% of the alleles located at a distance $\leq 0.3 \mu\text{m}$ from the periphery demonstrating a positive signal, indicating a transient anchoring of the super-enhancer at the nuclear pores.

Taking into consideration the Nodewalk data, we then wanted to examine what are the relative positions of *MYC* and the OSE, which demonstrated the highest frequency of interaction with

*MYC*¹⁴⁹, in relation to the nuclear periphery. Using 3D DNA FISH in both HCEC and HCT116 cells, we observed that the OSE is closer to the periphery in cancer cells than in normal cells, while the respective position of *MYC* did not show dramatic differences between the two cell lines. Following up on this, we next aimed to determine the polarity in the localization of the two loci in respect to the periphery of the nucleus in HCT116 cells. The data resulting from that analysis indicated that *MYC* and the OSE tend to occupy random positions in the inner space of the nucleus ($>1\ \mu\text{m}$ distance from the nuclear periphery), but as they approach closer to the periphery, the OSE comes most frequently in closer proximity to the periphery/pores than *MYC* does. This positioning suggests that the movement of the two loci is initially stochastic, but follows a specific polarity near the periphery, in which the OSE seems to juxtapose the active *MYC* gene to the nuclear pore.

To validate these findings with a higher resolution, ChrISP was once again used, by assessing the proximities between *MYC* and the OSE, but as well as an in-between enhancer named Enhancer D (EnhD), which did not interact with either NUP153 or NUP133. In addition, a region known to not interact with *MYC* was also used as a negative control. Quantitation of the ChrISP signal in each case showed that, as expected, *MYC* was generally in closer proximity with EnhD than with the OSE, while almost no signal was detected between *MYC* and its negative site. However, further review of this observation, in respect to the nuclear periphery, demonstrated that *MYC*-OSE proximity is much more intense at the periphery ($\leq 0.3\ \mu\text{m}$ distance) compared to the *MYC*-EnhD proximities, which are increased at a more distal position in relation to the periphery ($0.9\text{-}1.7\ \mu\text{m}$ distance).

Finally, using RNA FISH assays followed by DNA FISH on the *MYC* locus, we examined how the *MYC*-OSE proximities are spread in the nucleus in relation to ongoing *MYC* transcription. Using two different RNA FISH probes, covering either only the first intron or the entire *MYC* gene, to score for unspliced or spliced transcripts, respectively, we observed that the transcriptionally active alleles were located mainly either near the periphery/pore ($0.3\text{-}0.7\ \mu\text{m}$) or more distally ($1.1\text{-}1.7\ \mu\text{m}$). Interestingly, these data correlated with the aforementioned ChrISP results, showing that the *MYC*-OSE proximity overlapped mainly with the accumulation of processed *MYC* transcripts near the periphery, whereas the EnhD-*MYC* highest proximity overlapped with the localization of unspliced transcripts.

4.3. *MYC* expression is regulated post-transcriptionally by a gene gating mechanism

Following up these data, we questioned whether the action of the OSE could be explained by a potential gene gating mechanism of regulation, which was previously known to exist in other organisms^{51,52}. To test this hypothesis, we first determined the levels of *MYC* transcription in HCEC and HCT116 cells. For that purpose, cells were incubated with 5'-ethynyl uridine (EU), which would allow the labeling of nascent transcripts and their distinguishment from the already existing at the time of the analysis *MYC* mRNA. Surprisingly, qRT-PCR analyses revealed that even though *MYC* transcription levels were higher in HCEC than HCT116 cells,

the total RNA levels, as well as the levels of cytoplasmic *MYC* mRNA, were significantly increased in cancer cells, suggesting that *MYC* expression is regulated at the post-transcriptional level. Following a pulse-chase approach, where newly synthesized transcripts in HCT116 cells were initially labeled by EU and then, chased for 0-60 minutes, every 15 min, upon fractionation of the nuclear and cytoplasmic compartments (for more details see Methods), we interestingly observed that, while the ratio of cytoplasmic/nuclear *MYC* transcripts remained the same at all time points in HCECs, it dramatically increased, instead, over time in cancer cells (reaching its peak at 60 min). This reinforced our hypothesis on an mRNA export mechanism taking place in the regulation of *MYC* expression.

Taking these findings together, we next addressed whether the stability of *MYC* mRNAs would be different between the two cell types and thus, explain the observed differences in mRNA export. To answer this, we used actinomycin D, which blocks transcription, and tracked the mRNA degradation rate in total RNA preparations. Subsequent qRT-PCR analyses indicated that, even though the total *MYC* mRNA decay rate was the same between cancer and normal cells (average half-life of around 90 min), in both cases the degradation was much quicker in the nucleus than the cytoplasm. Moreover, focusing more on the mRNA decay of HCT116 cells, we then additionally used leptomycin B to inhibit mRNA export, which is known to block the export of any nuclear products¹⁹⁸, and see whether the measured half-life of *MYC* transcripts was influenced by the ongoing export. Interestingly, the results showed that this treatment had almost no effect on the measured decay rate. That would suggest that, in cancer cells, the OSE recruits *MYC* to the nuclear pores in order to facilitate the quick export of its transcripts to the cytoplasm and protect them from the enhanced mRNA degradation taking place in the nucleus. This would ultimately lead to the presence of more *MYC* transcripts in the cytoplasm and thus, higher total *MYC* expression levels in cancer cells, without increasing transcription, resulting in their survival and proliferation.

To finally confirm that, we used the data provided from the aforementioned experiments to generate a computer simulation mathematical model which would allow us to estimate the acquisition of steady-state cytoplasmic *MYC* mRNA levels in HCECs and HCT116 cells. The simulation results came in agreement with our experimental observations, showing that the entire time window needed to produce the experimentally observed differences in *MYC* mRNA accumulation levels cancer and normal cells might need to extend beyond the cytoplasmic and nuclear half-life rates of *MYC* mRNAs in order to achieve steady-state levels. Therefore, since the mRNA decay rates are very similar for both the nuclear and cytoplasmic compartments of HCT116 cells and HCECs, we suggest that indeed the cancer-cell-specific tethering of the active *MYC* gene to the pore might function as an escape mechanism from the rapid RNA degradation kinetics of the nucleus.

4.4. The super-enhancer-mediated gating of *MYC* is regulated by AHCTF1

After identifying this process, the next step in our reasoning was to detect molecular factors responsible for the recruitment of both *MYC* and the OSE to the pores. In that effort, we initially tried several potential candidates, such as members of the Mediator complex (CDK8), CDK9

and BRD4. However, downregulation by siRNA or inhibition by specific drugs, like JQ1, seemed to have no significant effect in the gating of *MYC*. Our focus then shifted to the NPC component AHCTF1 for the following reasons: a) it is a mobile nucleoporin, not permanently anchored to the NPC, b) it binds directly to chromatin³⁶ and c) it is the central factor of the NUP107 sub-complex, which had been previously shown to mediate the transport of a target plasmid from an intranuclear position to the pores²⁷⁴.

In order to examine a potential involvement of AHCTF1 in this mechanism, we first downregulated its expression by siAHCTF1 transfection and then, following a similar methodology as before, we observed its effect on *MYC* mRNA export, *MYC* transcription and binding of NUP133 on the OSE by ChIP-qPCR. The results demonstrated that even though knockdown of AHCTF1 did not significantly affect *MYC* transcription, it significantly reduced the nuclear export rate of *MYC* mRNA, when compared to the siGFP control. Moreover, downregulation of AHCTF1 induced a decrease of OSE occupancy by NUP133, suggesting that it binds to the OSE indirectly *via* AHCTF1.

Finally, we aimed at investigating how this set of data could correlate with the nuclear architecture. Again, we used the ChrISP technique to analyze the distribution of OSE-NUP133 proximities upon siAHCTF1 transfection. As our results indicated, knockdown of AHCTF1 diminished the ChrISP signal significantly, emphasizing the need of this nucleoporin for the recruitment and anchoring of the OSE to the pores. On the contrary, the absence of AHCTF1 did not seem to affect the polarity in the *MYC*-OSE positioning towards the periphery, neither their in-between proximities, suggesting that the OSE can establish contact and recruit *MYC* to perinuclear locations, but it ultimately requires AHCTF1 in order to finally reach the nuclear pore and promote the facilitated export of *MYC* transcripts.

4.5. WNT signaling mediates the recruitment of *MYC* to the nuclear pores and the export rate of *MYC* transcripts through β -catenin

To better understand the gating mechanism underlying *MYC* expression in HCT116 cells, we benefit from the information provided by previous reports observing that in this cell type WNT signaling seems to converge on TCF4-binding sites at enhancers controlling *MYC* expression^{131,240,275}. In addition, these cells exhibit a WNT3a autocrine loop²⁷⁶, making the pathway constantly active, while they produce at the same time a variety of WNT ligands which promote both the canonical and non-canonical signaling pathway²⁷⁷. To explore such a scenario, we targeted the canonical pathway by using the TCF4/ β -catenin protein complex inhibitor, BC21. This drug inhibits canonical WNT signaling by evicting β -catenin from chromatin. First, we investigated what was the relationship of the two main factors of the pathway, TCF4 and β -catenin, with AHCTF1, and whether this would be affected by inhibition of WNT. For that, we performed in parallel *in situ* proximity ligation assays and co-immunoprecipitation experiments. Interestingly, both methods indicated that AHCTF1 physically interacts with both TCF4 and β -catenin, while BC21 induced the loss of interaction

with the former, but not the latter, suggesting that the TCF4-AHCTF1 complex formation is likely indirect and dependent on β -catenin.

Next, we wanted to decipher what the effect of BC21 treatment, and therefore WNT/ β -catenin inhibition, would be on the OSE-mediated gating of *MYC* in HCT116 cells. To answer that question, we first applied a ChrISP analysis between NUP133, which is recruited to chromatin by AHCTF1, and the OSE upon BC21 treatment. In agreement with our hypothesis, BC21 diminished the produced ChrISP signal, compared to control cells treated with DMSO, indicating the suppression of both the potential for interaction between NUP133-OSE and anchoring of the OSE to the nuclear pore. To further examine the function of the direct physical interactions between β -catenin, AHCTF1 and the OSE, we chose to focus on a particular prominent TCF4-binding site included in the super-enhancer, known as WNT-response element 520 (WRE520)²⁴⁰. ChIP-qPCR analyses showed that AHCTF1 binds to WRE520 several-fold more efficiently than to *MYC* promoter. Furthermore, the use of BC21 reduced the occupancy of the WRE520 region equally by all three AHCTF1, NUP133 and β -catenin factors, suggesting that the binding of AHCTF1 (and thus, NUP133) on WRE520 might be β -catenin-dependent. Finally, new mRNA export assays, upon nuclear and cytoplasmic fraction separation, as well as qRT-PCR analyses of total and newly synthesized, EU-labeled *MYC* RNA showed that BC21 treatment significantly prevented the nuclear export of *MYC* mRNAs in HCT116 cells. This in turn led to a reduction in the total expression levels, while *MYC* transcription rates remained unaffected. In conclusion, these findings taken together suggest that β -catenin regulates the OSE-mediated gating of *MYC* in cancer cells, by promoting the anchoring of the OSE to the pore and thus, the nuclear export of *MYC* transcripts to the cytoplasm.

4.6. Colon cancer cells gain an excessive growth advantage due to the binding of CTCF to the OSE of *MYC*

Following up the findings of Paper I, we aimed to examine whether the genome organizer CTCF plays a role in the OSE-mediated gating of *MYC* in HCT116 cells. A main reason behind this supposition was the observation that one of the most prominent CTCF binding sites in the genome (CTCFBS) is located within the OSE-specific eRNA gene *CCAT1* and proximal to the TCF4 binding site WRE520. Moreover, this CTCFBS within the OSE was previously found (Paper I) to be in physical proximity with *MYC* in colon cancer cells.

To assess the potential involvement of this *CCAT1*-specific CTCFBS in the gene gating process, we took advantage of the CRISPR technology to edit a sequence of 8 bps within its site and generated two different cell clones with a mutated OSE allele, named D3 and E4. First, we tested the efficiency and validity of our model by performing CTCF ChIP analyses in WT and mutant HCT116 cells, as well as in HCECs. As expected, the recovery of CTCF binding in WT cells was immense (>2% of input DNA on average), but almost completely absent in both D3 and E4 mutant clones (~0.2% of input, i.e. more than 10-fold decreased compared to

wild-type). Moreover, the same analysis in HCECs showed that CTCF bound on the region corresponding to the CTCFBS in HCT116 cells at a similar level as in D3 and E4 mutants, whereas the CTCF binding to the *H19* ICR, which was used as a positive control, showed no significant difference in recovery between WT or mutant HCT116 cells and HCECs. That indicated that the CRISPR-induced mutation of the CTCFBS was sufficient to prevent CTCF binding at the OSE, while binding to other key regions was not affected. Moreover, it revealed that cancer cells with a mutated CTCFBS demonstrate the same phenotype as normal cells, suggesting that the CTCFBS might play a key role in this process.

In order to make sure that the applied CRISPR-Cas9 editing was specific for the CTCFBS and did not affect other CTCF binding sites across the genome, we used a CTCF ChIP-seq analysis in both wild-type and mutant HCT116 cells. The results exhibited no additional differences in the binding of CTCF genome-wide, confirming that our following observations would be specifically due to the mutated CTCFBS and not any off-target effects.

Given of these data, we next aimed to elucidate whether the mutation of the CTCFBS would attribute a phenotypic trait to the mutated cancer cells. For that purpose, we developed primers annealing specifically on the mutated sequence and thus, allowing us to distinguish mutated from wild-type HCT116 cell populations in co-culturing conditions. This strategy would provide an internal control and thus, permit the direct comparison of the proliferation rate between the cell lines. Therefore, co-culture experiments followed by genomic DNA purification and qPCR analysis showed that WT cells displayed a growth advantage over both D3 and E4 cells, which increased over time during a culture period of 14 days. Interestingly, administration of BC21 reduced significantly the relative growth rate of both WT/D3 and WT/E4 cells, compared to the DMSO control, and this effect became more intense again over time during a period of 28 days as compared to 14 days of culture. This important finding suggested that the CTCFBS of the OSE confers a proliferative advantage to colon cancer cells in a WNT signaling-dependent manner.

