

Transcriptional control in cancer

Ilaria Cherchi

Telegram: @ilariacherchi

Elisa Pettinà

Telegram: @elispettina

Github: <https://github.com/ilariache/Transcriptional-control-in-cancer>

December 19, 2022

Contents

1 Transcription	5
1.1 Transcription initiation	5
1.2 Chromatin remodeling	7
1.3 Nucleosome organization	8
1.3.1 Nucleosome remodelling	8
1.3.2 Chromatin Immunoprecipitation	10
1.3.3 Main nucleosome modifications in promoters	10
1.4 Transcription elongation	14
1.4.1 Structural basis of the nucleosome transition during RNA pol II passage	14
1.5 Transcription termination	18
1.5.1 Promoter reporter assay	18
1.5.2 Promoter types	19
1.5.3 Transcription factor binding	20
2 Transcription regulation and its deregulation in cancer	22
2.1 Mediator complex	22
2.2 Wnt signalling pathway	24
2.2.1 β -catenin structure	25
2.2.2 The canonical Wnt β -catenin pathway is crucially involved in colorectal cancer	26
2.3 CBP/p300	28
2.3.1 CBP/p300 structure	28
2.3.2 CBP/p300 in cancer	29
2.4 TSSa-RNAs transcription	30
2.5 Enhancers	31

CONTENTS

2.5.1	Properties of eRNAs	31
2.5.2	eRNAs proposed functions	32
2.5.3	Assessing whether the CRC 8q24 risk region acts as an enhancer	32
2.5.4	The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling	33
2.5.5	CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer	34
3	Chromosome domains and oncogene activation	37
3.1	Genome organization and transcription regulation	37
3.1.1	CTCF	37
3.1.2	Cohesin mediates transcriptional insulation by CCCTC-binding factor	38
3.1.3	CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation	40
3.2	Chromosome Conformation Capture (3C)	42
3.3	Hi-C approach	42
3.3.1	Optimization of Hi-C approach	43
3.4	Mediator and cohesin connect gene expression and chromatin architecture	46
3.4.1	Promoter-Enhancer Communication Occurs Primarily within Insulated neighbourhoods	47
3.4.2	Depletion of Mediator causes loss of PIC assembly	48
3.5	Super-enhancers	50
3.5.1	How is a SE generated during tumorigenesis?	51
3.5.2	How cancer cells take advantage of super-enhancers	53
3.6	The role of MYC in cancer cells	54
3.6.1	Key aspects of MYC in tumorigenesis	54
3.6.2	By which major mechanism or pathway do MYC proteins promote tumorigenesis?	56
4	Targeting transcription in cancer	58
4.1	Age-related remodelling of oesophageal epithelia by mutated cancer drivers	58
4.2	The risk of suffering any cancer increases during aging	60
4.3	Targeted therapeutics generally lead to resistance	61
4.3.1	Mutations in cancer	61
4.4	Target transcription in cancer	62
4.5	BRD4	64

CONTENTS

4.5.1	Selective inhibition of Tumor Oncogenes by Disruption of Super-Enhancers	64
4.6	YAP/TAZ	65
4.7	TEAD	67
4.7.1	YAP/TAZ/TEAD	67
4.8	Nuclear receptors	68
4.8.1	Estrogen receptor α	69
4.8.2	ER α in breast cancer	70
5	Transcription and genome instability	74
5.1	R-loops	74
5.1.1	R-loop degradation	75
5.1.2	VIM-antisense1	76
5.1.3	HOTTIP-dependent R-loop formation regulates CTCF boundary activity and TAD integrity in leukemia	77
5.1.4	HOTTIP lncRNA Promotes Hematopoietic Stem Cells Self-Renewal Leading to AML-like Disease in Mice	79
5.2	DNA damage	79
5.3	R loop modified bases	79
5.3.1	Apical kinases	81
5.3.2	Topoisomerases	83
6	Role of transcription in DNA damage response and repair	86
6.1	Biogenesis of canonical miRNA	86
6.2	DICER and DROSH regulation of DDR and checkpoint enforcement	88
6.3	DDRNAs	88
6.3.1	DDRNAs and dilncRNAs regulate DDR and DNA repair	90
6.4	RNA Pol II recruitment at DSBs	90
7	Telomeres and telomere transcription in cancer	92
7.1	What are telomeres?	92
7.2	Telomeres mask chromosome ends from being recognized as double-strand breaks	92
7.3	Telomere dysfunction in a p53 and Rb mutant setting does not induce cell cycle block	93
7.4	End of replication problem	94
7.4.1	Hayflick limit	95

CONTENTS

7.4.2	Triggering senescence	96
7.5	Cell crisis	97
7.6	Telomerase enzyme	97
7.6.1	Usage of telomerase	97
8	Journal Clubs	99
8.1	Enhancers are activated by p300/CBP activity-dependent PIC assembly, RNAPII recruitment, and pause release	99
8.2	SPT5 stabilization of promoter - proximal RNA polymerase II	100
8.3	MYC Recruits SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation	104
8.4	BRD4-directed super-enhancer organization of transcription repression programs links to chemotherapeutic efficacy in breast cancer	105
8.5	HSF2 cooperates with HSF1 to drive a transcriptional program critical for the malignant state	106

Chapter 1

Transcription

1.1 Transcription initiation

RNA polymerases are *highly processive enzymes*, able to continuously extend transcripts without releasing them, yet unable to resume transcription if they are detached from the nascent RNA.

RNA pol II is responsible for the transcription of all protein-coding genes, plus snoRNA genes, miRNA genes, siRNA genes, lncRNA genes and most snRNA genes. RNA pol II enzyme is composed of 12 subunits.

The **core promoter** is a short sequence encompassing ~ 50bp upstream and ~ 50bp downstream of the TSS consisting of core promoter elements.

We start from the unbound promoter and **TFIID complex**, formed by TAFs (14-15 subunits) and TBP. **TBP** is responsible for the recognition of the TATA box, TAF1-2 for Inr region and TAF6-9 for MTE and DPE.

Once TFIIA and TFIIB reach TFIID, the upstream promoter complex forms. In particular, the promoter region becomes bent. Lastly, Pol II and TFIIF are recruited at the promoter region to initiate transcription.

Once ATP is released, promoter clearance is activated allowing the PIC complex to reach the open conformation and to start transcription. **TFIIC** induces torsional stress by rotating with respect to a fixed site (TBP bound), melting DNA leading to the open complex state (OC). TFIIC promotes

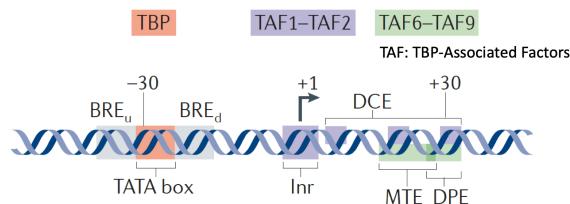


Figure 1.1: Sainsbury et al., Nature Review Mol. Cell. Biol. 2015

1.1. TRANSCRIPTION INITIATION

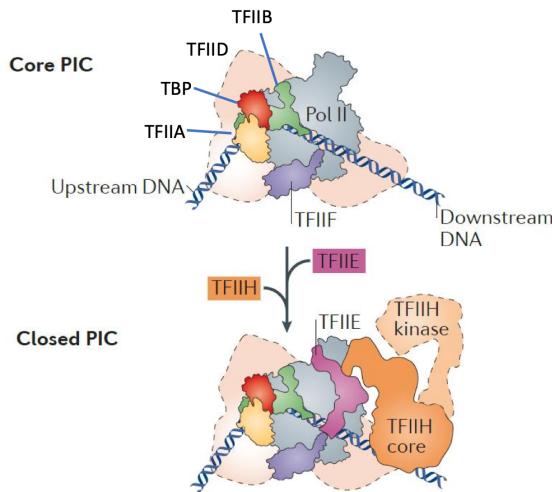


Figure 1.2: Sainsbury et al., Nature Review Mol. Cell. Biol. 2015

unwinding of approximately 10 bp of promoter DNA. When Pol II locates the TSS, it transits into a processive elongation mode. Dissociation of general transcription factors enables the formation of the Pol II elongation complex (addition of elongation factors).

Summarising, transcription initiation follows:

1. TFIID + TBP promoter recognition → PIC assembly
2. TFIIA recruitment stabilises the TBP-DNA complex
3. TFIIB stabilises TBP-DNA and TBP-DNA-Pol II complexes, binds the template strand to position DNA for initiation of RNA synthesis (setting the TSS); stimulates RNA synthesis by allosterically rearranging active site residues and stabilizing closed polymerase clamp
4. TFIIF heterodimer associates with polII and prevents non-specific interaction of polII with DNA; stabilizes the PIC by interacting also with TFIIB; influences TSS selection; stimulates phosphodiester bonds formation
5. TFIIE binds pol II (clamp) and facilitates the recruitment of TFIH to the initiation complex; provides a bridge between polII and TFIH; binds single stranded DNA; TFIIE stabilizes the open promoter; facilitates ATPase/kinase activities of TFIH
6. TFIH opens promoter DNA, phosphorylates RNA pol II CTD; promotes translocation of ssDNA into polII cleft
7. Factors binding modify the structural conformation of the preassembled complex TFIID-TBP-promoter structure becomes visible to TFIIB and A, the hole structure is visible to RNAPol II-TFIIF which is a complex and so on...
8. Pol II is positioned at the core promoter by the combination of TFIID, TFIIA and TFIIB through protein-protein interactions and protein-DNA interactions
9. TFIH then melts 10-15 bp of DNA in order to position the single strand template in the polII cleft (open complex) to initiate RNA synthesis
10. The CTD of PolII will be phosphorylated by the TFIH kinase subunit during the first 30 bp of transcription when PolII loses its contacts with GTFs before proceeding onto the elongation stage

1.2. CHROMATIN REMODELING

How does this start? *What makes a promoter all of a sudden accessible to TBP and so to TFIID?* The DNA sequence is always there, what makes it visible/accessible to the general transcription factors?