4.7. The increased nuclear export rate of mRNAs produced from *MYC* and *FAM49B* genes is controlled by the *CCAT1*-specific CTCFBS within the OSE

The next step in our study was to explore the potential role the CTCFBS might have in the OSE-mediated gating of *MYC*. With that regard and following the same methodology as in Paper I, we performed qRT-PCR analysis of EU-labeled transcripts to determine the ratio of newly synthesized cytoplasmic over nuclear mRNA. Our assay revealed that the mRNA export rate was approximately 3-fold higher in WT than in D3 or E4 mutant cells. In addition, use of BC21 reduced the export rate of WT cells, compared to the DMSO control, to the level of the mutant cells, while it did not decrease further the export rate of D3 and E4 clones. Apart from *MYC*, we also determined the nuclear export rate of an additional mRNA, *FAM49B*. The reasons behind this were a) the well-known functional interplay between the *MYC* and *FAM49B* gene products^{278,279} and b) the data provided by previous Nodewalk analyses from

our group identifying physical interaction between the OSE and an enhancer proximal to *FAM49B*¹⁴⁸. Interestingly, the results on the export rate of *FAM49B* mRNA between wild-type and mutant cells, as well as the effect of BC21 on them, were in line with the data obtained on *MYC*.

Despite the apparent effect on the nuclear export rates, analyses of transcription levels based on *MYC* and *FAM49B* introns, normalized to β -actin as a control, showed that the CTCFBS did not control the transcription of either of the genes, since no differences could be observed between WT and mutant cells. We next aimed to determine the total expression levels of both genes in order to confirm whether differences in export were indeed linked with a decrease in the total RNA levels of the mutants. Indeed, qRT-PCR analyses of the steady-state levels of cytoplasmic *FAM49B* and *MYC* mRNAs in wild-type and D3, E4 mutant cells displayed, upon normalization to the expression levels of TBP and external “spike in” RNA controls, an almost 50% decrease of total expression in mutant cells, for both genes. Finally, as in Paper I, a computer simulation model was constructed based on the export, transcriptional rate and the kinetics of mRNA degradation in the nuclear and cytosolic compartments, in order to compare the observed versus the simulated levels of *MYC* RNA between the WT and D3/E4 cells. The results of this method indicated that the simulated RNA levels followed a similar pattern to the observed ones, with total *MYC* expression being almost 2-fold higher over time in WT than in mutant cells. Taking these data together, we concluded that the CTCFBS plays a key role in the export of both nuclear *MYC* and *FAM49B* transcripts and thus, it controls the gating and subsequent expression of both genes in HCT116 cells.

4.8. β -catenin and CTCF recruit AHCTF1 to the OSE to facilitate its ability to reach the nuclear pore

As we had previously found that the recruitment of the OSE to the nuclear pore during the gene gating mechanism is mediated by AHCTF1, we then wanted to explore the potential relationship of CTCF with AHCTF1. Co-immunoprecipitation experiments showed that CTCF physically interacts with AHCTF1 in HCT116 cells, but also with other two factors involved in the gating of *MYC*, i.e. the NPC component NUP133 and β -catenin. Moreover, the use of BC21 revealed that the CTCF-AHCTF1 interaction is most likely WNT-dependent, since this drug induced the disruption of that complex formation (~50% decrease compared to the DMSO control). Based on this set of data, we reasoned that AHCTF1 might be recruited to chromatin upon interaction with CTCF and thus, specifically bind the CTCFBS. ChIP analyses verified this hypothesis, demonstrating that AHCTF1 bound not only this region of the OSE, but also to the *CCAT1* promoter. In line with the observed effect of BC21 on the CTCF-AHCTF1 protein-protein interaction, the use of this drug evicted AHCTF1 from chromatin at both the CTCFBS and *CCAT1*. To determine whether or not AHCTF1 binding to these loci was CTCF-dependent, we downregulated CTCF expression by transfection with siCTCF, followed by ChIP analyses. AHCTF1 displayed an approximately 50% reduction in its binding on the CTCFBS of WT HCT116 cells (compared to the siGFP control), indicating the need for the

presence of CTCF. Finally, to better understand the role of the CTCFBS in this process, we performed AHCTF1 ChIP analyses on both the CTCFBS and *CCAT1* gene in WT and D3/E4 mutant cells. The results pointed out that AHCTF1 requires the presence of a functional CTCFBS in order to efficiently bind chromatin on both loci, since its recruitment to chromatin was dramatically decreased in both mutant clones, compared to the wild-type cells.

In order to further examine how these data correlate with the nuclear architecture, we first downregulated AHCTF1 expression in WT HCT116 cells by siRNA transfection and observed the distribution of the OSE in relation to the nuclear periphery by 3D DNA FISH. Knockdown of AHCTF1 evicted the OSE from the periphery ($\leq 0.3 \mu\text{m}$ distance) by almost 50%, compared to the siGFP control, whereas localization in more inner positions of the nucleus (further than $0.3 \mu\text{m}$ from the periphery) was not significantly affected. This set of data indicated that AHCTF1 is needed for the final stretch of the OSE to the periphery/pore and since it has been known from a previous study that AHCTF1 locates also in the nucleoplasm³⁶, despite being an NPC component, that could also mean that CTCF and AHCTF1 most likely interact before the final anchoring of the OSE to the pore. Subsequent ISPLAs demonstrated highest proximity between the two factors within a region 1-2 μm distal to the periphery, which was in accordance with previous similar data from CTCF-NUP133 ISPLAs. Interestingly, the potential for CTCF-AHCTF1 interaction was significantly decreased upon BC21 treatment, indicating once more that a functional TCF4/ β -catenin complex formation might be needed for the establishment of this interaction.

4.9. The *CCAT1*-specific CTCFBS affects the *MYC*-OSE proximity at the nuclear periphery, but does not influence their overall interaction frequencies

At this point, a remarkable discovery of our study was how the CTCFBS regulates the proximities between *MYC* and its OSE in the 3D space of the nucleus. By performing 3D DNA FISH analyses in wild-type and mutant cells, we showed that the two loci approach the nuclear periphery in a coordinated manner in WT HCT116 cells, but both their relative positions towards the periphery and their in-between proximities were abolished in both D3 and E4 mutant cells. Moreover, by further stratifying our data in groups related to allele replication, in an effort to explore any relationship to the cell cycle, we succeeded in demonstrating that within a region spanning up to 1 μm from the periphery the proximity of both *MYC* and the OSE alleles to the periphery was significantly reduced in the CTCFBS-mutant cells and these results showed significance only in the case of un-replicated alleles (single-single), indicating that the gating principle might be taking place specifically during the G1 phase of the cell cycle. However, the absence of a functional CTCFBS did not affect the overall frequency of chromatin-chromatin interactions between *MYC* and the OSE in D3 and E4 cells, as provided by Nodewalk analysis, in comparison to wild-type cells. Consequently, CTCF influences the proximity between *MYC* and the OSE at the periphery, without, though, directly affecting their interactions, displaying a non-canonical function.

4.10. *CCAT1* eRNA expression is activated by WNT through the CTCFBS to induce the juxtaposition of the OSE to the nuclear periphery

Lastly, an interesting finding that emerged from our study in Paper II was the involvement of the long non-coding RNA *CCAT1* in the gating of *MYC*. The reasoning behind our choice of studying a potential implication of this particular lncRNA lies mainly in the aforementioned previous report by Xiang et al²⁵⁴ (see Introduction) showing that the *CCAT1* eRNA mediates *MYC*-OSE contacts, potentially through RNA-RNA interactions¹⁵⁹. Starting with an RNA-seq analysis, we first identified the *CCAT1* eRNA isoforms expressed in our model cell line, HCT116. As we found that only the long isoform *CCAT1-L* was expressed in these cells, but not the previously identified *CCAT1-5L*, we then designed specific primers for it and performed RNA FISH analyses, followed by DNA FISH on *MYC* and the OSE. Our observations showed that the distribution of transcriptionally active *CCAT1* alleles (RNA FISH signal) in WT HCT116 cells peaked at a position proximal to the nuclear periphery (~1 μ m distance) and declined when approached very close to the periphery/pore. However, our data did not provide any clear correlation between *CCAT1* eRNA transcription (strength or localization) and *MYC*-OSE proximities, which comes in line with our former results indicating closest proximity between the two loci at a distance <0.5 μ m from the periphery, where the presence of *CCAT1* eRNA seems to be decreasing. Even though the same pattern in the accumulation of the RNA FISH signal was followed in both D3 and E4 cells, the intensity of the observed signal was lower in both mutant cell line, but more pronounced in D3.

In an effort to further explore the relationship between the positioning of the OSE in respect to the periphery and the distribution of transcriptionally active *CCAT1* alleles, we stratified the localization of the OSE according to the strength of the RNA FISH signal. Through this analysis, we were therefore able to document that within a certain distance, close but not precisely at the periphery, closer proximity of the OSE to the periphery was combined with higher *CCAT1* transcription in wild-type cells, indicating a clear and significant correlation between these two factors. Despite maintaining the same trend, D3 and E4 mutants demonstrated a much weaker correlation, pointing out the need for a functional connection between the CTCFBS, the *CCAT1* gene and the ability of the OSE to juxtapose close to the nuclear periphery. The link between *CCAT1* transcription and CTCFBS was finally examined by qRT-PCR analyses of both total and newly synthesized *CCAT1* RNA. Our results showed that *CCAT1* expression levels were dramatically diminished (~4-fold) in both mutant clones compared to wild-type cells. Similarly, the transcriptional rate of *CCAT1*, upon normalization to TBP transcription, was found to be significantly lower in both D3 and E4 cells in comparison with WT HCT116 cells. Interestingly and in line with our previous findings, BC21 inhibited the transcription of *CCAT1* in wild-type cells, whereas no significant effect was observed in mutant cells. Taken together, these observations helped us define the role of *CCAT1* eRNA in the gene gating mechanism, which, upon WNT-dependent activation mediated by CTCF, drives the localization of the OSE at the nuclear periphery.

5 DISCUSSION

5.1. Gene gating mechanism in humans

The studies of the present thesis provided clear evidence for the existence of a gene gating mechanism, as a level of post-transcriptional regulation of gene expression, in humans. Our findings agree with previous observations in other organisms, like yeast and *Drosophila*^{28,51,52}, where gene gating is a well-established principle. They did not only provide the first genetic evidence for such a mechanism, via the specific mutation of the CTCFBS, but also suggested a well-defined multistep model for this process. We showed that G. Blobel's hypothesis⁵⁰ holds true also for human cells, although the principle does not involve a simple juxtaposition of a gene to the nuclear pore complex. Instead, it's a more complex procedure that includes members of the NPC, the genome organizer CTCF, the canonical WNT signaling pathway, a long non-coding RNA and finally, the function of an oncogenic super-enhancer. In particular, the mechanism of action of this OSE is of a significant interest, since it does not increase *MYC* expression by controlling its transcription, as it is more often observed in the majority of enhancers, but instead it conspires with the nuclear architecture to regulate the expression of *MYC* post-transcriptionally. Moreover, the use of gating as a mechanism to protect *MYC* transcripts from the rapid degradation rate of the nucleus demonstrates an example of the complexity a cancer cell works and suggests alternative ways for a cancer cell to adapt and proliferate. However, it still remains to be elucidated whether other cancer genes are gated as well following a similar procedure aiming to cancer cell survival and proliferation or if gene gating is used as a general mechanism to increase the expression of other disease-associated genes in the respective pathological conditions, such as metabolic diseases. Even though we have not yet investigated the potential gating of *FAM49B* in similar detail, our observations in Paper II increase the possibility for the CTCFBS to either directly or indirectly regulate the expression of that gene, in a manner similar to *MYC*. Therefore, its role as a reinforcer of the *MYC* function^{278,279} is in line with such a scenario.

5.2. The role of the NPC in the regulation of gene expression

Another important aspect of our research is offering new insights in the role of the nuclear pore complex at the regulation of gene expression. Despite our well-established knowledge regarding the role of certain nuclear compartments, such as the generally repressive environment of the lamina at the nuclear periphery and the transcriptionally permissive environment around the nucleolus, the exact role of the NPC was not well understood. Our studies have clarified how the NPC and its components are involved in the control of transcription and gene expression, showing that the nuclear pores might function as a transcriptionally permissive platform that promotes the rapid expression of certain genes.

In contrast to other nucleoporins, not much research has been previously done on AHCTF1 and its functions. Until this day, the only well-known role of AHCTF1 was in the assembly of the nuclear pore complex after mitosis³⁵. Our studies assigned a new function for it in the regulation of gene expression by facilitating chromatin movements and transitions within the nuclear architecture. In addition, we uncovered new interactions with other generally important factors for the cell, such as CTCF, β -catenin and TCF4. This discovery could pave the way for further research, as for example on whether AHCTF1 might be involved in other, canonical, functions of CTCF, such as genome organization and the formation of chromatin loops or function as an insulator in the expression of certain target genes. Furthermore, given of the known role of CTCF in the regulation of circadian transcription¹⁸⁹, it would be interesting to examine whether AHCTF1 has a similar association with the circadian cycle, in parallel with its link to the cell cycle³⁵.

Finally, our studies pointed out as well the important role of another not well studied nucleoporin, NUP133. Even though its frequent interactor NUP98 has been previously implicated in a variety of studies on the control of gene expression^{23,25,26,37,39,43,44}, the exact role of NUP133 in this context was not well understood. Therefore, our data revealed its involvement in the post-transcriptional regulation of expression by a gating mechanism, through the establishment of contacts with the OSE and anchoring of the latter to the pores. Moreover, its prominent protein complex formation with CTCF could raise the same questions that apply for AHCTF1.

5.3. A novel role for CTCF

CTCF is generally known as a genome organizer protein. Its main function lies in the establishment of chromatin fiber interactions and the formation of chromatin loops that might promote transcription, by facilitating enhancer action, or inhibit it. However, our findings in Paper II uncovered a novel role for CTCF in the regulation of gene expression, independently of its canonical function. CTCF, through its binding to the CTCFBS of the OSE, is able to modulate the proximities between *MYC* and the OSE, as well as their positioning in relation to the nuclear periphery, without affecting at all their overall interactions. Therefore, its newly discovered role focuses on the upregulation of *MYC* gene expression by driving its OSE-mediated gating. That is in accordance with a previous study suggesting that CTCF can have alternative roles in the cell, apart from the well-established ones, such as the regulation of circadian gene transcription¹⁸⁹. In fact, having CTCF as a common central core in both studies, it becomes very tempting to speculate that a circadian gene gating mechanism might exist, too. That would mean that cells could employ this mechanism in order to rapidly change their expression programs during the day and cover certain, time-specific, needs. An indication that could reinforce even more this idea is that unpublished data from our research have shown that both AHCTF1 expression and binding to chromatin (CTCFBS, in particular) might be under circadian control as well. Of course, that could pave the way for a whole new area of studies and open a Pandora's box.

5.4. Involvement of a lncRNA, *CCAT1*, in the post-transcriptional regulation of gene expression

As in detail elaborated in the Introduction of the present thesis, non-coding RNAs have been reported to play multiple roles in the regulation of gene expression, especially in cancer, but past studies were positioning them mostly in the control of the transcriptional process^{245,248,250,253,254}. However, Paper II demonstrates that a widely known lncRNA, *CCAT1*, whose coding gene is located in a region covered by the OSE controlling *MYC* expression through gating, interestingly, regulates expression at the post-transcriptional level. This provides new insights into the general mechanism of action and function of lncRNAs in the cell, while emphasizing even more on their importance in the epigenetic regulation. It is very intriguing that the interplay between *CCAT1* and the gating of *MYC* is a two-way relationship and might act more as a feedback loop. That is based on the observation that *CCAT1* needs the main factors involved in gating, i.e. the presence of a functional CTCFBS, as well as WNT signaling, in order to get transcriptionally active and promote gating, and at the same time gating needs the presence of *CCAT1* transcripts in order to efficiently take place. However, contrary to previous claims²⁵⁴, we have failed to see any possible involvement of that lncRNA in the binding of CTCF on *MYC* or its OSE. Potential RNA-protein interactions should be further investigated regarding this process, although the length of *CCAT1-L* RNA complicates the application of any immunoprecipitation methods.

5.5. A new example of how chromatin responds to external signals

The core idea of the general field of epigenetics is based on the impact external stimuli might have on chromatin conformation and thus, gene expression. Signals from the environment, like nutrients, temperature, hormones and their association to cell signaling pathways can induce a quick epigenetic activation or silencing, most often via the direct control of transcription, through the recruitment of histone modification or DNA methylation writer and reader factors, such as the repressive Polycomb complex^{1,7,8,13}. The studies included in the present thesis demonstrated a new example of how an external signal, like the presence of WNT ligands, can alter gene expression programs. Interestingly, this signal is translated not as changes in the primary chromatin structure or as large chromosomal movements within the nuclear architecture, but rather as the driving force that would promote transitional protein-protein interactions and chromatin binding by certain factors. That in turn will ultimately increase expression at the post-transcriptional level through gene gating. Our data clearly indicate that gating is WNT-dependent, highlighting once more the importance of WNT signaling in tumor development. Importantly, since excessive *MYC* expression induces WNT signaling in breast cancer²⁸⁰, this suggests the existence of complex feed-back and feed-forward loops between *MYC* and the WNT pathway.

Therefore, another intriguing question rising based on these findings is whether other cell signaling pathways might follow a similar gene gating mechanism for the regulation of

expression of their target genes. For instance, it has been known for years that TGF-beta signaling participates in the epithelial to mesenchymal transitions and thus, promotes metastasis in cancer²⁸¹. Consequently, it would be interesting to investigate whether this TGF-beta-based process would accordingly employ a gene gating procedure.

5.6. Potential involvement of other factors in gene gating

Another interesting aspect to examine would be whether and how gating is connected with other mechanisms and machineries, such as DNA repair or splicing. In fact, it is logical to speculate that factors, with normally different functions, could play a role in gating, too. For instance, unpublished data from our group indicate that PARP1, a well-known factor involved in the DNA repair machinery, physically interacts with both AHCTF1 and NUP133 members of the NPC, while its relationship with CTCF has been very well-established previously from our group¹⁸⁹ and other studies¹⁶⁴. It is quite important to mention at this point that the function of CTCF requires PARylation by PARP1 and *vice versa*, CTCF in turn promotes the self-PARylation of PARP1¹⁶⁴. Moreover, the same set of data show that, similarly to AHCTF1 and CTCF, PARP1 binds the CTCFBS as well. That would suggest that a whole new protein interactome might be hiding behind this mechanism and that the process might be more complicated than we think.

Another quite important candidate to examine is Topoisomerases and particularly, TOP1 and TOP2B. These factors induce DNA breaks in order to release torsional stress generated during transcription and DNA replication. Interestingly, it has been known that the activity of topoisomerase 1 (TOP1), an enzyme recently linked to super-enhancer activation, eRNA production, enhancer-promoter loop formation and efficient transcriptional elongation at gene bodies via releasing transcription-induced torsional stress^{282,283}, is inhibited by SUMOylation²⁸⁴ and that certain NUPs can bind to SUMO proteases (SEN1 and SEN2) to potentially increase TOP1 activity²⁸⁵. Therefore, it would be exciting to study whether the nuclear pores and nucleoplasmic NUPs provide a platform for the activation of TOP1 through the removal of SUMO modifications to facilitate gene expression in general and gene gating in particular. That idea could be further supported by preliminary data of our research detecting the presence of DNA breaks at *MYC* promoter.

Apart from TOP1, findings from previous studies have also provided evidence for a potential involvement of TOP2B in the gating of *MYC* or other genes. Precisely, *MYC* was recently shown to interact with both TOP1 and TOP2B, recruiting them at gene promoters and directly stimulating their levels and activities at promoters, enhancers and gene bodies²⁸⁶. In turn, TOP2B has also been documented to localize at *MYC* promoter and its enhancers, promoting its expression²⁸⁷. Interestingly, the TOP2B levels and genomic localization was linked with *MYC* expression in glioma specimens, but not in non-tumoral human brain tissue. Lastly, a link between TOP2 and the CTCFBS could derive by extrapolation, since a recent study by Gothe et al. showed that recurrent DNA double-strand breaks induced by TOP2 occur at CTCF

binding sites located at the bases of chromatin loops and their frequency positively associates with transcriptional output and directionality²⁸⁸.

Finally, another process with potential association or direct involvement in gene gating could be splicing. As in detail mentioned in the Introduction of this thesis, splicing can practically occur at several points between the transcription and the nuclear export of an RNA. During that multistep procedure, the transcripts usually follow a long movement within the nucleus. At the same time, the data provided by our first study indicate that most likely the transcription of *MYC* is initiated at a more inner space of the nucleus, while the elongation and final transcript release take place after the recruitment to the pores during gating. Moreover, an association with the nuclear speckles before the final anchoring to the nuclear pore cannot be excluded as a possibility, and since the speckles are enriched in components of the spliceosome, that hypothesis could be even further supported.

5.7. Reproducibility and stochastic chromatin networks driven by *MYC*

A strength of the research documented in this thesis is the high reproducibility of our findings between the two papers, as well as compared to previous publications from our research group. In fact, several observations from the first study, such as the closest *MYC*-OSE proximity at the periphery/pore (<0.5 μm distance), were successfully reproduced in Paper II, emphasizing on the solidity and re-validation of the produced data. Moreover, the Nodewalk technique was further validated after application in both studies, demonstrating its advantages in the use for the study of chromatin fiber interactions with high resolution and reproducibility. Our findings on the gating mechanism in the control of *MYC* expression in cancer cells are in agreement with the previous study by Sumida et al. uncovering an acquired ability of *MYC* to get dynamically activated as a response to a wide variety of environmental signals encountered by the cell during the tumorigenic process¹⁴⁸. Furthermore, we developed a standardized methodology for the study of nuclear mRNA export, newly synthesized transcripts and decay rate, which, upon use of the suggested simulation model, can predict the estimated steady-state cytoplasmic mRNA levels (i.e. levels of total expression) with high precision. In addition, the use of the Chromatin In Situ Proximity method in Paper I pointed out the necessity for such a technique due to its ability to link higher order chromatin structures in relation to structural hallmarks of the nuclear architecture.

5.8. Advances in the understanding of and the fight against cancer

Despite the aforementioned novelties and strengths of our research, the most valuable of its features is the advances it offers in the understanding of such a complex disease that is cancer. As evidently shown, gene gating is a mechanism of post-transcriptional regulation of gene expression taking place in human colon cancer, but not in normal primary colon epithelial cells. In fact, cancer cells could use it as a way of increasing the expression of oncogenes pivotal for

their survival and acute proliferation. Therefore, a logical question emerging is whether gating is a feature exclusive to colon cancer or it occurs in the cells of other cancer types as well. Since we are discussing an essential molecular biology mechanism appearing at the very core of the disease initiation or development, it is very tempting to speculate that this can be the case for more types of cancer. That could also be explained by the fact that *MYC* is an oncogene involved, in multiple ways, in almost the 70% of known cancers¹⁹³.

Several lines of evidence support this idea. First, preliminary data from our group show that some indications of gating are also identified in breast cancer tissue samples. Second, it has been known that excessive WNT signaling is linked to high-risk neuroblastomas that do not exhibit *MYCN* amplification²⁸⁹, which suggests that *MYC* paralogues might be gated too. Since the gating of *MYC* in our studies is a process mediated by the colorectal-specific oncogenic super-enhancer, it would be reasonable to consider a similar mechanism of function for other tissue-specific oncogenic super-enhancers, such as the ones identified recently in oesophageal squamous cell carcinoma (OSCC)²⁹⁰, the super-enhancer driving *MYCN* overexpression in neuroblastoma²⁹¹, or in medulloblastoma^{292,293}. In the concrete, cancer-associated super-enhancers mark putative oncogenes and lineage-restricted genes that under normal conditions serve as essential regulators of cell proliferation and apoptosis, like *MYC*. Such super-enhancers have been also detected in small-cell lung carcinoma (SCLC)²⁹⁴, breast²⁹⁵ and gastric²⁹⁶ cancers, as well melanoma²⁹⁷. As seen from the example of HCT116 cells versus HCECs in our studies, these super-enhancers are unique to cancer cells and generally absent in healthy, untransformed cells of identical lineage, which suggests that they might be acquired during the process of tumorigenesis and manifest the oncogenic state in fundamental ways^{293,298}.

Finally, it remains to be elucidated whether gene gating could occur also in healthy cells, as, for instance, a mechanism of maintaining gene expression programs or quick adaptation to changes of their environment. *MYC* is certainly not gated in normal colon epithelial cells, but we cannot exclude the possibility that other genes that the cell requires in abundance, such as house-keeping genes, might be gated instead. Alternatively, we could question whether enhancer-mediated gating can take place in normal cells involving super-enhancers. Such an example is represented by the super-enhancers located within the *Igh* locus²⁹⁸. This idea is supported by the discovery that regions in the 5'-flank of *Igh* promote its positioning to the nuclear periphery in mouse plasma cells that express high levels of cytoplasmic *Igh* transcripts²⁹⁹. However, it is of remarkable interest that administration of WNT3a to primary cultures of HCECs in our study (Paper I), not only did not increase *MYC* expression, but instead it induced its downregulation and did not increase the presence of β -catenin in the nucleus.

5.9. Promises for new cancer treatment

In conclusion, it is noteworthy to mention the promises these studies hold for new therapies against cancer. Since the gating of *MYC* seems to be absent in healthy human colon epithelial

cells, our research has paved the way for the identification of not only diagnostic, but also therapeutic strategies to aim at the effects of pathological *MYC* expression without disrupting its normal function. For instance, a new diagnostic tool could be provided by the presence of CTCF within the *CCAT1* gene, in order to define if pathological *MYC* expression is associated with gene trafficking to nuclear pores. In addition, our discovery has provided the knowledge for several key steps involved in this process that could function as plausible therapeutic targets in order to downregulate *MYC* expression in cancer. Such an example could be inhibitors targeting the partners of CTCF, since its binding to the OSE represents a critical and central point in the WNT pathway.

Moreover, our data indicate potential solutions to the well-established problem faced in cancer treatment due to chemoresistance. Interestingly, it has been previously argued that cancer cells can become addicted to super-enhancer-driven oncogenic transcription and that dependence could function as an Achille's heel for their chemoresistance³⁰⁰. Therefore, the discovery of gene gating mechanism in humans may help in the development of new drugs or in the reconsideration of existing ones to improve the specificity and sensitivity of cancer cells to chemotherapy drugs. In this context, our data showed that BC21 or its derivatives consist a very prominent option for cancer therapy. Furthermore, another promising drug is represented by Olaparib, since it has been shown to inhibit the physical interaction between CTCF and PARP1 at the nuclear periphery¹⁸⁹. Additionally, CDK inhibitors, such as Roscovitine or Flavopiridol, could also work as possible drugs for cancer treatment targeting the gating of *MYC*. Interestingly, Roscovitine, a CDK inhibitor that reduces *MYC* protein levels³⁰¹, inhibited the facilitated export of *MYC* mRNAs, according to unpublished data from our group. In fact, in these experiments we observed using 3D DNA FISH that although Roscovitine antagonized the facilitated export of *MYC* transcripts in colon cancer cells, it did not interfere with the recruitment of the super-enhancer and the *MYC* loci to the nuclear pores. Moreover, it did not inhibit the binding of NUP133 to the super-enhancer. Therefore, as Roscovitine and its derivatives have shown promising effects in the therapy of cancer patients in clinical trials, these results might provide novel perspectives for therapeutic strategies aiming at reducing *MYC* expression. Taken together, these suggestions for treatment are reinforced by our observation on the ability of the CTCFBS to provide a WNT-dependent excessive growth advantage to human colon cancer cells.