1.2 Chromatin remodeling

Nucleosomes are the functional units of protein. In nucleosomes the position of each major groove facing the histone octamer is designed as a **superhelix location** (SHL), numbered from 0 at the Dyad to ± 7 .

Proteins cannot easily associate with DNA sequences that touch the nucleosomal histone surface. Nucleosomal DNA is bent around the histone octamer – target sequences can be distorted or unrecognizable. GC-rich sequences with AA, TT or TA dinucleotides spaces by 10 bp bend more easily and display higher affinity for the histone octamer “nucleosome positioning sequences”.

Nucleosome unwrapping is influenced by histone PTMs. In particular, certain combinations of PTMs display synergistic effects. For instance, covalent modifications of histone proteins can significantly alter the organization and function of chromatin. Examples:

- acetylated H3K56 enhances the unwrapping of the DNA at the entry/exit sites of the nucleosome.
- acetylated H4K91 leads to nucleosome instability - H4K91 lays in the H3-H4 and H2A-H2B interaction surface

Histone modifications may also recruit proteins and enzyme to regulate the chromatin state. Remember that nucleosomal fibers can fold into higher order structures that may be even less accessible.

With the use of energy nucleosomes can be remodelled:

- Histone octamers can slide along DNA
- Can be fully or partially disassembled
- Histones can be replaced with histone variants and post-translationally modified

These processes are carried out by the synergic activity of **chromatin remodeling complexes** and **histone chaperones**.

1.2.0.1 Techniques to study chromatin remodelling

- **MNase:** micrococcal nuclease is able to digest linker DNA, therefore to isolate nucleosomes. We have no information about the sequence, but we can assess the bp length of isolated DNA. Cons: some mapped entities are not nucleosomal, preference for cutting A/T rich regions.
- **ChIP-seq:** after DNA sonication, immunoprecipitation and immunocomplex purification are performed (with protein-specific antibodies). DNA and proteins are cross-linked and purified, then bound DNA is analyzed by massively parallel short-read sequencing. ChIP-seq is relatively simple (technically), but does not have a high resolution due to non-uniform DNA fragmentation. Resolution also depends on library/sequencing setting.

1.3. NUCLEOSOME ORGANIZATION

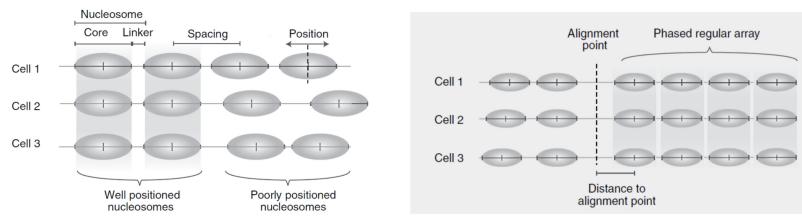


Figure 1.3: Nucleosome occupancy

1.3 Nucleosome organization

Nucleosome organization can be described as a combination of nucleosome occupancy and positioning:

- *Occupancy*: average number of nucleosomes measured within a specific genomic region within a cell population
- *Positioning*: probability of a nucleosome reference point being at a specific genomic coordinate (same dyad)

Most nucleosomes in animal genomes are poorly positioned, but there are regions with **phased arrays** of well positioned nucleosomes (Figure 1.3).

Most promoters contain a nucleosome free region (NFR) and a defined nucleosomal architecture that aids the recruitment of Pol II. A subset of genes may have a nucleosome in their promoter region, which is depleted during gene activation. DNA sequences containing relatively high GC content tend to wrap nucleosomes with higher affinity in vitro.

ATP-dependent chromatin remodelers act at promoter regions. **Remodels the structure of chromatin (RSC)** is a chromatin remodelling factor acting on core promoters. It retains a cavity which can accommodate nucleosomes.

1.3.1 Nucleosome remodelling

The unpeeling of DNA segments from the histone surface can lead to a delocalization of nucleosomes (their sliding), removal or replacement of histones or to the complete eviction of the nucleosome, also dependent on associated histone chaperones (Figure 1.4).

ATP-dependent chromatin remodelling enzymes can act cooperatively with **histone chaperones** to generate or disassemble nucleosomes. Chaperones associate with histones upon their synthesis, escort them into the nucleus, and aid in their specific association with DNA during different processes such as DNA replication, repair, or transcription. Histone Chaperones participate in assembly or disassembly nucleosomes in vitro or in vivo without using the energy of ATP.

Example: SWR and Inositol-requiring protein 80 (INO80) chromatin remodelling complexes regulate histone turnover and histone variant H2A.Z deposition. HIRA complex regulates H3.3 deposition at genes and regulatory elements.

1.3. NUCLEOSOME ORGANIZATION

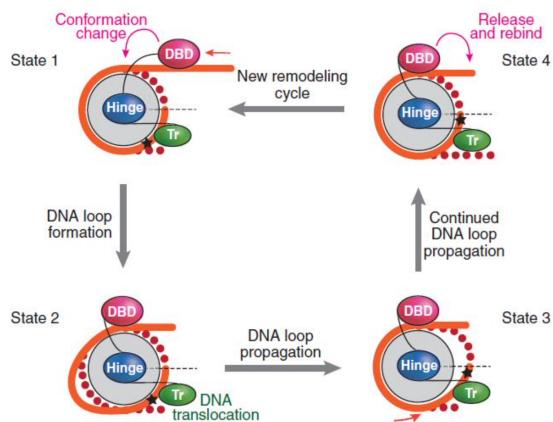


Figure 1.4: Becker and Workman, Cold Spring Harbor Laboratory, 2013

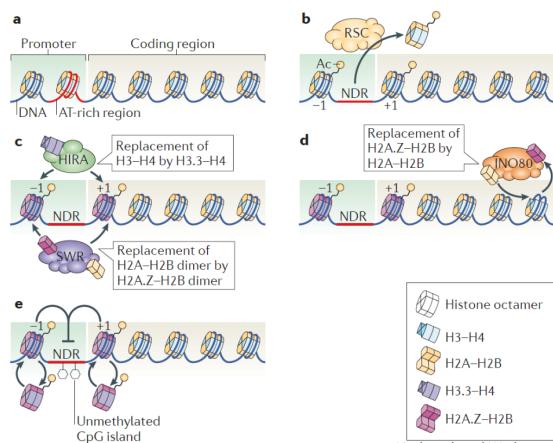


Figure 1.5: Venkatesh and Workman, Nature Rev Mol Cell Biol, 2015

H2A.Z has extended acidic patches on the surface which stimulate remodeling activities and weaken DNA interaction and may play a role in Pol II recruitment - it is lost at promoters upon RNA PolII loading. H2A.Z comprises 15% of total H2A and it is post translationally modified (acetylated, polyubiquitinilate...).

The most rapid **nucleosome turnover** occurs over promoters, tRNA, and small nucleolar RNA genes (often studied in over expression settings).

PIC formation at a core promoter requires a nucleosome depleted region (Figure 1.5). This could be achieved by **RSC** binding to the AT-rich region in the promoter, freeing the NDR. Next, **HIRA** performs the replacement of H3-H4 by H3.3-H4 and **SWR** replaces the H2A-H2B dimer by H2A.Z-H2B dimer. Finally, **INO80** replaces H2A.Z-H2B by H2-A and H2B. The resulting scenario exhibits unmethylated CpG islands in NDR region, allowing for PIC formation.

1.3. NUCLEOSOME ORGANIZATION

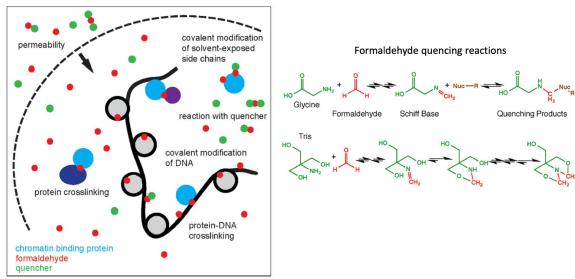


Figure 1.6: Cross linking

1.3.2 Chromatin Immunoprecipitation

(ChIP) experiments to determine protein-DNA interactions require a *crosslinking* step.

1. Crosslinking: formaldehyde and a quencher are required to covalently modify DNA
2. Lysis
3. Antibody binding
4. Immunoprecipitation
5. Wash steps
6. Crosslink reversal
7. DNA purification and quantitative PCR
8. DNA and protein analysis

Chromatin Immunoprecipitation coupled to sequencing analyses can be used to determine protein-DNA interactions.

1.3.3 Main nucleosome modifications in promoters

Most promoters (98%) occupied by Pol II are also occupied by histone H3K4me3 and acetylated H3K9 and H3K14 (Gunter et al., 2007).

- **H3K4me3**-modified nucleosomes are found at sites of transcription. H3K4me3 enrichment is detectable in almost 80% of all protein-coding genes within 1 kb of known or predicted transcript start sites. Only 30-40% of protein coding genes are expressed in human embryonic stem cells. Histone H3K4me3 is observed also at the promoters of genes for which there is no evidence of transcription in ES cells - typically lower than in active genes. Genes not enriched in H3K4me3 are found in clusters along the genome; The vicinity of these genes on the genome suggest that their expression they may be regulated sinergically.
- **H3K9-14Ac** were enriched at the promoters of nearly 70% of genes, including both transcriptionally active and inactive genes. Nearly all of the promoters (>95%) acetylated on H3K9 and H3K14 were also enriched for H3K4me3.
- **H3K36me3** and **H3K79me2** modifications occur almost exclusively downstream of promoters that produce detectable transcripts in ES cells

1.3. NUCLEOSOME ORGANIZATION

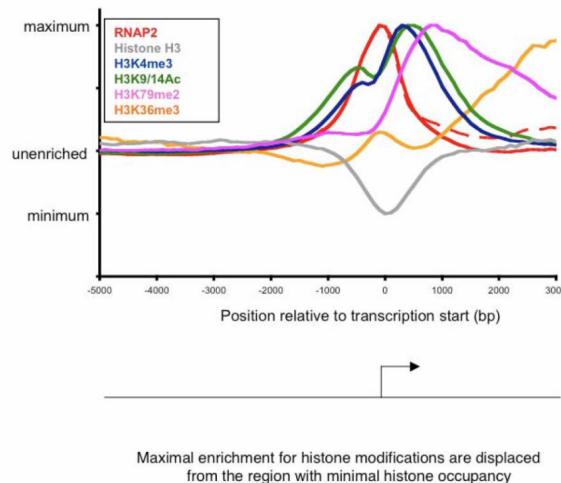


Figure 1.7: Guenther et al., Cell 2007

1.3.3.1 RT-qPCR based detection of 5' RNA transcripts of inactive genes

1. Assay components and DNA template: forward primer + probe + reverse primer
2. Denatured template and annealing assay components:
3. Polymerization and signal generation

Genes that do not produce full length mRNAs can experience transcription initiation. The majority of all genes contain H3K4me3-modified nucleosomes in both ES and differentiated cells. Differential H3K4 methylation is found in 25% of genes. These genes have cell type-specific expression pattern and cell type-specific function.

Summary: Results suggest that most protein-coding genes in human cells, including most genes thought to be transcriptionally inactive, experience the hallmarks of transcription initiation. H3K4me3 and H3K9-14Ac modifications, together with RNA Pol II, occupy the promoters of approximately 79% of protein-coding genes in ES and differentiated cells, but only about half of these produce detectable transcripts. The co-occupancy of Pol II with H3K4me3- and H3K9,14Ac-modified nucleosomes at genes without detectable levels of transcription suggests that a large fraction of human genes experience transcription initiation without transcript completion.

Protein-coding genes can fall into three groups of regulatory behaviour:

- *Productive*: the actively transcribed genes are occupied by nucleosomes with histone modifications that are hallmarks of both initiation (H3K4me3 and H3K9-14Ac) and elongation (H3K36me3 and H3K79me2)
- *Non-productive*: experiences transcription initiation without evidence of transcript elongation or accumulation.
- *No initiation*: genes that are excluded from experiencing transcription initiation, where mechanisms that prevent transcription initiation must predominate.

1.3. NUCLEOSOME ORGANIZATION

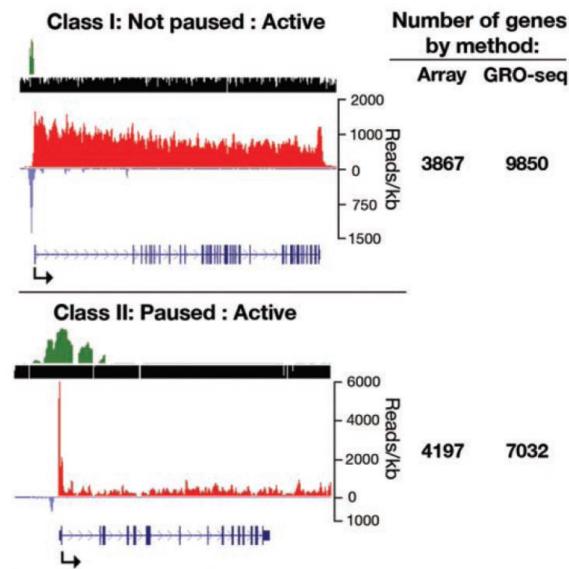


Figure 1.8: Core et al, 2008

1.3.3.2 Global run-on sequencing (GRO-seq)

Global run-on sequencing (GRO-Seq) enables to map the presence of elongating RNA pol in the chromatin context by detection of the nascent transcripts. The *Sarkosyl treatment* permeabilizes nuclear membranes, prevents RNA Pol from binding DNA (if not already bound), disassembles chromatin and dissociates pausing factors.

Results: only few paused genes are transcriptionally inactive. **Pausing** can be seen as a mechanism for tuning expression from active genes rather than a means of gene inactivation. The paused Pol II remains stably associated with the nascent RNA and is fully capable of resuming elongation; however, further signals are needed to elicit the transition to a productive elongation complex.

Pausing index: ratio of Pol II density on promoter to gene body.

The **CTD** undergoes dynamic changes in phosphorylation and dephosphorylation during transcription. The CTD is a long unstructured domain consisting of $Y_1S_2P_3T_4S_5P_6S_7$ tandem repeats. **Cyclin-dependent kinases** (CDKs) are serine/threonine specific kinases; cell cycle-associated CDKs regulate cell cycle progression. Transcription-associated CDKs are key regulators of gene expression. PTMs of the PolII CTD coordinate transcription and RNA processing

The differential phosphorylation of RNA Pol II CTD during transcription is reported in Figure 1.9.

1.3.3.3 Transcription pausing mechanism

NELF is recruited on PolII by **DSIF**, which contributes to pausing by forming 2 nucleic acid clamps. CDK9 with cyclin T, referred to as **P-TEFb** complex, enables pause-release leading to transcription elongation (Figure 1.10).

1.3. NUCLEOSOME ORGANIZATION

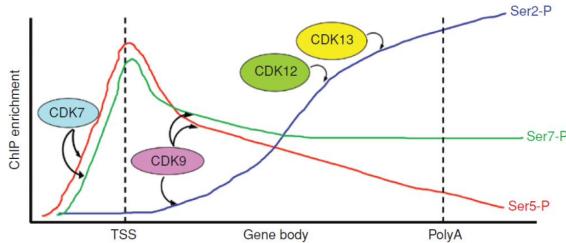


Figure 1.9: Differential phosphorylation of RNA Pol II CTD during transcription.

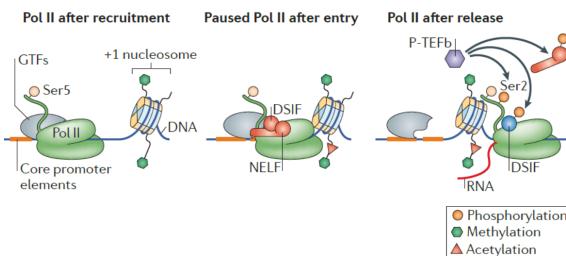


Figure 1.10: Jonkers & Lis, Nature Rev Mol Cell Biol, 2015

Pol II distribution profiles surrounding all promoters reveal a general decrease in promoter occupancy upon NELF depletion. Knockdown of NELF and consequent reduced Pol II pausing decreases gene expression, because nucleosomes occupy the promoter regions and thus block the expression of these genes by preventing TF and GTF binding.

Promoter-proximal pausing represents an additional layer of regulation to accommodate increased demands for precise and rapid gene regulation during organism development and responses to stress. It might be beneficial to maintain highly regulated promoters poised in an open chromatin state, to prevent their incorporation into the more inaccessible, condensed heterochromatin that exists in metazoans. Housekeeping genes lack pronounced pol II pausing.

Main hypotheses for the function of pol II pausing:

- establishing permissive chromatin: paused Pol II helps to maintain the nucleosome-deprived structure by blocking nucleosome assembly over promoter sequences. Pausing would thus keep the promoter region accessible for activator and transcription factor binding.
- rapid or synchronous activation: at a gene with paused Pol II, gene activation could proceed simply through recruitment of positive transcription elongation factor b (P-TEFb), thereby triggering the rapid release of paused Pol II into productive elongation without the need of the PIC assembly.
- integration of signals: pausing represents a separate step in the transcription cycle for factors to act and allows for combinatorial control between transcription factors that recruit the transcription machinery (TF1) and those that trigger pause release (TF2), where both would be necessary for gene activation.
- checkpoint in early elongation: on the left, arrows depict interactions between the capping enzyme complex (CEC) and DSIF–NELF as well as the Ser5 phosphorylation of the carboxy-terminal heptapeptide repeat domain (CTD) of Pol II, which is thought to stimulate capping

1.4. TRANSCRIPTION ELONGATION

activity. The hat represents the 5 RNA cap. In the centre, P-TEFb-dependent phosphorylation events release paused Pol II and create a platform for binding of RNA-processing factors (RPFs) on the Ser2-phosphorylated CTD of Pol II, as shown on the right.