6 CONCLUSIONS

The present thesis has uncovered a novel mechanism of post-transcriptional regulation of *MYC* gene expression in humans termed gene gating. We have shown that in human colon cancer cells, HCT116, the oncogenic super-enhancer (OSE) of *MYC* increases the expression of its target by mediating the recruitment of *MYC* to the nuclear pores and promoting the rapid nuclear export of *MYC* mRNA to the cytoplasm. This process is cancer-specific, since it does not occur in normal human colon epithelial cells (HCECs), where the OSE is absent. Moreover, we have demonstrated that the OSE-mediated gating of *MYC* is regulated by the mobile component of the nuclear pore complex AHCTF1 and it is a WNT signaling/ β -catenin-dependent procedure, since it can be inhibited by the β -catenin and TCF inhibitor BC21. Finally, as the degradation of *MYC* transcripts is more rapid in the nucleus than in the cytoplasm, our findings suggest that gating functions as a post-transcriptional mechanism to facilitate their escape from the nuclear degradation system, ultimately increasing their total levels in the cytoplasm and thus, expression.

Following up these findings, the research of this thesis has also provided the first genetic evidence of a gating mechanism, as well as deciphered the molecular factors and steps involved in it. More precisely, we have shown that the WNT-dependent gating of *MYC* is controlled by a CTCF binding site (CTCFBS) within the OSE that confers a growth advantage to colon cancer cells. To achieve this, first, the CTCFBS mediates the transcription of the *CCAT1* eRNA, which in turn will move the OSE to a peripheral position. Then, the CTCFBS is bound by AHCTF1 to promote the final anchoring to the nuclear pore. Thus, a WNT/ β -catenin-AHCTF1-CTCF-eRNA circuit allows the OSE to coordinate the trafficking of the transcriptionally active *MYC* gene within the 3D space of the nucleus and promote pathological cell growth, in a procedure that requires a non-canonical function of CTCF.

In conclusion, our studies offer new insights into the processes underlying tumor development and assist the better understanding of the complex cancer disease. In addition, they indicate novel plausible therapeutic targets for cancer treatment, highlighting new promises in the medical science.

7 POINTS OF PERSPECTIVE

The content of this thesis has shed light upon several long-last enigmas in biomedical research and the field of epigenetics in particular. However, it also paves the way for future research and a variety of applications that derive from it.

In this regard, it is first of all really important to examine the possibility of a genome-wide gene gating mechanism. Using advanced microscopy methods and NGS approaches, it is now possible to investigate what part of the genome is gated and its underlying functional meaning. For instance, the construction of a map with all the cancer-related genes that are being gated in every different type of cancer and the comparison of those maps between them, as well as with their respective healthy cells, can provide us with a more clear picture regarding the heterogeneity of tumors and potential patterns emerging in the disease. Therefore, the need for analysis of gene gating possibilities in clinical samples from cancer patients is really prominent. That would open a whole new era in molecular medicine and the fight against cancer. In parallel with that, further research to uncover whether and what part of the genome in normal cells is being gated would broaden our horizons about how we perceive the biology underlying the general function of a cell or tissue.

Another cancer-related outcome resulting from our research is the clinical application of BC21 or similar acting drugs. In one possibility, new pharmacological studies could apply the combination of BC21 with other known drugs, for example used in chemotherapies, to specifically target and inhibit the gating of *MYC*, without affecting its normal function or having toxic effects on healthy cells. The issue at hand is that this strategy would allow for much lower doses of normally harmful pharmaceuticals. In addition, the advantage of such a combinatorial approach for clinical trials is that a cocktail of two or more drugs targeting different pathological pathways in the cell, could dramatically decrease the fitness of cancer cells and their ability to adapt and survive the treatment. Alternatively, new similar research could be established in order to develop new inhibitors for other key players or steps of the gating mechanism, such as against AHCTF1 binding on the CTCFBS specifically, either against pathological *CCAT1* eRNA transcription or CTCF protein-protein interactions with AHCTF1 or β -catenin/TCF4.

Apart from the direct medical interest, an additional, different possible application of our findings lies in the possibility of a circadian gene gating mechanism, or, in other words, the regulation of circadian gene expression through a gating/trafficking function. It has been known for some time that *MYC* expression is under the control of the circadian clock^{302,303}. It might therefore be rewarding to examine whether the nuclear pore provides a platform for the circadian activation of *MYC* or, alternatively, whether the gating of *MYC* follows circadian oscillations during a certain time period. In fact, since not much previous research on the connection between the cell cycle and the circadian cycle is currently available, such a study would be of particular interest as *MYC* has been alleged to regulate both of them. In conclusion,

our studies encourage the examination of whether other genes whose expression is under circadian control or genes coding the main factors of the circadian core clock machinery, such as CLOCK, BMAL1, PER and CRY, are gated too.

8 ACKNOWLEDGEMENTS

My PhD was a long journey, the end of which I would have never reached without the help of certain people that I need to thank and specially mention.

First of all, I would like to thank my main supervisor, **Anita Göndör**. It was an honor and great pleasure to work with you and be your student. Thank you for teaching me, inspiring me, appreciating my work and training me into becoming an independent researcher, but as well as for supporting my career and always being there to hear and discuss my scientific and general life problems. You were one of the best supervisors I could have ever asked for and I really valued our communication. Finally, thank you for always helping me troubleshoot my methods, letting me develop my creativity and taking into consideration my ideas.

Second, I would like to thank Professor **Rolf Ohlsson**. Your passion about science really inspired me through this long journey and made me believe in my dream of having an academic career. Moreover, your great experience and knowledge taught me to always work hard to keep improving and educating myself as a scientist. But most importantly, I would like to thank you for believing in me and giving me in first place the opportunity to join Anita's group and make this beautiful experience come true. I will really miss our interesting scientific -and not only- discussions.

Moreover, I would really like to thank my co-supervisor, Professor **Galina Selivanova**. I particularly enjoyed our talks and getting precious scientific and career advice from you. Thank you for agreeing to be my co-supervisor, but also for all the interest you showed about my progress during the whole period of my PhD. Your great knowledge and academic experience have been an important inspiration for me.

A person I really owe a lot to is my second co-supervisor, colleague and friend, Assistant Professor **Barbara Scholz**. I think words are not enough to describe how thankful and grateful I am to you. Thank you for teaching me many of the things I know today as a researcher, for extensively helping me with my experiments, for educating me inside and outside the lab. But most importantly, thank you for always being there and supporting me from the very first day of my PhD. Our everyday talks (which I will miss pretty much!) motivated me and helped me many times to get through the hard periods of my studies, as well as to focus on my goals.

Next, I would like to especially thank my mentor and coordinator of the Marie S. Curie "Chromatin3D" consortium, that me and my group participated in during the first 3 years of my PhD, Professor **Charalampos (Babis) Spilianakis**. I was very lucky and honored to be knowing you and be your student in Biochemistry and Epigenetics at the University of Crete, since over a decade ago. Our common experience at Chromatin3D was very helpful to me and I particularly enjoyed and appreciated the useful feedback and advice I took from you in all of our consortium's meetings. You were one of the best professors I've ever had and I feel I learnt a lot from you, both as an undergraduate and a PhD student.

Of course, the completion of my PhD would have never been accomplished without the help of all the amazing colleagues and past members of Anita Göndör's and Rolf Ohlsson's research group I had the luck and honor to work with through the years.

Mirco Martino. My friend and "scientific brother", I consider myself very lucky and privileged to have met and worked side by side with you. From the very first day we started together our common PhD journey, you have been more than helpful, inside and outside the lab, and "saved" me more times than I can remember. Your patience, deep knowledge, organization skills, dedication and good heart are some of the things I really admire on you. Keep going and good luck with finishing your PhD soon as well!

Ilyas Chachoua. Thank you for our beautiful scientific talks and for all the hard work we put in together from the very beginning for the Nat Com paper! Your motivation, curiosity and strong dedication to science inspired me a lot during my PhD, as well as for the next steps of my career. It was a great pleasure working next to you both at the lab and office space.

Jia Pei Lim. You have been a great colleague and friend. Thank you for all the help you provided with my experiments, microscope analysis and general project discussions. There are no words to describe your amazing character and kind heart. I really enjoyed working with you and I wish you all the best for the future!

Carolina Diettrich Mallet de Lima. You were one of the first friends I made in the lab. It was a great experience working with you. You have always been so helpful and supportive! Thank you particularly for your help regarding my experiments for the Nat Com paper.

Marta Imreh. There are no words to describe how helpful and kind you have been from the first moment I joined our research group. Thank you for taking care perfectly of everything in our lab management and especially my/our orders. Your patience has been remarkable and it was always a pleasure taking advice from you about life in Sweden!

Felipe Beccaria Casagrande. You have been one of the kindest colleagues I have ever had. Thank you for all the help you offered me inside and outside the lab, as an equally good friend. I wish you all the best in your career and I'm sure you will do great things!

Andrej Nikoshkov. My office-mate and my favourite Russian, it was a great pleasure working and talking with you. I enjoyed sitting next to you in the office space and having discussions about everything. You are one of the kindest people I have ever met and the positive energy you bring to the group is really unique!

Johanna Vestlund. Thank you for all your help and patience at the lab. I really admire your responsible and organized way of working as a PhD student and I'm sure you will very soon be defending your thesis as well. Keep up the good work and I wish you all the best for the rest of your PhD, as well as for your new role as a mother!

Deeksha Bhartiya & Rashid Mehmood. You are two of the kindest and most helpful colleagues I've ever had. Thank you so much for all your patience, great energy and willingness to teach me and explain to me bioinformatics, R and everything else relevant to the dry lab. Keep up the great work!

Apart from the current members of our research group, it is really important to remember and always appreciate the past members who contributed greatly in all this effort through the years.

The co-authors of the two papers: **Noriyuki Sumida, Emmanouil (Manos) Sifakis, Honglei Zhao, Hao Dai, Anna Lewandowska-Ronnegren, Shuangyang Wu.** Thank you for your great efforts, hard work and dedication to make this happen. I would also like to especially thank you for teaching me methods in the lab, for the nice team work and the wonderful friendship I developed with all of you! I wish you the best in your careers.

Lluís Millan-Arino, Agnieszka (Aga) Walczak, Farzaneh Shahin Varnoosfaderani. Thank you for all your help during the first years of my PhD. It was a pleasure working with you in the lab and office, as well as learning from you during my first steps as a doctoral researcher.

Next, since my PhD project was part of the **Marie Skłodowska-Curie Actions “Chromatin3D” consortium**, funded by the EU, I would really like to thank all the people working in this (PIs, coordinators, manager), as well as the committee funding my research. Especially, the PhD students who are part of the Chromatin3D network and turned out to be some of my best friends all over Europe: **Ourania, Nefeli, Ioannis-Marios, Tomas, Thodoris, Alessandro, Florian, Iwona, Nastia, Nina, David, Nikolas, Himanshu.** Thank you for the great support, helpful scientific comments and the beautiful memories at our annual meetings and conferences.

As part of the consortium, I would specifically like to thank Professor **George Garinis** and post-doctoral researcher **Kalliopi (Kallina) Stratigi** for hosting me in their group during my secondment at the IMBB, FORTH in Greece and teaching me the BLESS method I could later use in my projects.

Besides my colleagues, I would like to acknowledge the contribution and help of the people I would have achieved nothing without: my friends.

First, all the people from the 6th floor in Bioclinicum, that I had the great pleasure to spend a lot of time with both during and outside my working hours: **Konstantina, Amineh (Amy), Matko, Sonia, Sylvy, Angelos, Mireia.** I really enjoyed our afterwork drinking sessions, our lunch conversations and coffee breaks at the kitchen! A special mention to the people of the Lars Holmgren research group in OnkPat, that I had the great pleasure to work with during my time spent in the cell culture lab: **Aarren, Pedro, Vicky.** You made the cell culture lab work a bit more fun and I especially loved talking with the guys about football, Liverpool and life in general. Thank you for all the great time and I wish you all the best in your studies and future careers!

My Greek friends, **Dimitris Kanellis, Sophia Miliara, Georgia Tsikala, Evi Tserga, Maria Sali, Konstantinos Toskas, Dimitris Chioureas.** Ήσασταν οι πρώτοι φίλοι που έκανα στην Στοκχόλμη και πάντα θα έχετε μία ξεχωριστή θέση στην καρδιά μου, σαν μια οικογένεια, όπως σας αποκαλούσα πάντα. Σας ευχαριστώ απεριόριστα που είστε φίλοι μου, που με βοηθήσατε να προσαρμοστώ και να ξεκινήσω τη ζωή μου εδώ, για όλες τις όμορφες αναμνήσεις και την επιστημονική βοήθεια. **Γιωργίτσα,** θα είσαι πάντα στο μυαλό και την καρδιά μας και δεν θα ξεχάσω ποτέ την ευγενική σου ψυχή, το καλοσυνάτο χαμόγελο και την πάντα ευχάριστη διάθεση και προθυμία σου για να με βοηθήσεις με όλα, από τα προπτυχιακά μου κιόλας χρόνια στην Κρήτη. Σε ευχαριστώ για όλα.

My friends, **Vasilis G, Nikos P, Matina R, Marcus S, Christina T, Ioanna T, Dimitris S, Giannis M, Athena E, Konstantinos F,** or as I like to call you, my brothers and sisters. You are the best friends I could ever wish for in Stockholm. Thank you for always being there for me and supporting me in my most difficult moments. Thank you for all your patience hearing me nagging about my problems, for the most amazing memories and time we spent together. I will really miss you and especially miss our gatherings and game nights, our fun times at my place, the barbecue sessions, the brunches, and so much more. You are all amazing people, really dedicated to science, extremely smart and talented, and I am sure you will have great careers in science! I love you all. It's not a goodbye, it's just a see you later.

Mireia Cruz De los Santos. I will really miss our talks and fika breaks at Bioclinicum. Thank you for always being there for me and for being a wonderful friend. And of course, thank you so much for making this amazing artwork for the cover of my thesis! I wish you the best of luck and great success in your PhD.

Nerea Martín Almazán. Thank you for your great support and for tolerating my mood swings while I was working on the Nature Communications paper (Paper II). I deeply appreciate and will never forget that. You have a beautiful heart, a great mind and a big love for science. I am really happy I met you. Keep up the good work, don't give up on your dreams and goals and I'm sure you will achieve great things! Best of luck for your PhD and remember to always MUTIM.

Elin, Klaudia, Orlinda. You have been more than great friends, even though we came close just at the end of this very long journey. But I couldn't be more grateful for all your help and mental support when things in my life were really rough. Thank you! Plus: Elin, thank you so much for your help with the translation of the popular science summary in Swedish!

My quiz "Mediterranean" friends. Gabriele, Nuria, Alba, Ramon. Your friendship and support helped me stay focused on my goal and made the hard work of my research much easier to go through and more enjoyable. I will really miss our quiz nights and victories at Kappa!

My best friends in Greece:

Panos, Spilios, Christos. Σας ευχαριστώ απεριόριστα που μου στέκεστε όλα αυτά τα χρόνια και μου δίνετε δύναμη και κουράγιο να προχωράω σε αυτό το δύσκολο ταξίδι της επιστήμης. Είστε ένας από τους λόγους για να γυρίζω πάντα στην Ελλάδα.

Dionysis, Antonis, Giorgos. «Αααα, πήρε διδακτορικό». Αδέρφια μου, σας ευχαριστώ πολύ για αυτή την όμορφη διαδρομή από το Ηράκλειο μέχρι σήμερα. Ήσασταν πάντα εκεί, κάτι παραπάνω από απλοί φίλοι. Σε όλα μαζί, στα όμορφα και στα άσχημα, πάνω από 12 χρόνια φίλοι και συνεχίζουμε. Σας αγαπώ. Να είστε πάντα έτσι.

It's very important to always remember how essential coffee is in our work. With that, I would like to especially thank **Gast** and **Kaffeverket** cafés in Stockholm for supporting with many cups of coffee and wonderful pastries the writing of the present thesis, on an almost daily basis.

Last, but definitely not least, I would like to thank the three people to whom I owe everything and none of my dreams in life would have ever become true without them: **my father, mother and sister. Μπαμπά, μαμά, Γωγώ:** σας ευχαριστώ για όλα όσα κάνατε για μένα όλα αυτά τα χρόνια, από το Γυμνάσιο και Λύκειο μέχρι σήμερα. Σας ευχαριστώ που μου δώσατε τα εφόδια για να γίνω ο επιστήμονας που πάντα ονειρευόμουν και για την στήριξη και τόση αγάπη που τόσο απλόχερα μου προσφέρατε σε κάθε δύσκολη στιγμή. Ήταν ένα πολύ μακρύ ταξίδι, αλλά επιτέλους τα κατάφερα και έφτασα στο τέλος του με επιτυχία! Μα πάνω απ' όλα, σας ευχαριστώ που πιστέψατε σε μένα από την πρώτη στιγμή και δεν αμφιβάλατε ποτέ για τις δυνατότητες και την αξία μου. Σας οφείλω τα πάντα. **Γιαγιά, παππού:** και σε εσάς το ίδιο. Μπορεί να μην είστε σήμερα μαζί μας, αλλά αυτή μου η επιτυχία είναι και δική σας. Σας ευχαριστώ για όλα. **Μπαμπά:** Ήσουν και είσαι ένας από τους λόγους που πάντα θα προσπαθώ για το καλύτερο. Μου έμαθες να πιστεύω στον εαυτό μου, να μην τα παρατάω και να στοχεύω πάντα την κορυφή. Ελπίζω να σε έκανα υπερήφανο. Όλο μου το διδακτορικό είναι εσύ και από 'δω και στο εξής, κάθε επιτυχία μου στην επιστήμη και κάθε δημοσίευση θα είναι αφιερωμένη σε εσένα.

I got my first data for Paper II on 2017 and the person I called that day to share my great happiness with was my father. Since that day, we had been discussing about it, he was excited about my findings and was looking forward to seeing it published. Unfortunately, he's not here to see it anymore, but this work is dedicated to his memory. Thank you for everything.

9 REFERENCES

1. Felsenfeld, G. A brief history of epigenetics. *Cold Spring Harbor Perspectives in Biology* **6**, (2014).
2. Almouzni, G. & Cedar, H. Maintenance of epigenetic information. *Cold Spring Harbor Perspectives in Biology* **8**, (2016).
3. Kim, K., Lee, K., Bang, H., Kim, J. Y. & Choi, J. K. Intersection of genetics and epigenetics in monozygotic twin genomes. *Methods* **102**, 50–56 (2016).
4. Barchuk, A. R. *et al.* Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*. *BMC Developmental Biology* **7**, (2007).
5. Noble, D. Conrad Waddington and the origin of epigenetics. *Journal of Experimental Biology* **218**, 816–818 (2015).
6. Pujadas, E. & Feinberg, A. P. Regulated noise in the epigenetic landscape of development and disease. *Cell* vol. 148 1123–1131 (2012).
7. Suganuma, T. & Workman, J. L. Signals and Combinatorial Functions of Histone Modifications. *Annual Review of Biochemistry* **80**, 473–499 (2011).
8. Tzelepis, I., Martino, M. & Göndör, A. A Brief Introduction to Chromatin Regulation and Dynamics. in *Chromatin Regulation and Dynamics* 1–34 (Elsevier, 2017). doi:10.1016/B978-0-12-803395-1.00001-0.
9. Berdasco, M. & Esteller, M. Crosstalk Between Non-Coding RNAs and the Epigenome in Development. in *Chromatin Regulation and Dynamics* 211–234 (Elsevier, 2016). doi:10.1016/B978-0-12-803395-1.00009-5.
10. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* vol. 403 41–45 (2000).
11. Liang, G. *et al.* Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 7357 (2004).
12. Teissandier, A. & Bourc’his, D. Gene body DNA methylation conspires with H3K36me3 to preclude aberrant transcription. *The EMBO Journal* **36**, 1471 (2017).
13. Feinberg, A. P., Koldobskiy, M. A. & Göndör, A. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nature Reviews Genetics* vol. 17 284–299 (2016).

14. Cremer, T. & Cremer, M. Chromosome Territories. *Cold Spring Harbor Perspectives in Biology* **2**, a003889–a003889 (2010).
15. Cremer, T. *et al.* Chromosome territories - a functional nuclear landscape. *Current Opinion in Cell Biology* vol. 18 307–316 (2006).
16. Parada, L. A., McQueen, P. G. & Misteli, T. Tissue-specific spatial organization of genomes. *Genome biology* **5**, R44 (2004).
17. Girelli, G. *et al.* GPSeq reveals the radial organization of chromatin in the cell nucleus. *Nature Biotechnology* **38**, 1184–1193 (2020).
18. Guetg, C. & Santoro, R. Formation of nuclear heterochromatin the nucleolar point of view. *Epigenetics* **7**, 811–814 (2012).
19. Reddy, K. L. & Feinberg, A. P. Higher order chromatin organization in cancer. *Seminars in Cancer Biology* **23**, 109–115 (2013).
20. Deng, B., Melnik, S. & Cook, P. R. Transcription factories, chromatin loops, and the dysregulation of gene expression in malignancy. *Seminars in Cancer Biology* vol. 23 65–71 (2013).
21. Scholz, B. A., Millán-Ariño, L. & Göndör, A. 3D Nuclear Architecture and Epigenetic Memories: Regulators of Phenotypic Plasticity in Development, Aging and Cancer. in *Chromatin Regulation and Dynamics* 417–460 (Elsevier, 2017). doi:10.1016/B978-0-12-803395-1.00017-4.
22. 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. *Nature Genetics* **41**, 246–250 (2009).
23. 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).
24. 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. *Developmental Cell* **22**, 446–458 (2012).
25. Liang, Y., Franks, T. M., Marchetto, M. C., Gage, F. H. & Hetzer, M. W. Dynamic Association of NUP98 with the Human Genome. *PLoS Genetics* **9**, (2013).
26. Light, W. H. *et al.* A Conserved Role for Human Nup98 in Altering Chromatin Structure and Promoting Epigenetic Transcriptional Memory. *PLoS Biology* **11**, (2013).

27. Jacinto, F. v, Benner, C. & Hetzer, M. W. The nucleoporin Nup153 regulates embryonic stem cell pluripotency through gene silencing. *Genes and Development* **29**, 1224–1238 (2015).
28. Burns, L. T. & Wentz, S. R. From Hypothesis to Mechanism: Uncovering Nuclear Pore Complex Links to Gene Expression. *Molecular and Cellular Biology* **34**, 2114–2120 (2014).
29. Björk, P. & Wieslander, L. Integration of mRNP formation and export. *Cellular and Molecular Life Sciences* **74**, 2875–2897 (2017).
30. Goldberg, M. W. & Allen, T. D. High resolution scanning electron microscopy of the nuclear envelope: demonstration of a new, regular, fibrous lattice attached to the baskets of the nucleoplasmic face of the nuclear pores. *Journal of Cell Biology* **119**, 1429–1440 (1992).
31. Rout, M. P. *et al.* The Yeast Nuclear Pore Complex: Composition, Architecture, and Transport Mechanism. *The Journal of Cell Biology* vol. 148 <http://www.jcb.org> (2000).
32. Alber, F. *et al.* The molecular architecture of the nuclear pore complex. *Nature* **450**, 695–701 (2007).
33. Bilokapic, S. & Schwartz, T. U. 3D ultrastructure of the nuclear pore complex. *Current Opinion in Cell Biology* vol. 24 86–91 (2012).
34. Fernandez-Martinez, J. & Rout, M. P. A jumbo problem: Mapping the structure and functions of the nuclear pore complex. *Current Opinion in Cell Biology* vol. 24 92–99 (2012).
35. Rasala, B. A., Orjalo, A. v., Shen, Z., Briggs, S. & Forbes, D. J. ELYS is a dual nucleoporin/kinetochore protein required for nuclear pore assembly and proper cell division. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 17801–17806 (2006).
36. Gillespie, P. J., Khoudoli, G. A., Stewart, G., Swedlow, J. R. & Blow, J. J. ELYS/MEL-28 Chromatin Association Coordinates Nuclear Pore Complex Assembly and Replication Licensing. *Current Biology* **17**, 1657–1662 (2007).
37. Raices, M. & D'Angelo, M. A. Nuclear pore complexes and regulation of gene expression. *Current Opinion in Cell Biology* vol. 46 26–32 (2017).
38. Pascual-Garcia, P. & Capelson, M. Nuclear pores as versatile platforms for gene regulation. *Current Opinion in Genetics and Development* vol. 25 110–117 (2014).

39. Pascual-Garcia, P., Jeong, J. & Capelson, M. Nucleoporin Nup98 associates with Trx/MLL and NSL histone-modifying complexes and regulates Hox gene expression. *Cell Reports* **9**, 433–442 (2014).
40. Buchwalter, A. L., Liang, Y. & Hetzer, M. W. Nup50 is required for cell differentiation and exhibits transcription-dependent dynamics. *Molecular Biology of the Cell* **25**, 2472–2484 (2014).
41. Kehat, I., Accornero, F., Aronow, B. J. & Molkentin, J. D. Modulation of chromatin position and gene expression by HDAC4 interaction with nucleoporins. *Journal of Cell Biology* **193**, 21–29 (2011).
42. Brown, C. R., Kennedy, C. J., Delmar, V. A., Forbes, D. J. & Silver, P. A. Global histone acetylation induces functional genomic reorganization at mammalian nuclear pore complexes. *Genes and Development* **22**, 627–639 (2008).
43. Kalverda, B. & Fornerod, M. Characterization of genome-nucleoporin interactions in *Drosophila* links chromatin insulators to the nuclear pore complex. *Cell Cycle* **9**, 4812–4817 (2010).
44. Franks, T. M. *et al.* Evolution of a transcriptional regulator from a transmembrane nucleoporin. *Genes and Development* **30**, 1155–1171 (2016).
45. Arib, G. & Akhtar, A. Multiple facets of nuclear periphery in gene expression control. *Current Opinion in Cell Biology* **23**, 346–353 (2011).
46. Tan-Wong, S. M., Wijayatilake, H. D. & Proudfoot, N. J. Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes and Development* **23**, 2610–2624 (2009).
47. Brickner, D. G. *et al.* Transcription Factor Binding to a DNA Zip Code Controls Interchromosomal Clustering at the Nuclear Periphery. *Developmental Cell* **22**, 1234–1246 (2012).
48. Breuer, M. & Ohkura, H. A negative loop within the nuclear pore complex controls global chromatin organization. *Genes and Development* **29**, 1789–1794 (2015).
49. van de Vosse, D. W., Wan, Y., Wozniak, R. W. & Aitchison, J. D. Role of the nuclear envelope in genome organization and gene expression. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* vol. 3 147–166 (2011).
50. Blobel, G. Gene gating: a hypothesis. *Proceedings of the National Academy of Sciences* **82**, 8527–8529 (1985).
51. Ben-Yishay, R., Ashkenazy, A. J. & Shav-Tal, Y. Dynamic Encounters of Genes and Transcripts with the Nuclear Pore. *Trends in Genetics* **32**, 419–431 (2016).