7SK snRNP represses transcription by sequestering and inhibiting P-TEFb. Super elongation complex (**SEC**) and Bromodomain containing protein 4 (**BRD4**) complexes represent active forms of P-TEFb that can promote the release of Pol II from pausing.

BRD4 competes for binding to P-TEFb with the P-TEFb inhibitory complex HEXIM–7SK, and it is recruited to TSSs by multiple means, including histone acetylation. The SEC can interact with a subset of co-activators such as Mediator, polymerase-associated factor 1 (PAF1) and Integrator, the latter of which is a complex that interacts with the CTD of Pol II. The SEC and BRD4 predominantly mediate the recruitment P-TEFb, but their regulatory importance and composition with regard to paused Pol II release seems to vary across different genes, cell types and stimuli. Ser2 phosphorylated CTD promotes the recruitment of other transcription elongation factors, including SPT6.

A large repertoire of transcription activators crucially involved in cancer, including MYC and NF- κ B, target P-TEFb, DSIF or NELF to gene promoters.

1.4 Transcription elongation

During the elongation phase, RNA pol II faces nucleosomes. In order to proceed with elongation, RNA pol II needs to overcome the issue: *histone turnover* is predominantly observed at transcription start sites.

1.4.1 Structural basis of the nucleosome transition during RNA pol II passage

The aim of the research paper was to reconstitute RNA pol II and a nucleosome separately in vitro and analyze their interaction. The DNA stretch is required to be around 153 bp to achieve histone wrapping for obtaining a nucleosome. Keep in mind that is not easy to recapitulate the PIC formation in vitro. The first step involves unwinding and exposing DNA. For bypassing this step, it is possible to use an already “melted” initiation site ready for transcription. Once RNA pol II and **TFIIS** elongation factor are added, transcription can start.

1.4.1.1 Nucleosome structure

A single base pair is centred on the nucleosome *dyad*, which defines the pseudo 2-fold symmetry axis of the nucleosome. The major groups in the DNA stretch facing the dyad are *superhelical location* (SHL) -7 and -1. In nucleosomes, SHLs are formed at regular intervals (10bp) from the dyad interacting with histone K and R residues.

Kujirai and Kurumizaka, *Curr Op. Struct. Biol.*, 2020

1.4. TRANSCRIPTION ELONGATION

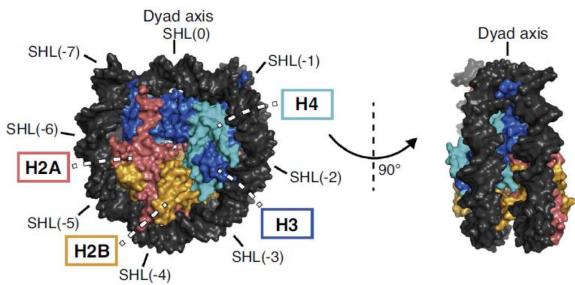


Figure 1.11: Kujirai and Kurumizaka, *Curr Op. Struct. Biol.*, 2020



Figure 1.12: Screenshot 2022-10-05 at 22.57.06.png

1.4.1.2 Cryo-EM

Cryo-Electron Microscopy is technique applied for molecule structure determination, which allows to reconstruct structures almost at the atomic level. We freeze samples in vitro (-80°C , -90°C) with liquid ethane. The sample is hit with a photon beam instead of light from different orientations, obtaining a pool of 2D images. Finally, a 3D image is reconstructed through an algorithm, providing a resolution in Armstrongs.

Most of the times, RNA Pol II was associated with the nucleosome at position SHL(-5) or SHL(-1), as well as SHL(-6) and SHL(-2). Other sites were quickly encountered and passed by Pol II, whereas the aforementioned ones required a bit more time. Looking at the electrophoresis results without the addition of TFIIS, RNA pol II can only reach SHL(-5), while when we add it also SHL(-1) is observed.

RNA Pol II elongates RNA up to SHL(-5) in the absence of the elongation factor and until reaching SHL(-1) in the presence of TFIIS. Pausing at SHL (-5) and (-1) represent rate limiting steps of nucleosomal transcription in vitro.

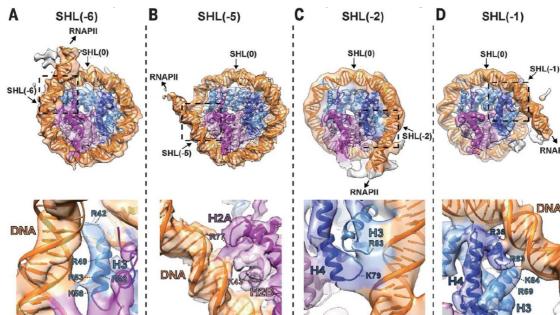


Figure 1.13: Kujirai et al., *Science*, 2018

1.4. TRANSCRIPTION ELONGATION

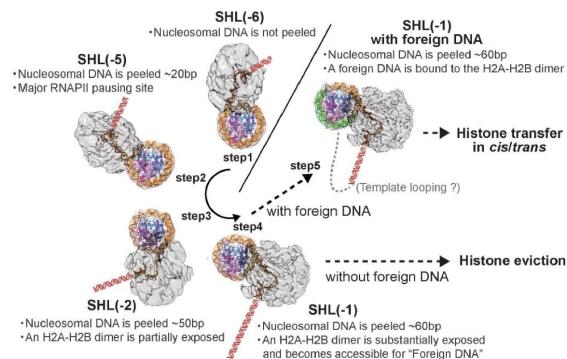


Figure 1.14: Screenshot 2022-10-05 at 23.02.04.png

Kujirai et al., *Science*, 2018

SHL(-6) is at the entry site, DNA has not been “peeled” yet from the nucleosome. At these sites H3 and H4 interact with DNA, we can appreciate the presence of modifications.

The dimer H2A and H2B is completely exposed after the “peeling” of the first DNA turn. This results in the uncovering of a positively charged region, which can be easily accessed by other compounds and rewrapped by “foreign DNA” i.e. not the DNA which was there previously.

1.4.1.3 Summary of the main results

- Nucleosomes can inhibit initiation and elongation of transcription.
- Nucleosome free regions are present within core promoters.
- Transcription pausing occurs at the first well-positioned nucleosome (+1).
- RNA polII transiently pauses when it transcribes nucleosomal DNA.
- Elongation factors are required for productive transcription.
- The transcription efficiency is modulated by histone post-translational modifications, histone variants and histone exchange which may affect histone-DNA contacts in the nucleosome re-leaving the major RNAPII pausing sites.

1.4.1.4 Template looping

DNA upstream of Pol II starts wrapping around the upstream nucleosome, pushing pol II to continue. This process does not require disassembly.

Kujirai and Kurumizaka., *Curr Op. Struct. Biol.*, 2020 Template looping results in histone transfer from ahead of the transcribing RNAPol II to behind

It is not only required to have TFIIS, other factors are involved: Spt4/5 and Paf1C (elongation factors), **FACT** and **Spt6** (histone chaperones) with acidic patches balancing the positively charged histones exposed in unwrapping.

FACT(facilitate chromatin transcription) interacts with the H2A-H2B dimer and we observe the formation of the *hexasome* (only 8 subunits) in the process of histone turnover. FACT activity

1.4. TRANSCRIPTION ELONGATION

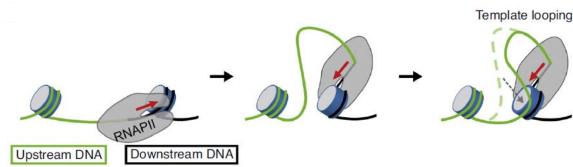


Figure 1.15: Kujirai and Kurumizaka., *Curr Op. Struct. Biol.*, 2020 Template looping results in histone transfer from ahead of the transcribing RNAPol II to behind

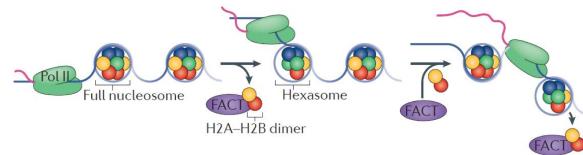


Figure 1.16: Lai and Pugh, *Nature Review Mol Cell Biol.* 2017

is also influenced by H2A-H2B dimers ubiquitylation. Nap1 (HC) stabilizes hexasomes favouring transcription elongation.

Lai and Pugh, *Nature Review Mol Cell Biol.* 2017

In a similar fashion, H3.3 is bound by HIRA, TF for “recycling”, regulated by acetylation.

Paf1C elongation factor interacts directly with Pol II, in particular with s5 phosphorylation. It also interacts with DSIF and FACT. When pol II is elongating, it promotes the phosphorylation of NELF and DSIF. Ctr9 is a subunit of Paf.

Van Oss *et al.*, *TRENDS Biochem Sci.* 2017

1.4.1.5 Structural basis of nucleosome disassembly and reassembly by RNA P II elongation complex with FACT - Ehara et al.

Recent article: once H2A-H2B are partially exposed, FACT starts interacting with the nucleosome. When the full exposure of the nucleosome is reached, FACT promotes a transit from downstream to upstream, assisting the template looping procedure.

Chromatin remodeling factors also participate in transcription elongation:

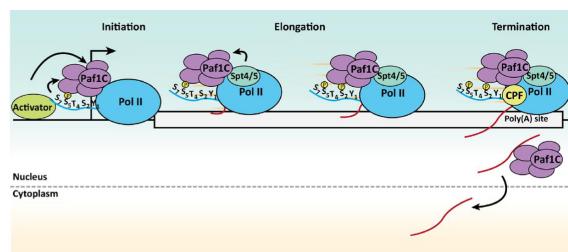


Figure 1.17: Van Oss *et al.*, *TRENDS Biochem Sci.* 2017

1.5. TRANSCRIPTION TERMINATION

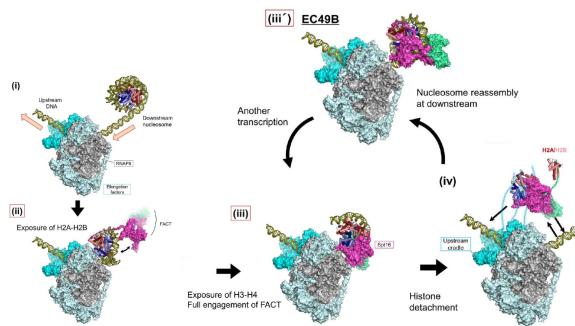


Figure 1.18: Screenshot 2022-10-10 at 10.35.12.png

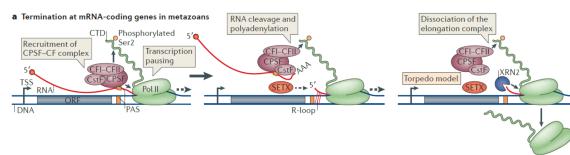


Figure 1.19: Porrua & Libri, Nature Review Mol Cell Biol, 2015

- **RSC:** arrangement of NDR between +1 and -1 nucleosomes
- **CHD1** and **ISWI:** arrangement of nucleosome spacing and separation of closely packed nucleosomes
- **Fun30:** facilitation of nucleosome disassembly

1.5 Transcription termination

Transcription termination occurs when the elongation complex is disassembled, and both RNA Pol II and the RNA are released. In our cells a signal on DNA is required, a *polyA site* followed by U rich/G rich regions. Once polyA is transcribed, we observe the SPSF complex recruitment, which recognizes the AAA site and proceeds to RNA cleavage and polyadenylation. Meanwhile, RNA pol II will continue transcription (slowly, probably due to allosteric changes in pol II structure bound to the complex) of mRNA which will be degraded due to the exposed 5' end. **Torpedo model:** XRN2 degrades this RNA and leads to the dissociation of pol II from DNA.

Porrua & Libri, Nature Review Mol Cell Biol, 2015

1.5.1 Promoter reporter assay

Since transcription at core promoters can be induced by distal elements, we can perform a promoter reporter assay to investigate which regions actively regulate transcription.

In order to perform the assay, we need to clone the upstream region of a known gene e.g. GFB. *Cloning vector:* enhancer + CNV (promoter and enhancer) + multiple cloning sites (**MCS**) + GFB (coding sequence + polyA site) + OriC + amp.

1.5. TRANSCRIPTION TERMINATION

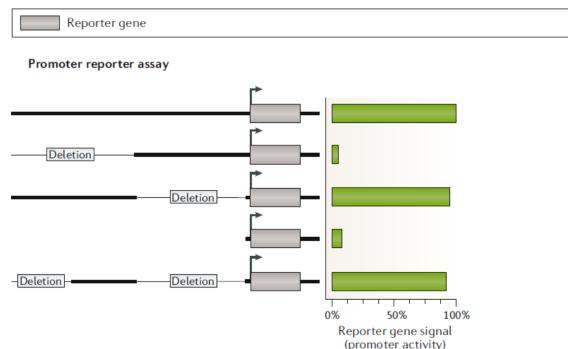


Figure 1.20: Anderson & Sandelin, *Nature Review Gen.* 2019

1. Linearize a plasmid by cutting it with restriction enzyme (HIND) and run it on the gel for extraction
2. Ligation
3. Transformation through heat shock
4. Integration screening (check amp resistance)

After cloning, we transfet the compound in cells. GFB will be barely detectable. We could start again with a longer sequence and clone it in from of GFB, but it would still be undetectable. It is required to employ a longer gene for observing signal. By playing with deletions, we understand that we only need small regions: 100 bp at TSS and a proximal enhancer region.

Anderson & Sandelin, *Nature Review Gen.* 2019

The enhancer can be hundreds of kilobases away from the promoter, could be upstream or downstream, even in introns. In order to check if a gene is an enhancer we need to place it around a reporter gene with different combinations.

- *Core promoter*: sequence that enables to initiate transcription (immediate vicinity of a TSS) by docking the pre-initiation complex (PIC)
- *Promoter*: refers to a sequence which can autonomously drive high levels of productive transcription (proximal enhancer + core promoters)
- *Enhancer*: distal element. The enhancer can be transcribed, but the mRNA is not stable and not codifying for protein.

1.5.2 Promoter types

Not all promoters have a TATA box, we have 3 different kinds of promoters:

1. **Type I ('adult')**: TATA box and sharp TSS. No CpG, nucleosomes are not well positioned → minority of promoters. Tissue-specific expression.
2. **Type II ('ubiquitous')**: no TATA box and CpG islands. Broad TSS, ordered nucleosome configuration.

1.5. TRANSCRIPTION TERMINATION

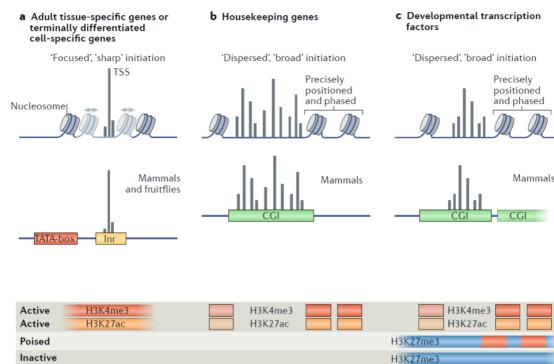


Figure 1.21: Lenhard *et al.*, Nature Rev Cancer 2012

3. **Type III ('developmentally regulated')**: featured by polycomb regulation, large CpG islands extending into the body of gene.

Lenhard *et al.*, Nature Rev Cancer 2012

1.5.3 Transcription factor binding

Most of the times, TFs require co-factors and direct binding; sometimes the *effector domain* can be either a ligand-binding domain or PPI domain.

TFs cooperate and synergize to gain access to their binding sites. For instance, factor A can bind alone to the unwinded first part, then B binds. In general we can observe:

- passive cooperativity
- TF-TF interactions
- enhancementosome model

Most TF binding occurs in clusters (<1kb) on nucleosome depleted regions. Regulatory networks are usually created with the same TF regulating different genes depending on other TF and cofactors availability.

Pioneer TFs are able to access their own binding sites independently of nucleosome/chromatin binding. For instance, **FoxA** HTH domain is really similar to H1 linker histone. HTH binds DNA minor groove: it has a trans activator domain and it is able to bind core histone. Once pioneer TF bind, they are able to pull DNA from the nucleosome creating a more accessible stretch. Examples:

- SOX2 exposes DNA for OCT4 → *sequential binding mechanism* for recruiting multiple TFs, rather than simultaneous
- FoxA1 enables Estrogen Receptor alpha recruitment

Other mechanisms we can observe are the recruitment of acetylases e.g. p300 to enhance nucleosome destabilization.

1.5. TRANSCRIPTION TERMINATION

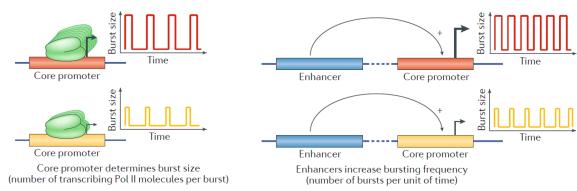


Figure 1.22: Haberle & Stark, *NatRevMolCelBiol.* 2018

Transcription is dynamically regulated but not constantly occurring. The core promoter will accommodate a number of pol II, which will then transcribe (burst). Different core promoters can have a different burst size. Enhancers do not change this size (which is given by promoter structure), but increase the number of bursts per time.

Chapter 2

Transcription regulation and its deregulation in cancer

Enhancer regulation is controlled by specific transcription factors with a DBD. Most of the times, TFs interact with PIC through co-factors. In particular, the Mediator is a key co-activator complex of RNA pol II transcription regulation.

2.1 Mediator complex

The Mediator complex is a multi subunit complex, composed of almost 30 subunits. CDK8 is the only module with enzymatic activity.