52. Pascual-Garcia, P. *et al.* Metazoan Nuclear Pores Provide a Scaffold for Poised Genes and Mediate Induced Enhancer-Promoter Contacts. *Molecular Cell* **66**, 63-76.e6 (2017).
53. Raices, M. & D'Angelo, M. A. Nuclear pore complex composition: A new regulator of tissue-specific and developmental functions. *Nature Reviews Molecular Cell Biology* vol. 13 687–699 (2012).
54. Zhu, L. & Brangwynne, C. P. Nuclear bodies: The emerging biophysics of nucleoplasmic phases. *Current Opinion in Cell Biology* vol. 34 23–30 (2015).
55. Mao, Y. S., Zhang, B. & Spector, D. L. Biogenesis and function of nuclear bodies. *Trends in Genetics* vol. 27 295–306 (2011).
56. Shevtsov, S. P. & Dundr, M. Nucleation of nuclear bodies by RNA. *Nature Cell Biology* **13**, 167–173 (2011).
57. Sleeman, J. E. & Trinkle-Mulcahy, L. Nuclear bodies: New insights into assembly/dynamics and disease relevance. *Current Opinion in Cell Biology* vol. 28 76–83 (2014).
58. Papantonis, A. & Cook, P. R. Transcription factories: Genome organization and gene regulation. *Chemical Reviews* vol. 113 8683–8705 (2013).
59. Zaborowska, J., Egloff, S. & Murphy, S. The pol II CTD: new twists in the tail. *Nature Structural & Molecular Biology* 2016 23:9 **23**, 771–777 (2016).
60. Saldi, T., Cortazar, M. A., Sheridan, R. M. & Bentley, D. L. Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *Journal of Molecular Biology* **428**, 2623–2635 (2016).
61. Naftelberg, S., Schor, I. E., Ast, G. & Kornblihtt, A. R. Regulation of Alternative Splicing Through Coupling with Transcription and Chromatin Structure. <http://dx.doi.org/10.1146/annurev-biochem-060614-034242> **84**, 165–198 (2015).
62. Venkatesh, S. & Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nature Reviews Molecular Cell Biology* 2015 16:3 **16**, 178–189 (2015).
63. Venkatesh, S. *et al.* Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature* 2012 489:7416 **489**, 452–455 (2012).
64. Sims, R. J. *et al.* Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Post-Initiation Factors and pre-mRNA Splicing. *Molecular cell* **28**, 665 (2007).

65. Zhou, H. L., Luo, G., Wise, J. A. & Lou, H. Regulation of alternative splicing by local histone modifications: potential roles for RNA-guided mechanisms. *Nucleic Acids Research* **42**, 701 (2014).
66. Pan, T. & Sosnick, T. RNA FOLDING DURING TRANSCRIPTION. <http://dx.doi.org/10.1146/annurev.biophys.35.040405.102053> **35**, 161–175 (2006).
67. Lai, D., Proctor, J. R. & Meyer, I. M. On the importance of cotranscriptional RNA structure formation. *RNA* **19**, 1461–1473 (2013).
68. Zhao, P., Zhang, W. & Chen, S. J. Cotranscriptional folding kinetics of ribonucleic acid secondary structures. *The Journal of chemical physics* **135**, (2011).
69. Zamft, B., Bintu, L., Ishibashi, T. & Bustamante, C. Nascent RNA structure modulates the transcriptional dynamics of RNA polymerases. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 8948–8953 (2012).
70. Visa, N., Izaurralde, E., Ferreira, J., Daneholt, B. & Mattaj, I. W. A nuclear cap-binding complex binds Balbiani ring pre-mRNA cotranscriptionally and accompanies the ribonucleoprotein particle during nuclear export. *The Journal of cell biology* **133**, 5–14 (1996).
71. Gonatopoulos-Pournatzis, T. & Cowling, V. H. Cap-binding complex (CBC). *The Biochemical journal* **457**, 231–242 (2014).
72. Osheim, Y. N., O.L. Miller, J. & Beyer, A. L. RNP particles at splice junction sequences on Drosophila chorion transcripts. *Cell* **43**, 143–151 (1985).
73. Baurén, G. & Wieslander, L. Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell* **76**, 183–192 (1994).
74. Wetterberg, I., Baurén, G. & Wieslander, L. The intranuclear site of excision of each intron in Balbiani ring 3 pre-mRNA is influenced by the time remaining to transcription termination and different excision efficiencies for the various introns. *RNA* **2**, 641 (1996).
75. Singh, J. & Padgett, R. A. Rates of in situ transcription and splicing in large human genes. *Nature structural & molecular biology* **16**, 1128–1133 (2009).
76. Howard, J. M. & Sanford, J. R. The RNAissance family: SR proteins as multifaceted regulators of gene expression. *Wiley interdisciplinary reviews. RNA* **6**, 93–110 (2015).
77. Boehm, V. & Gehring, N. H. Exon Junction Complexes: Supervising the Gene Expression Assembly Line. *Trends in genetics : TIG* **32**, 724–735 (2016).
78. Hir, H. le, Saulière, J. & Wang, Z. The exon junction complex as a node of post-transcriptional networks. *Nature reviews. Molecular cell biology* **17**, 41–54 (2016).

79. Dower, K. & Rosbash, M. T7 RNA polymerase-directed transcripts are processed in yeast and link 3' end formation to mRNA nuclear export. *RNA (New York, N.Y.)* **8**, 686–697 (2002).
80. Ahn, S. H., Kim, M. & Buratowski, S. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Molecular cell* **13**, 67–76 (2004).
81. Licatalosi, D. D. *et al.* Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Molecular cell* **9**, 1101–1111 (2002).
82. Sagawa, F., Ibrahim, H., Morrison, A. L., Wilusz, C. J. & Wilusz, J. Nucleophosmin deposition during mRNA 3' end processing influences poly(A) tail length. *The EMBO journal* **30**, 3994–4005 (2011).
83. Proudfoot, N. J. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science (New York, N.Y.)* **352**, (2016).
84. Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A. & Burge, C. B. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science (New York, N.Y.)* **320**, 1643–1647 (2008).
85. Yeh, H. S. & Yong, J. Alternative Polyadenylation of mRNAs: 3'-Untranslated Region Matters in Gene Expression. *Molecules and cells* **39**, 281–285 (2016).
86. Eberle, A. B. & Visa, N. Quality control of mRNP biogenesis: networking at the transcription site. *Seminars in cell & developmental biology* **32**, 37–46 (2014).
87. Schmid, M. & Jensen, T. H. Transcription-associated quality control of mRNP. *Biochimica et biophysica acta* **1829**, 158–168 (2013).
88. Custódio, N. *et al.* Inefficient processing impairs release of RNA from the site of transcription. *The EMBO journal* **18**, 2855–2866 (1999).
89. Libri, D. *et al.* Interactions between mRNA export commitment, 3'-end quality control, and nuclear degradation. *Molecular and cellular biology* **22**, 8254–8266 (2002).
90. Davidson, L., Kerr, A. & West, S. Co-transcriptional degradation of aberrant pre-mRNA by Xrn2. *The EMBO journal* **31**, 2566–2578 (2012).
91. Kallehauge, T. B., Robert, M. C., Bertrand, E. & Jensen, T. H. Nuclear retention prevents premature cytoplasmic appearance of mRNA. *Molecular cell* **48**, 145–152 (2012).
92. Spector, D. L. & Lamond, A. I. Nuclear speckles. *Cold Spring Harbor perspectives in biology* **3**, 1–12 (2011).

93. Girard, C. *et al.* Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nature communications* **3**, (2012).
94. Meyer, K. D. & Jaffrey, S. R. The dynamic epitranscriptome: N6-methyladenosine and gene expression control. *Nature reviews. Molecular cell biology* **15**, 313–326 (2014).
95. Hoernes, T. P. & Erlacher, M. D. Translating the epitranscriptome. *Wiley interdisciplinary reviews. RNA* **8**, (2017).
96. Gilbert, W. v., Bell, T. A. & Schaening, C. Messenger RNA modifications: Form, distribution, and function. *Science (New York, N.Y.)* **352**, 1408–1412 (2016).
97. Shav-Tal, Y. & Gruenbaum, Y. Single-molecule dynamics of nuclear mRNA. *F1000 biology reports* **1**, (2009).
98. Shav-Tal, Y. *et al.* Dynamics of single mRNPs in nuclei of living cells. *Science (New York, N.Y.)* **304**, 1797–1800 (2004).
99. Politz, J. C. & Pederson, T. Review: movement of mRNA from transcription site to nuclear pores. *Journal of structural biology* **129**, 252–257 (2000).
100. Bassell, G. J., Oleynikov, Y. & Singer, R. H. The travels of mRNAs through all cells large and small. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **13**, 447–454 (1999).
101. Mor, A. *et al.* Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells. *Nature cell biology* **12**, 543–552 (2010).
102. Braga, J., McNally, J. G. & Carmo-Fonseca, M. A reaction-diffusion model to study RNA motion by quantitative fluorescence recovery after photobleaching. *Biophysical journal* **92**, 2694–2703 (2007).
103. Ma, J. *et al.* High-resolution three-dimensional mapping of mRNA export through the nuclear pore. *Nature communications* **4**, (2013).
104. Siebrasse, J. P., Kaminski, T. & Kubitscheck, U. Nuclear export of single native mRNA molecules observed by light sheet fluorescence microscopy. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 9426–9431 (2012).
105. Grünwald, D. & Singer, R. H. In vivo imaging of labelled endogenous β -actin mRNA during nucleocytoplasmic transport. *Nature* **467**, 604–607 (2010).
106. Smith, C. *et al.* In vivo single-particle imaging of nuclear mRNA export in budding yeast demonstrates an essential role for Mex67p. *The Journal of cell biology* **211**, 1121–1130 (2015).

107. Kelich, J. M. & Yang, W. High-resolution imaging reveals new features of nuclear export of mRNA through the nuclear pore complexes. *International journal of molecular sciences* **15**, 14492–14504 (2014).
108. Knockenhauer, K. E. & Schwartz, T. U. The Nuclear Pore Complex as a Flexible and Dynamic Gate. *Cell* **164**, 1162–1171 (2016).
109. Lund, M. K. & Guthrie, C. The DEAD-box protein Dbp5p is required to dissociate Mex67p from exported mRNPs at the nuclear rim. *Molecular cell* **20**, 645–651 (2005).
110. Lejeune, F., Ishigaki, Y., Li, X. & Maquat, L. E. The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. *The EMBO journal* **21**, 3536–3545 (2002).
111. Katahira, J. mRNA export and the TREX complex. *Biochimica et biophysica acta* **1819**, 507–513 (2012).
112. Köhler, A. & Hurt, E. Exporting RNA from the nucleus to the cytoplasm. *Nature Reviews Molecular Cell Biology* 2007 8:10 **8**, 761–773 (2007).
113. Carmody, S. R. & Wentz, S. R. mRNA nuclear export at a glance. *Journal of Cell Science* **122**, 1933–1937 (2009).
114. Fornerod, M., Ohno, M., Yoshida, M. & Mattaj, I. W. CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* **90**, 1051–1060 (1997).
115. Culjkovic, B., Topisirovic, I., Skrabanek, L., Ruiz-Gutierrez, M. & Borden, K. L. B. eIF4E is a central node of an RNA regulon that governs cellular proliferation. *The Journal of cell biology* **175**, 415–426 (2006).
116. Heath, C. G., Viphakone, N. & Wilson, S. A. The role of TREX in gene expression and disease. *The Biochemical journal* **473**, 2911–2935 (2016).
117. Jani, D., Valkov, E. & Stewart, M. Structural basis for binding the TREX2 complex to nuclear pores, GAL1 localisation and mRNA export. *Nucleic acids research* **42**, 6686–6697 (2014).
118. Umlauf, D. *et al.* The human TREX-2 complex is stably associated with the nuclear pore basket. *Journal of cell science* **126**, 2656–2667 (2013).
119. Wickramasinghe, V. O. *et al.* Selective nuclear export of specific classes of mRNA from mammalian nuclei is promoted by GANP. *Nucleic acids research* **42**, 5059–5071 (2014).
120. Lai, M. C. & Tarn, W. Y. Hypophosphorylated ASF/SF2 binds TAP and is present in messenger ribonucleoproteins. *The Journal of biological chemistry* **279**, 31745–31749 (2004).

121. Huang, Y., Gattoni, R., Stévenin, J. & Steitz, J. A. SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Molecular cell* **11**, 837–843 (2003).
122. Delaleau, M. & Borden, K. Multiple Export Mechanisms for mRNAs. *Cells* **4**, 452–473 (2015).
123. Wickramasinghe, V. O. & Laskey, R. A. Control of mammalian gene expression by selective mRNA export. *Nature Reviews Molecular Cell Biology* **16**, 431–442 (2015).
124. Göndör, A. & Ohlsson, R. Chromosome crosstalk in three dimensions. *Nature* vol. 461 212–217 (2009).
125. Gómez-Díaz, E. & Corces, V. G. Architectural proteins: Regulators of 3D genome organization in cell fate. *Trends in Cell Biology* vol. 24 703–711 (2014).
126. Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nature Reviews Genetics* **17**, 772–772 (2016).
127. Holwerda, S. & de Laat, W. Chromatin loops, gene positioning, and gene expression. *Frontiers in Genetics* **3**, 1–13 (2012).
128. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84–98 (2012).
129. Blackwood, E. M. & Kadonaga, J. T. Going the distance: A current view of enhancer action. *Science* vol. 281 60–63 (1998).
130. Alpsoy, A., Sood, S. & Dykhuizen, E. C. At the Crossroad of Gene Regulation and Genome Organization: Potential Roles for ATP-Dependent Chromatin Remodelers in the Regulation of CTCF-Mediated 3D Architecture. *Biology 2021, Vol. 10, Page 272* **10**, 272 (2021).
131. Hnisz, D. *et al.* Convergence of Developmental and Oncogenic Signaling Pathways at Transcriptional Super-Enhancers. *Molecular Cell* **58**, 362–370 (2015).
132. Bahar Halpern, K. *et al.* Bursty gene expression in the intact mammalian liver. *Molecular Cell* **58**, 147–156 (2015).
133. Lam, A. & Deans, T. L. A noisy tug of war: The battle between transcript production and degradation in the liver. *Developmental Cell* vol. 33 3–4 (2015).
134. Calo, E. & Wysocka, J. Modification of Enhancer Chromatin: What, How, and Why? *Molecular Cell* vol. 49 825–837 (2013).
135. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).