The Mediator complex can interact with hundreds of TFs and integrate their function, allowing cooperation and regulation of multiple sites in enhancer regions. Examples: SOX,c-Myc,...

Since the Mediator is a multi-subunit structure, a conformational change in one subunit propagates to the whole complex.

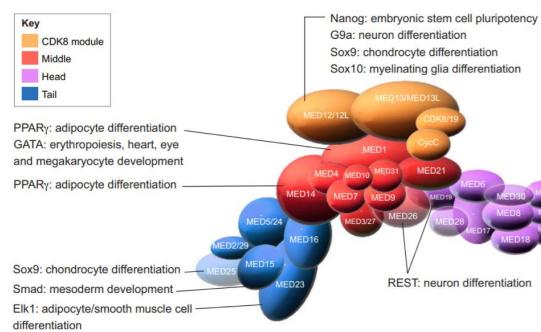


Figure 2.1: The Mediator acts as a hub integrating signals from multiple pathways

2.1. MEDIATOR COMPLEX

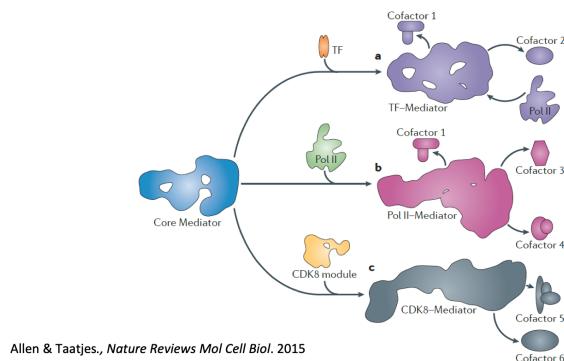


Figure 2.2: Allen and Taatjes , *Nature Review Mol Cell Biol.* 2015

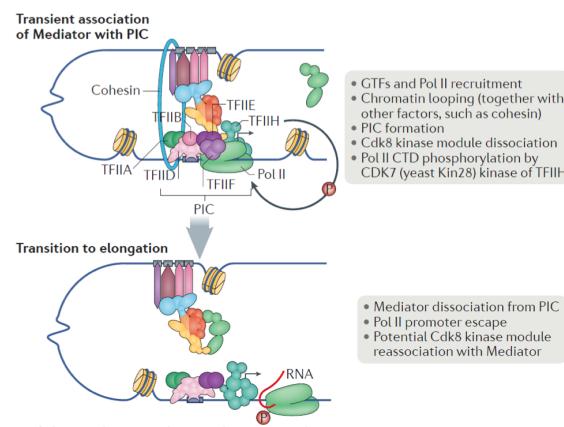


Figure 2.3: Soutourina, *Nature Review Mol Cell Biol.* 2018

The interaction with Pol II prevents the catalytic interaction: the core mediator has mutually exclusive interactions.

The Mediator functions as a bridge between the TFs bound to the enhancers and the PIC at core promoters; DNA looping involves *cohesins* (circle DNA) [Figure 2.3]. The interaction with enhancers occurs first and is more stable than the core promoter one.

The Mediator stimulates transcription at several levels: assembly of the PIC, chromatin organization, transcription elongation, nucleosome displacement at TATA-box promoters. After Pol II promoter escapes, the Mediator interacts with Cdk8 to activate pause release. The Mediator presence at TATA regions inversely correlates with nucleosome occupancy : indeed, it is able to bind acetylated histones and interact with chromatin remodelling complexes contributing in histone eviction.

The Mediator has been defined as **GTF** (general TF), however it preferentially associates with enhancers and its occupancy at core promoters is generally low and transient.

Once the Mediator is interacting with the PIC, Cdk8 is displaced; afterwards, the Mediator is able to recruit it again to activate Cdk9, allowing the pausing mechanism (Figure 2.4).

The multi-subunit and modular structure of the Mediator enables to respond to multiple signalling

2.2. WNT SIGNALLING PATHWAY

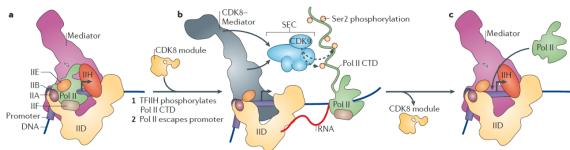


Figure 2.4: Allen and Taatjes, *Nature Reviews Mol Cell Biol.* 2015

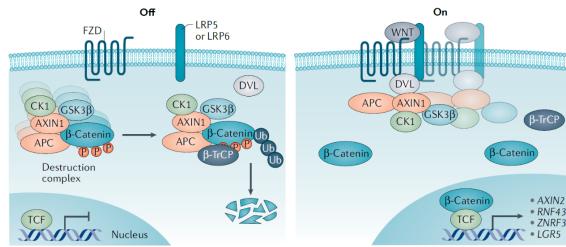


Figure 2.5: Bugter et al., *Nature Review Cancer*, 2021

cascades activating transcription. Different subunits can be engaged in PPIs; we are interested in many of them, in particular oncogenes usually interact with the mediator.

2.2 Wnt signalling pathway

Wnt signalling pathway is a relevant pathway for morphogenesis and embryonic development, as well as adult tissue homeostasis. Its activation is associated with cell proliferation and cell renewal.

The pathway occurs at a plasma membrane, where ~20 ligands interact among themselves and with a core receptor (either LRP5 or LRP6).

- Off: AXIN1 forms a *destruction complex* with APC, CK1, GSK, which recruits **β -catenin**, a relevant co-activator. The activation is performed through phosphorylation on Ser and Thr residues, which will then recruit btrCP ubiquitinylase and lead to **β-catenin degradation**.
- On: when Wnt is present, the complex is sequestered and therefore **β-catenin** is not phosphorylated and ubiquitinylylated, it remains present in the cell. Among target genes, AXIN2 promotes the recruitment and formation of destruction complex, others are plasma membrane proteins which promote *endosome mediated degradation* of WNT.

The Wnt pathway is schematized in Figure 2.5.

Wnt is synthesized in the endoplasmic reticulum, lipid modified and released from cells through EVs (or can remain attached to membrane of secreting neighbouring cells). It can spread and reach different sites.

Excessive **β -catenin** activity in cancer can occur due to:

- inactivating mutations in members of the destruction complex

2.2. WNT SIGNALLING PATHWAY

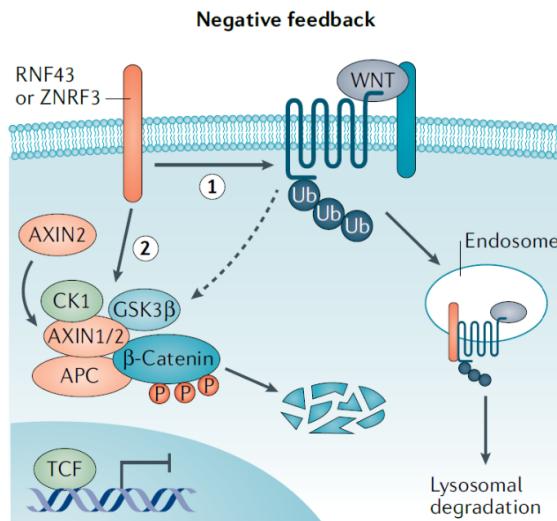


Figure 2.6: Bugter et al. , *Nature Review Cancer*, 2021

- negative feedback pathway (RNF43 or ZNRF3) and activating mutations β -catenin [Figure 2.6]

Genetic alterations in the WNT pathway occur in most cancer types with some components of the pathway showing tissue specificity. All WNT mutations converge in enhancing β -catenin activity, although with different extent, and cooperate with other driver mutations to promote tumorigenesis. Mutation are considered early (APC) or late (RNF43) events and associate with different tumor subtypes. APC mutant tumor leads to many clinical features: depending on the function and position in the pathway of the gene, the downstream effect can be very different.

In absence of β -catenin (Wnt pathway off), TLE and HDAC bind TCF and stop transcription. β -catenin subcellular translocation is presumed to function thanks to the binding of adaptor proteins. β -catenin needs to interact with different TFs, among which TCF family members are the most common and expressed in different forms. They all recognize the same binding site. In particular, different isoforms of TCF7 and LEF1 are highly expressed in colorectal cancer.

2.2.1 β -catenin structure

β -catenin structure acts as a binding platform for numerous interactions. It is composed of 781 AA residues in humans and divided in:

- central highly repetitive sequence with *Armadillo sequence* important for nuclear activity
- Helix-C conserved structure close to CTD
- C-terminal transcription activator domain (CTTA)

Key interactors: E-cadherin and A-catenin \rightarrow structural function, focal adhesion. Decrease in E-cadherin is found in EMT. E-cadherin cad is stabilized by β -catenin, therefore with low levels of β -catenin we observe an E-cadherin decrease in EMT.

2.2. WNT SIGNALLING PATHWAY

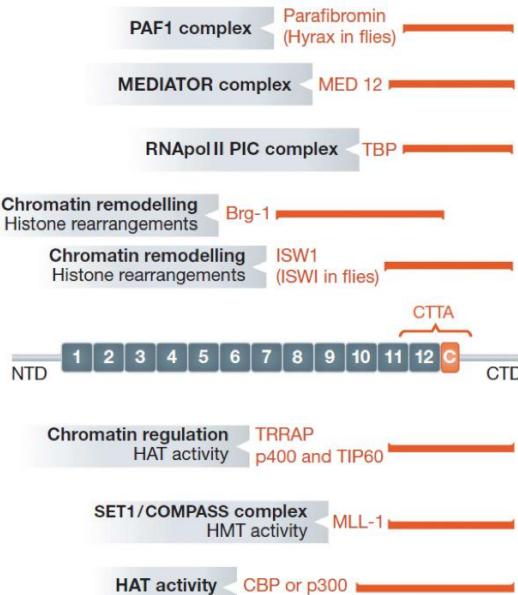


Figure 2.7: Valenta *et al.*, EMBO J. 2012

The function of β -catenin, both structural and signaling, is regulated by post-translational modifications.

RNA pol II is also loaded on inactive genes; to foster elongation we require β -catenin, which can recruit PAF1 complex. The majority interact with the CTTA domain, which serves as a binding platform or trans-activation domain (Figure 2.7, 2.8).

2.2.2 The canonical Wnt β -catenin pathway is crucially involved in colorectal cancer

The **APC gene** was first identified by being mutated in a hereditary colon cancer syndrome termed familiar adenomatous polyposis. Most cases of sporadic colorectal cancer result from loss of both APC alleles. Loss of APC function leads to the inappropriate stabilization of β -catenin and the formation of constitutive complexes between β -catenin and the intestinal TCF family member TCF7l2/TCF4. Patients with hereditary Axin2 mutations display a predisposition to colon cancer. Aberrant activation of the canonical Wnt/ β -catenin pathway occurs in almost all colorectal cancers, contributing to their growth, invasion and survival.

Luciferase reporter genes: firefly and renilla Luc, of the two only one has a responsive site to TCF. It is a control for the efficiency of transfection.

Let's see which genes are involved in the regulation of β -catenin through RNAi in DLD1 cells: 34 genes are necessary for β -catenin activity and 166 candidate genes are necessary for HCT116 cells proliferation. By merging the two groups, we can identify a set of 9 potential regulators of colon cancer proliferation and β -catenin activity: CDK8, CSK1G3, CSNK1E, DKC1, MAP3K14, MLLT7, PLK4, TAOK1 and ZAK.

2.2. WNT SIGNALLING PATHWAY

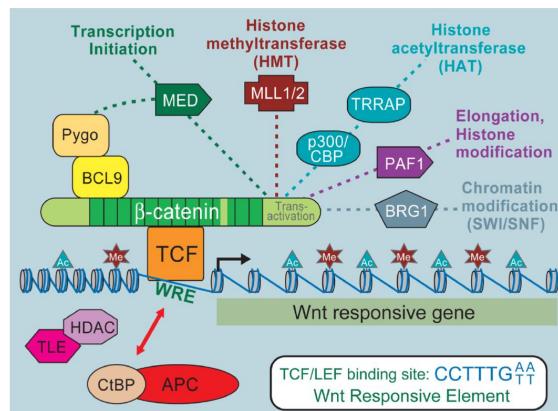


Figure 2.8: McDonald *et al.*, *Dev Cell*, 2009

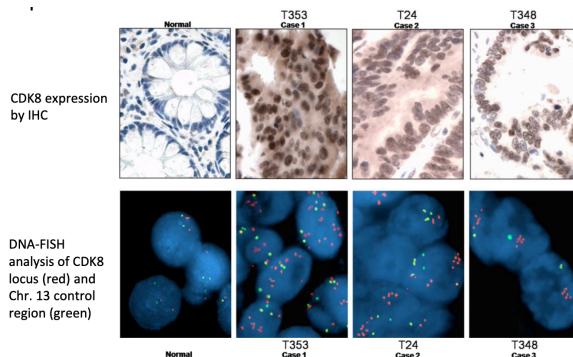


Figure 2.9: Firestein *et al.*, *Nature* 2008

Genome wide analysis of chromosome copy number alterations (CNA) in human CRC biopsies revealed amplification of chr. 13 region including CDK8. Immunohistochemical analysis of CDK8 expression in the same 50 specimens in Figure 2.9 revealed elevated protein levels in 26% colon cancer samples, including those that showed CDK8 CN gain [the brownish colour in IHC indicates that CDK8 expression increases].

These observations indicate that CDK8 is amplified and overexpressed in a substantial fraction of colon cancers. In particular, CDK8 action is not a passenger effect, it is a *driving proliferation*. There is a huge difference from in vitro and in vivo tumor setting: in order to obtain more suitable experiments, we can look at the transforming potential (as not all proliferating cells are necessarily tumor cells). If we combine β -catenin in catalytically dead CDK8 we drop the effect, but it is not completely abolished - as β -catenin can bypass the need engaging interactions with others.

Main conclusions of the article

- CDK8 acts as an oncogene in a substantial fraction of colorectal cancers
- CDK8 kinase activity is essential to regulate β -catenin dependent transcription and transformation

2.3. CBP/P300

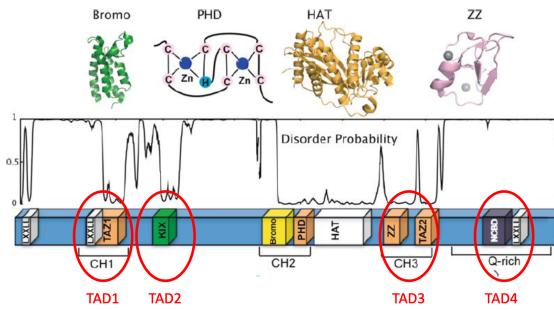


Figure 2.10: Wang *et al.*, *Cell Mol Life Sci.* 2013

- CDK8 acts in part by co-activating β -catenin driven transcription in colon cancer with high CDK8 expression and β -catenin activity

Therapeutic interventions that target CDK8 kinase activity may be of clinical value in colorectal cancer

Almost all Mediator subunits have been found to be mutated/deregulated in a huge list of cancers e.g. MED12 subunit in prostate cancer.

2.3 CBP/p300

CREB-binding protein (CBP) and p300 is a heavily regulated TF. Differently from the Mediator, it is a monomer and has several trans-activation domain, which enable it to interact with TF and PIC. It has histone acetyltransferase activity (HAT) or lysine acetyltransferase activity (KAT). Chromatin will be modified by increasing accessibility and also TF regulation.

2.3.1 CBP/p300 structure

CBP/p300 domains are reported in Figure 2.10:

- 4 transactivation domains
- histidine rich or glutamine rich regions
- zinc fingers domain for PPIs
- bromodomain and PHD interact with histones
- lysine acetyltransferase domain (HAT)

There are no intrinsically disordered regions, they fold after interacting with intrinsically disordered proteins.

Note that among CBP/p300 binding partners we find both oncosuppressor genes and oncogenes, as well as co-factors.

2.3. CBP/P300

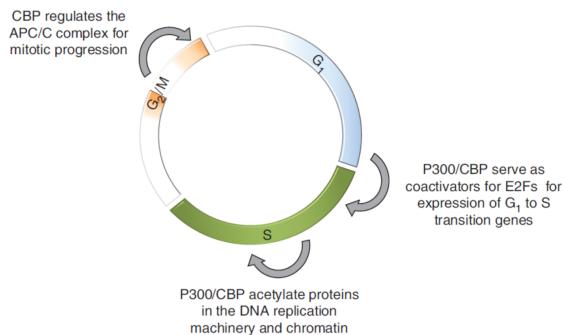


Figure 2.11: CBP/p300 regulate cell cycle at multiple time points

TF interactions are mediated by TADs. There is a combined interaction among multiple TFs and CBP, enabling the integration of the function of several TFs. The interaction with enhancers does not necessarily correlate with transcription activation; only when the appropriate number and identity is recruited on the enhancer, it becomes functional.

CBP/p300 levels in cells are limiting and different genomic sites or pathways compete for their activity.

In the case of p300, we observe the *co-binding mechanism* with common cofactors or common complexes (**transcriptional synergy**).

Co-activators facilitate transcription by numerous mechanisms:

- Adaptors function: bridging TFs to the PIC, recruitment of polII to core promoters, recruitment of chromatin remodelling and modifying enzymes to core promoters
- Scaffolding function: facilitating protein-protein and protein-DNA interactions, promoting/stabilizing PIC formation
- Enzymatic activities: Kinase (mediator), KAT (CBP/p300) targeting histones, TFs and non-TF proteins (acetylation of P-TEFb enhances its activity).

Their activity can vary from gene to gene depending on TF involved and regulated by PTMs and chromatin accessibility.

2.3.2 CBP/p300 in cancer

In particular, CBP/p300 can regulate pro-proliferative pathways, e.g., cell cycle regulation (Figure 2.11).

Indeed, CBP and p300 gene are frequently mutated in cancer - mutually exclusive. Most are missense mutations (single nucleotide), some deletions and insertions. Mutation hotspots are found at the catalytic domain (KAT). We always need to consider the context of genetic instability characterizing cancer: depending on which signal is activated/repressed, a particular mutated form can either exert proliferative or oncosuppressor activity [often loss of heterozygosity].