136. Creighton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences* **107**, 21931–21936 (2010).
137. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–285 (2011).
138. Zentner, G. E., Tesar, P. J. & Scacheri, P. C. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Research* **21**, 1273–1283 (2011).
139. Natoli, G. & Andrau, J.-C. Noncoding Transcription at Enhancers: General Principles and Functional Models. *Annual Review of Genetics* **46**, 1–19 (2012).
140. Plank, J. L. & Dean, A. Enhancer function: Mechanistic and genome-wide insights come together. *Molecular Cell* vol. 55 5–14 (2014).
141. Noordermeer, D. *et al.* Variegated gene expression caused by cell-specific long-range DNA interactions. *Nature Cell Biology* **13**, 944–951 (2011).
142. Smith, E. & Shilatifard, A. Enhancer biology and enhanceropathies. *Nature Structural and Molecular Biology* vol. 21 210–219 (2014).
143. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
144. Dostie, J. *et al.* Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Research* **16**, 1299–1309 (2006).
145. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing Chromosome Conformation. *Science* **295**, 1306–1311 (2002).
146. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genetics* **38**, 1341–1347 (2006).
147. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
148. Sumida, N. *et al.* MYC as a driver of stochastic chromatin networks: Implications for the fitness of cancer cells. *Nucleic Acids Research* **48**, 10867–10876 (2020).
149. Sumida, N. *et al.* The ultra-sensitive Nodewalk technique identifies stochastic from virtual, population-based enhancer hubs regulating MYC in 3D: Implications for the fitness of cancer cells. *bioRxiv* 286583 (2018) doi:10.1101/286583.

150. Söderberg, O. *et al.* Proximity ligation: a specific and versatile tool for the proteomic era. *Genetic engineering* **28**, 85–93 (2007).
151. Roussis, I. M., Guille, M., Myers, F. A. & Scarlett, G. P. RNA whole-mount in situ hybridisation proximity ligation assay (rISH-PLA), an assay for detecting RNA–protein complexes in intact cells. *PLoS ONE* **11**, 1–13 (2016).
152. Serebryanny, L. A. & Misteli, T. HiPLA: High-throughput imaging proximity ligation assay. *Methods* **157**, 80–87 (2019).
153. Chen, X. *et al.* The visualization of large organized chromatin domains enriched in the H3K9me2 mark within a single chromosome in a single cell. *Epigenetics* **9**, 1439–1445 (2014).
154. Chen, X. *et al.* Chromatin in situ proximity (ChrISP): Single-cell analysis of chromatin proximities at a high resolution. *BioTechniques* **56**, 117–124 (2014).
155. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods* **10**, 1213–1218 (2013).
156. Greil, F., Moorman, C. & van Steensel, B. [16] DamID: Mapping of In Vivo Protein–Genome Interactions Using Tethered DNA Adenine Methyltransferase. *Methods in Enzymology* **410**, 342–359 (2006).
157. Horike, S. I., Cai, S., Miyano, M., Cheng, J. F. & Kohwi-Shigematsu, T. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nature Genetics* **2004 37:1** **37**, 31–40 (2004).
158. Liu, X. *et al.* In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell* **170**, 1028–1043.e19 (2017).
159. Cai, Z. *et al.* RIC-seq for global in situ profiling of RNA–RNA spatial interactions. *Nature* **582**, 432–437 (2020).
160. Klenova, E. M. *et al.* CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Molecular and cellular biology* **13**, 7612–24 (1993).
161. 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. *Molecular and Cellular Biology* **16**, 2802–2813 (1996).

162. Wood, A. M., Garza-Gongora, A. G. & Kosak, S. T. A Crowdsourced nucleus: Understanding nuclear organization in terms of dynamically networked protein function. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* vol. 1839 178–190 (2014).
163. Ong, C.-T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nature Reviews Genetics* **15**, 234–246 (2014).
164. Yu, W. *et al.* Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. *Nature Genetics* **36**, 1105–1110 (2004).
165. MacPherson, M. J., Beatty, L. G., Zhou, W., Du, M. & Sadowski, P. D. The CTCF Insulator Protein Is Posttranslationally Modified by SUMO. *Molecular and Cellular Biology* **29**, 714–725 (2009).
166. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
167. Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO Journal* **32**, 3119–3129 (2013).
168. Seitan, V. C. *et al.* Cohesin-Based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Research* **23**, 2066–2077 (2013).
169. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci USA* **111**, 996–1001 (2014).
170. Gong, Y. *et al.* Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nature Communications* **9**, (2018).
171. Dunlap, J. C. Molecular bases for circadian clocks. *Cell* vol. 96 271–290 (1999).
172. Oishi, K., Higo-Yamamoto, S., Yamamoto, S. & Yasumoto, Y. Disrupted light-dark cycle abolishes circadian expression of peripheral clock genes without inducing behavioral arrhythmicity in mice. *Biochemical and Biophysical Research Communications* **458**, 256–261 (2015).
173. Abitbol, K. *et al.* Clock-dependent and system-driven oscillators interact in the suprachiasmatic nuclei to pace mammalian circadian rhythms. *PLoS ONE* **12**, (2017).
174. Melanson, E. L. *et al.* Daytime bright light exposure, metabolism, and individual differences in wake and sleep energy expenditure during circadian entrainment and misalignment. *Neurobiology of Sleep and Circadian Rhythms* **4**, 49–56 (2017).

175. Tahara, Y. & Shibata, S. Entrainment of the mouse circadian clock: Effects of stress, exercise, and nutrition. *Free Radical Biology and Medicine* (2017) doi:10.1016/j.freeradbiomed.2017.12.026.
176. West, A. C. *et al.* Misalignment with the external light environment drives metabolic and cardiac dysfunction. *Nature Communications* **8**, (2017).
177. Woelders, T., Beersma, D. G. M., Gordijn, M. C. M., Hut, R. A. & Wams, E. J. Daily Light Exposure Patterns Reveal Phase and Period of the Human Circadian Clock. *Journal of Biological Rhythms* **32**, 274–286 (2017).
178. Sheikh-Ali, M. & Maharaj, J. Circadian clock desynchronisation and metabolic syndrome. *Postgraduate Medical Journal* vol. 90 461–466 (2014).
179. Ferrell, J. M. & Chiang, J. Y. L. Circadian rhythms in liver metabolism and disease. *Acta Pharmaceutica Sinica B* vol. 5 113–122 (2015).
180. Vadnie, C. A. & McClung, C. A. Circadian Rhythm Disturbances in Mood Disorders: Insights into the Role of the Suprachiasmatic Nucleus. *Neural Plasticity* **2017**, 1–28 (2017).
181. Shostak, A. Circadian clock, cell division, and cancer: From molecules to organism. *International Journal of Molecular Sciences* **18**, (2017).
182. Paganelli, R., Petrarca, C. & di Gioacchino, M. Biological clocks: Their relevance to immune-allergic diseases. *Clinical and Molecular Allergy* vol. 16 (2018).
183. Koike, N. *et al.* Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science* **338**, 349–354 (2012).
184. Benegiamo, G., Brown, S. A. & Panda, S. RNA dynamics in the control of circadian rhythm. in *Advances in Experimental Medicine and Biology* vol. 907 107–122 (2016).
185. Aguilar-Arnal, L. *et al.* Cycles in spatial and temporal chromosomal organization driven by the circadian clock. *Nature Structural and Molecular Biology* **20**, 1206–1215 (2013).
186. Xu, Y. *et al.* Long-Range Chromosome Interactions Mediated by Cohesin Shape Circadian Gene Expression. *PLoS Genetics* **12**, (2016).
187. Chen, H. *et al.* Functional organization of the human 4D Nucleome. *Proceedings of the National Academy of Sciences* **112**, 8002–8007 (2015).
188. Buhr, E. D. & Takahashi, J. S. Molecular Components of the Mammalian Circadian Clock. In *Circadian clocks* 3–27 (2013) doi:10.1016/S0092-8674(03)00076-X.

189. Zhao, H. *et al.* PARP1- and CTCF-Mediated Interactions between Active and Repressed Chromatin at the Lamina Promote Oscillating Transcription. *Molecular Cell* **59**, 984–997 (2015).
190. Beneke, S. Regulation of chromatin structure by poly(ADP-ribosyl)ation. *Frontiers in Genetics* **3**, 1–16 (2012).
191. Thomas, C. & Tulin, A. v. Poly-ADP-ribose polymerase: Machinery for nuclear processes. *Molecular Aspects of Medicine* vol. 34 1124–1137 (2013).
192. Yagita, K. *et al.* Development of the circadian oscillator during differentiation of mouse embryonic stem cells in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 3846–51 (2010).
193. Amjadi-Moheb, F., Paniri, A. & Akhavan-Niaki, H. Insights into the links between MYC and 3D chromatin structure and epigenetics regulation: Implications for cancer therapy. *Cancer Research* **81**, 1925–1936 (2021).
194. Huppi, K., Pitt, J. J., Wahlberg, B. M. & Caplen, N. J. The 8q24 gene desert: An oasis of non-coding transcriptional activity. *Frontiers in Genetics* **3**, 69 (2012).
195. Dang, C. v., O'Donnell, K. A. & Juopperi, T. The great MYC escape in tumorigenesis. *Cancer Cell* **8**, 177–178 (2005).
196. Meyer, N. & Penn, L. Z. Reflecting on 25 years with MYC. *Nature Reviews Cancer* 2008 8:12 **8**, 976–990 (2008).
197. Kress, T. R., Sabò, A. & Amati, B. MYC: connecting selective transcriptional control to global RNA production. *Nature Reviews Cancer* 2015 15:10 **15**, 593–607 (2015).
198. Amati, B., Frank, S. R., Donjerkovic, D. & Taubert, S. Function of the c-Myc oncoprotein in chromatin remodeling and transcription. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1471**, M135–M145 (2001).
199. Poole, C. J. & van Riggelen, J. MYC—Master Regulator of the Cancer Epigenome and Transcriptome. *Genes* 2017, Vol. 8, Page 142 **8**, 142 (2017).
200. Knoepfler, P. S. *et al.* Myc influences global chromatin structure. *The EMBO Journal* **25**, 2723–2734 (2006).
201. Amati, B. *et al.* Oncogenic activity of the c-Myc protein requires dimerization with Max. *Cell* **72**, 233–245 (1993).
202. Amati, B. *et al.* Transcriptional activation by the human c-Myc oncoprotein in yeast requires interaction with Max. 24. Hille, B. *Ionic Channels of Excitable Membranes* **11**, 597–603 (1968).

203. Desbarats, L., Gaubatz, S. & Eilers, M. Discrimination between different E-box-binding proteins at an endogenous target gene of c-myc. *Genes & Development* **10**, 447–460 (1996).
204. Sabò, A. *et al.* Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis. *Nature* 2014 511:7510 **511**, 488–492 (2014).
205. Shostak, A. *et al.* MYC/MIZ1-dependent gene repression inversely coordinates the circadian clock with cell cycle and proliferation. *Nature Communications* **7**, (2016).
206. Huber, A.-L. *et al.* CRY2 and FBXL3 Cooperatively Degrade c-MYC. *Molecular Cell* **64**, 774–789 (2016).
207. Bachmann, A. S. & Geerts, D. Polyamine synthesis as a target of MYC oncogenes. *The Journal of Biological Chemistry* **293**, 18757 (2018).
208. Dang, C. v., Reddy, E. P., Shokat, K. M. & Soucek, L. Drugging the “undruggable” cancer targets. *Nature Reviews Cancer* 2017 17:8 **17**, 502–508 (2017).
209. Davis, T. L., Firulli, A. B. & Kinniburgh, A. J. Ribonucleoprotein and protein factors bind to an H-DNA-forming c-myc DNA element: Possible regulators of the c-myc gene (transcription element/DNA conformer/gene regulation). *Proc. Naid. Acad. Sci. USA* **86**, 9682–9686 (1989).
210. Berberich, S. & Postel, E. PuF/NM23-H2/NDPK-B transactivates a human c-myc promoter-CAT gene via a functional nuclease hypersensitive element. *undefined* (1995).
211. Gonzalez, V. & Hurley, L. H. The c-MYC NHE III1: Function and Regulation. <http://dx.doi.org/10.1146/annurev.pharmtox.48.113006.094649> **50**, 111–129 (2010).
212. Lancho, O. & Herranz, D. The MYC Enhancer-ome: Long-Range Transcriptional Regulation of MYC in Cancer. *Trends in Cancer* **4**, 810–822 (2018).
213. Thandapani, P. Super-enhancers in cancer. *Pharmacology & Therapeutics* **199**, 129–138 (2019).
214. Schuijers, J. *et al.* Transcriptional Dysregulation of MYC Reveals Common Enhancer-Docking Mechanism. *Cell Reports* **23**, 349–360 (2018).
215. Sotelo, J. *et al.* Long-range enhancers on 8q24 regulate c-Myc. *Proceedings of the National Academy of Sciences* **107**, 3001–3005 (2010).
216. 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 Research* **20**, 1191 (2010).

217. Jia, Y., Chng, W. J. & Zhou, J. Super-enhancers: critical roles and therapeutic targets in hematologic malignancies. *Journal of Hematology & Oncology* 2019 12:1 **12**, 1–17 (2019).
218. Bahr, C. *et al.* A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature* 2018 553:7689 **553**, 515–520 (2018).
219. Chung, C. C. & Chanock, S. J. Current status of genome-wide association studies in cancer. *Human Genetics* 2011 130:1 **130**, 59–78 (2011).
220. Sahasrabudhe, R. *et al.* The 8q24 rs6983267G variant is associated with increased thyroid cancer risk. *Endocrine-Related Cancer* **22**, 841–849 (2015).
221. Mitchell, J. S. *et al.* Genome-wide association study identifies multiple susceptibility loci for multiple myeloma. *Nature Communications* 2016 7:1 **7**, 1–9 (2016).
222. Zhang, M. *et al.* Three new pancreatic cancer susceptibility signals identified on chromosomes 1q32.1, 5p15.33 and 8q24.21. *Oncotarget* **7**, 66328–66343 (2016).
223. Pulikkan, J. A. *et al.* CBF β -SMMHC Inhibition Triggers Apoptosis by Disrupting MYC Chromatin Dynamics in Acute Myeloid Leukemia. *Cell* **174**, 172-186.e21 (2018).
224. Park, J. H., Pyun, W. Y. & Park, H. W. Cancer Metabolism: Phenotype, Signaling and Therapeutic Targets. *Cells* **9**, (2020).
225. Hoffman, B. & Liebermann, D. A. Apoptotic signaling by c-MYC. *Oncogene* **27**, 6462–6472 (2008).
226. Mosimann, C., Hausmann, G. & Basler, K. β -Catenin hits chromatin: regulation of Wnt target gene activation. *Nature Reviews Molecular Cell Biology* **10**, 276–286 (2009).
227. Kohn, A. D. & Moon, R. T. Wnt and calcium signaling: β -Catenin-independent pathways. *Cell Calcium* **38**, 439–446 (2005).
228. Clevers, H. Wnt/ β -Catenin Signaling in Development and Disease. *Cell* **127**, 469–480 (2006).
229. Klaus, A. & Birchmeier, W. Wnt signalling and its impact on development and cancer. *Nature Reviews Cancer* 2008 8:5 **8**, 387–398 (2008).
230. Bilić, J. *et al.* Wnt induces LRP6 signalosomes and promotes dishevelled-dependent LRP6 phosphorylation. *Science* **316**, 1619–1622 (2007).
231. Huang, H. & He, X. Wnt/ β -catenin signaling: new (and old) players and new insights. *Current opinion in cell biology* **20**, 119 (2008).

232. Huber, A. H., Nelson, W. J. & Weis, W. I. Three-Dimensional Structure of the Armadillo Repeat Region of β -Catenin. *Cell* **90**, 871–882 (1997).
233. Xing, Y. *et al.* Crystal Structure of a Full-Length β -Catenin. *Structure (London, England : 1993)* **16**, 478 (2008).
234. Brannon, M., Gomperts, M., Sumoy, L., Moon, R. T. & Kimelman, D. A β -catenin/XTcf-3 complex binds to the siamois promoter to regulate dorsal axis specification in *Xenopus*. *Genes & Development* **11**, 2359 (1997).
235. Fang, M. *et al.* C-terminal-binding protein directly activates and represses Wnt transcriptional targets in *Drosophila*. *The EMBO Journal* **25**, 2735 (2006).
236. Daniels, D. L. & Weis, W. I. β -catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nature Structural & Molecular Biology* 2005 12:4 **12**, 364–371 (2005).
237. Li, J. *et al.* CBP/p300 are bimodal regulators of Wnt signaling. *The EMBO Journal* **26**, 2284–2294 (2007).
238. Parker, D. S., Ni, Y. Y., Chang, J. L., Li, J. & Cadigan, K. M. Wingless Signaling Induces Widespread Chromatin Remodeling of Target Loci. *Molecular and Cellular Biology* **28**, 1815–1828 (2008).
239. Kalkat, M. *et al.* MYC Deregulation in Primary Human Cancers. *Genes* 2017, Vol. 8, Page 151 **8**, 151 (2017).
240. Rennoll, S. & Yochum, G. Regulation of MYC gene expression by aberrant Wnt/ β -catenin signaling in colorectal cancer. *World journal of biological chemistry* **6**, 290–300 (2015).
241. Posternak, V., Ung, M. H., Cheng, C. & Cole, M. D. MYC Mediates mRNA Cap Methylation of Canonical Wnt/ β -Catenin Signaling Transcripts By Recruiting CDK7 and RNA Methyltransferase. *Molecular Cancer Research* **15**, 213–224 (2017).
242. Iaccarino, I. lncRNAs and MYC: An Intricate Relationship. *International Journal of Molecular Sciences* 2017, Vol. 18, Page 1497 **18**, 1497 (2017).
243. Tao, J., Tao, J. & Zhao, X. c-MYC–miRNA circuitry. <http://dx.doi.org/10.4161/cc.27646> **13**, 191–198 (2014).
244. Hart, J. R. *et al.* MYC regulates the non-coding transcriptome. *Oncotarget* **5**, 12543–12554 (2014).
245. Gupta, R. A. *et al.* Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 2010 464:7291 **464**, 1071–1076 (2010).

246. Tripathi, V. *et al.* Long Noncoding RNA MALAT1 Controls Cell Cycle Progression by Regulating the Expression of Oncogenic Transcription Factor B-MYB. *PLOS Genetics* **9**, e1003368 (2013).
247. Gutschner, T. *et al.* The Noncoding RNA MALAT1 Is a Critical Regulator of the Metastasis Phenotype of Lung Cancer Cells. *Cancer Research* **73**, 1180–1189 (2013).
248. Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409–419 (2010).
249. Yoon, J. H. *et al.* LincRNA-p21 Suppresses Target mRNA Translation. *Molecular Cell* **47**, 648–655 (2012).
250. Ling, H. *et al.* CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. *Genome research* **23**, 1446–1461 (2013).
251. Nissan, A. *et al.* Colon cancer associated transcript-1: a novel RNA expressed in malignant and pre-malignant human tissues. *International journal of cancer* **130**, 1598–1606 (2012).
252. Chung, S. *et al.* Association of a novel long non-coding RNA in 8q24 with prostate cancer susceptibility. *Cancer science* **102**, 245–252 (2011).
253. Kim, T. *et al.* Long-range interaction and correlation between MYC enhancer and oncogenic long noncoding RNA CARLo-5. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 4173–4178 (2014).
254. Xiang, J. F. *et al.* Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Research* **24**, 513–531 (2014).
255. Wang, C. *et al.* C-Myc-activated long non-coding RNA PVT1 enhances the proliferation of cervical cancer cells by sponging miR-486-3p. *The Journal of Biochemistry* **167**, 565–575 (2020).
256. Cho, S. W. *et al.* Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell* **173**, 1398-1412.e22 (2018).
257. Zhao, X. *et al.* Disruption of the MYC-miRNA-EZH2 loop to suppress aggressive B-cell lymphoma survival and clonogenicity. *Leukemia* 2013 27:12 **27**, 2341–2350 (2013).
258. Grossmann, T. N. *et al.* Inhibition of oncogenic Wnt signaling through direct targeting of β -catenin. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 17942–17947 (2012).

259. Kudo, N. *et al.* Leptomycin B Inhibition of Signal-Mediated Nuclear Export by Direct Binding to CRM1. *Experimental Cell Research* **242**, 540–547 (1998).
260. Göndör, A. *et al.* Window into the complexities of chromosome interactomes. *Cold Spring Harbor symposia on quantitative biology* **75**, 493–500 (2010).
261. Khan, A. & Zhang, X. dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic acids research* **44**, D164–D171 (2016).
262. Sur, I. K. *et al.* Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* **338**, 1360–1363 (2012).
263. Jackson, D. A., Pombo, A. & IBORRA Sir William, F. The balance sheet for transcription: an analysis of nuclear RNA metabolism in mammalian cells; The balance sheet for transcription: an analysis of nuclear RNA metabolism in mammalian cells. doi:10.1096/fasebj.14.2.242.
264. Singh, S. B., Genilloud, O. & Peláez, F. Terrestrial Microorganisms – Filamentous Bacteria. in *Comprehensive Natural Products II* vol. 2 109–140 (Elsevier, 2010).
265. Hansen, M. M. K., Desai, R. v., Simpson, M. L. & Weinberger, L. S. Cytoplasmic Amplification of Transcriptional Noise Generates Substantial Cell-to-Cell Variability. *Cell systems* **7**, 384–397.e6 (2018).
266. Zuo, E. *et al.* GOTI, a method to identify genome-wide off-target effects of genome editing in mouse embryos. *Nature Protocols* 2020 15:9 **15**, 3009–3029 (2020).
267. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology* 2013 31:3 **31**, 213–219 (2013).
268. Kim, S. *et al.* Strelka2: fast and accurate calling of germline and somatic variants. *Nature Methods* 2018 15:8 **15**, 591–594 (2018).
269. Wilm, A. *et al.* LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research* **40**, 11189 (2012).
270. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology* 2020 38:3 **38**, 276–278 (2020).
271. Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnology* 2011 29:1 **29**, 24–26 (2011).
272. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* 2001 409:6822 **409**, 860–921 (2001).
273. Ibarra, A., Benner, C., Tyagi, S., Cool, J. & Hetzer, M. W. Nucleoporin-mediated regulation of cell identity genes. *Genes and Development* **30**, 2253–2258 (2016).

274. Schwartzy, M., Travesa, A., Martell, S. W. & Forbes, D. J. Analysis of the initiation of nuclear pore assembly by ectopically targeting nucleoporins to chromatin. *Nucleus (Austin, Tex.)* **6**, 40–54 (2015).
275. Togel, L. *et al.* Dual Targeting of Bromodomain and Extraterminal Domain Proteins, and WNT or MAPK Signaling, Inhibits c-MYC Expression and Proliferation of Colorectal Cancer Cells. *Molecular Cancer Therapeutics* **15**, 1217–1226 (2016).
276. 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).
277. Voloshanenko, O. *et al.* β -catenin-independent regulation of Wnt target genes by RoR2 and ATF2/ATF4 in colon cancer cells. *Scientific Reports* 2018 8:1 **8**, 1–14 (2018).
278. Benitah, S. A., Frye, M., Glogauer, M. & Watt, F. M. Stem cell depletion through epidermal deletion of Rac1. *Science (New York, N.Y.)* **309**, 933–935 (2005).
279. Fort, L. *et al.* Fam49/CYRI interacts with Rac1 and locally suppresses protrusions. *Nature Cell Biology* 2018 20:10 **20**, 1159–1171 (2018).
280. Cowling, V. H. & Cole, M. D. Turning the Tables: Myc Activates Wnt in Breast Cancer. <http://dx.doi.org/10.4161/cc.6.21.4880> **6**, 2625–2627 (2007).
281. McDonald, O. G., Wu, H., Timp, W., Doi, A. & Feinberg, A. P. Genome-scale epigenetic reprogramming during epithelial-to-mesenchymal transition. *Nature Structural and Molecular Biology* **18**, 867–874 (2011).
282. Puc, J. *et al.* Ligand-dependent enhancer activation regulated by Topoisomerase-I activity. *Cell* **160**, 367–380 (2015).
283. Baranello, L. *et al.* RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription. *Cell* **165**, 357–371 (2016).
284. Li, M., Pokharel, S., Wang, J.-T., Xu, X. & Liu, Y. RECQ5-dependent SUMOylation of DNA topoisomerase I prevents transcription-associated genome instability. *Nature Communications* **6**, 6720 (2015).
285. Chow, K.-H., Elgort, S., Dasso, M. & Ullman, K. S. Two distinct sites in Nup153 mediate interaction with the SUMO proteases SENP1 and SENP2. *Nucleus* **3**, 349–358 (2012).
286. Das, S. K. *et al.* MYC assembles and stimulates topoisomerases 1 and 2 in a “topoisome”. *Molecular cell* **82**, 140-158.e12 (2021).

287. Gonzalez-Buendia, E. *et al.* TOP2B Enzymatic Activity on Promoters and Introns Modulates Multiple Oncogenes in Human Gliomas. *Clinical Cancer Research* **27**, 5669–5680 (2021).
288. Gothe, H. J. *et al.* Spatial Chromosome Folding and Active Transcription Drive DNA Fragility and Formation of Oncogenic MLL Translocations. *Molecular Cell* **75**, 267–283.e12 (2019).
289. Liu, X. *et al.* Deregulated Wnt/ β -catenin program in high-risk neuroblastomas without MYCN amplification. *Oncogene* 2008 27:10 **27**, 1478–1488 (2007).
290. Jiang, Y. Y. *et al.* Targeting super-enhancer-associated oncogenes in oesophageal squamous cell carcinoma. *Gut* **66**, 1358–1368 (2017).
291. Chipumuro, E. *et al.* CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer. *Cell* **159**, 1126–1139 (2014).
292. Lin, C. Y. *et al.* Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature* **530**, 57 (2016).
293. Sengupta, S. & George, R. E. Super-enhancer-driven Transcriptional Dependencies in Cancer. *Trends in cancer* **3**, 269 (2017).
294. Christensen, C. L. *et al.* Targeting Transcriptional Addictions In Small Cell Lung Cancer With a Covalent CDK7 Inhibitor. *Cancer cell* **26**, 909 (2014).
295. Wang, Y. *et al.* CDK7-Dependent Transcriptional Addiction in Triple-Negative Breast Cancer. *Cell* **163**, 174 (2015).
296. Ooi, W. F. *et al.* Epigenomic profiling of primary gastric adenocarcinoma reveals super-enhancer heterogeneity. *Nature Communications* **7**, (2016).
297. Zhou, B. *et al.* INO80 governs superenhancer-mediated oncogenic transcription and tumor growth in melanoma. *Genes and Development* **30**, 1440–1453 (2016).
298. Hnisz, D. *et al.* Super-Enhancers in the Control of Cell Identity and Disease. *Cell* **155**, 934–947 (2013).
299. Yang, Q., Riblet, R. & Schildkraut, C. L. Sites that direct nuclear compartmentalization are near the 5' end of the mouse immunoglobulin heavy-chain locus. *Molecular and cellular biology* **25**, 6021–6030 (2005).
300. Li, G. H. *et al.* Super-enhancers: a new frontier for epigenetic modifiers in cancer chemoresistance. *Journal of Experimental & Clinical Cancer Research* 2021 40:1 **40**, 1–17 (2021).

301. Cassandri, M. *et al.* CDK9 as a Valuable Target in Cancer: From Natural Compounds Inhibitors to Current Treatment in Pediatric Soft Tissue Sarcomas. *Frontiers in Pharmacology* **11**, 1230 (2020).
302. 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 - Gene Regulatory Mechanisms* **1859**, 541–552 (2016).
303. 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. *Chronobiology International* **27**, 722–741 (2010).