2.4. TSSa-RNAS TRANSCRIPTION

The many functions of CBP/p300 can be differentially exploited in cancer depending on the context, leading the CBP/p300 proteins to act as oncosuppressors or oncogenes. Depending on cellular identity, dysregulated pathways and transcription rewiring, cancer cells may select for the CBP/p300 alterations most advantageous for survival and growth.

2.4 TSSa-RNAs transcription

Transcription by RNA Pol II is often thought to occur unidirectionally, however we also observe an enrichment of transcripts in the opposite direction with respect to forward transcription. TSSa-RNAs (TSS antisense) surround promoters in nonrandom and divergent orientations. Sense TSSa-RNAs map downstream of the associated promoter, overlapping genic transcripts and peaking in abundance between +0 and +50 nt downstream of the TSS. Antisense TSSa-RNAs peak between nucleotides -100 and -300 and can range from 20 to 90 nt (detected by Northern blot). The mere fact of having high abundance of TF can recruit pol II, which can act *spuriously*: it not only transcribes following the functional direction, but also in the opposite direction (with different pre-initiation complexes).

It is presumed that these small transcripts are abortive transcription tentatives; stalling can occur if RNA is not properly processed e.g. no recruitment of cap proteins, so after degradation and when RNA pol II is displaced, a short sequence (which was protected by pol II) remains at the site. In general, there is a bias towards the directionality of these transcripts, confirming the previous hypothesis: highly expressed genes all have TSSa-RNA.

Although divergent transcription initiation is widespread and might have functions, productive elongation by RNAP II occurs primarily unidirectionally, downstream of TSSs. Promoter upstream transcripts (PROMPTs) are longer than TSS-a and are more stable, while PASRs overlap on the TSS.

Different families of noncoding RNAs are transcribed at promoters of protein coding genes:

- Transcription start-site-associated (**TSSa**) RNAs are short transcripts (18-24 up to 90bp), do not overlap TSS, sense or antisense of the mRNA
- Promoter associated short transcripts (**PASRs**): capped bidirectional short RNA (20-100nt long) that overlap TSS
- Promoter upstream transcripts (**PROMPTs**): capped bidirectional ncRNAs, map upstream of active promoters (-500 and -2000 from TSS)

Potential functions:

- Help modulating chromatin structure due to the Pol II activity and/or binding with chromatin i.e. help sustaining nucleosome free regions
- Noncoding RNAs as precursors for short ncRNA (especially longer ones)

[RNase are degraded by the exosome, which is a multi-subunit complex which recognizes uncut RNase.]

ENCODE (Encyclopedia of DNA Elements) has the objective to map all RNAs, which are divided into long and short and by cellular location - either cytoplasmatic or nuclear.

2.5. ENHancers

- 74.7% of human genome is transcribed (primary transcripts)
- short RNAs detected: ~160,000 transcripts (~40,000 exonic, ~55,000 intronic, ~45,000 intergenic, ~23,000 gene-intergene boundaries)
- long RNA transcripts detected: ~60,000 transcripts from ~19,000 protein coding genes (GENCODE v7); 73,325 transcripts mapping within intergenic regions (also enhancers) and antisense elements (long noncoding RNAs and possible new genes)

2.5 Enhancers

TSS identified within enhancer regions with transcription proceeding for several kilobases giving rise to **eRNAs**.

How can we distinguish enhancers? By analysing epigenetic marks, TF binding and especially co-factor binding (ChIP-Seq).

Poly(A)- mostly enriched, but also Poly(A), detectable by CAGE hence capped. Transcribed enhancers on average show a significantly different pattern of chromatin modification than non-transcribed ones e.g. H3K27ac,... All the markers are associated with transcription initiations.

In addition to the fact that enhancers need to regulate transcription to allow RNA production, transcription can occur bidirectionally and the enhancer itself can be transcribed.

- **poised enhancer:** TF bound to the enhancer LDTF (lineage-determinant TF), which is a pioneer TF. We can see some histone modifications for activation, but they are not sufficient to fully activate the promoter. Even pol II can or cannot be recruited, it is not yet acting on a target promoter, but it is active.
- **active enhancer:** Mediator recruitment, HAT activity, nucleosome remodelling complexes and transcription occurs on the enhancer.

Example: the signalling of NFKB TF starts in cytoplasm, it only is shuttled to nucleosome upon a specific signal. Since it is not pioneer TF, it can only bind to poised enhancer → cell-line dependent mechanism.

An active enhancer shows high levels for H3K4me (especially me3), Pol II and Ser5P, H3K27ac, while other elongation markers are less observed. Example of cell-specific regulation: TAL1 TF, oncogene expressed in endothelial cells and erythroid cells, we witness enrichment of H3k4me2 at enhancer sites. Active enhancers associate with chromatin looping.

2.5.1 Properties of eRNAs

- eRNAs are transcribed from putative enhancer regions characterized by high levels of H3K27ac and the lack of the repressive H3K27me3 mark
- These genomic regions are bound by TFs and associate with transcriptional co-regulators including Mediator, histone acetyltransferase CBP/p300
- eRNA-expressing enhancers are also enriched with PIC
- eRNAs are 50-2000 nt in length and generally exhibit shorter half-lives compared to mRNAs and lncRNAs. They are generally not spliced or polyadenylated

2.5. ENHancers

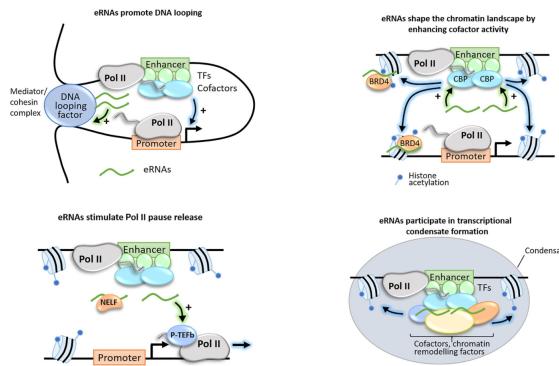


Figure 2.12: Studniarek *et al.*, *Trends in Genetics*, 2021

- Enhancer transcripts are preferentially enriched at enhancers engaged in chromatin looping with promoters of protein-coding genes and other enhancers, which is a feature correlated with enhancer activity

2.5.2 eRNAs proposed functions

The proposed functions of eRNAs are depicted in Figure 2.12.

Last function: Intrinsically disordered proteins → phase separation mechanism. We are enriching the density of enzymes in a particular region.

In particular, eRNA can promoter chromatin remodelling recruitment to reach appropriate configurations for chromatin at TSS.

2.5.3 Assessing whether the CRC 8q24 risk region acts as an enhancer

Most mutations in cancer are SNPs and mainly occur outside of coding regions. Driver mutations are commonly found at coding regions, but also outside of it. One way in which we can analyze non-coding region is assessing the frequency of mutation for a region. In 2007, two different studies found that some SNPs located in 8q24 were highly associated with colorectal cancer.

In particular, the rs6983267 SNP G allele (over T allele) on 8q24 associates with increased risk of colorectal cancer. Homozygosity for the risk G allele in cancer increases CRC risk 1.5 folds.

One way to study non-coding SNP is to understand whether it can be a regulatory region:

- ChIP-Seq: the risk locus with G shows an enrichment in RNAP II, methylation and p300
- Reporter assay: 1.5 kb long region including the risk allele can function as an enhancer. The G risk allele shows increase reporter activity than the WT allele for the risk allele. It is not a 10 fold difference, suggesting it is not a driver mutation.

Epigenetic and reporter assay data indicate that the CRC risk allele functions as an enhancer. The rs6983267 SNP is located with a **TCF** consensus binding site (A/T)(A/T)CAA(A/T)GG. This site

2.5. ENHancers

is not an exact consensus sequence, but if we have a G on the last position we achieve a higher affinity. By performing ChIP on TCF4, the peak for G allele is higher than T. There is no gene for hundreds of kbs, but if we move forward we find Myc. Since TCF4 is an effector of Wnt pathway and Myc is a beta-catenin target, the SNP can be placed in Myc enhancer region. Instead, no association between MYC mRNA expression and CRC risk allele was identified in either normal colon or tumor samples. However, 3C experiments show a physical interaction between the risk allele and MYC promoter was detected in CRC lines but not fibroblasts.

Conclusions:

- The CRC risk locus shows enhancer-like functions in CRC cell lines
- The CRC risk allele interacts with TCF4
- This locus physically interacts with the MYC promoter

Hypothesis: the rs6983267 risk allele may regulate MYC during tumorigenesis

2.5.4 The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling

Computational approach predicts that the rs6983267 would impact TCF4 binding site, since a change from T to G is predicted to considerably increase the affinity for TCF4 binding. They use a different approach to verify whether TCF4 is interacting with the risk region.

EMSA: in vitro incubation of recombinant proteins (oligonucleotide of interest with radioactive mark). Run electrophoresis: TCF4 with its own sequence, unrelated sequence and SNP alleles sequence. TCF4 can bind the risk allele G, much less affinity for T allele.

Activation of the WNT pathway through GSK3B inhibition indicates WNT responsive enhancer activity. Reporter assay to obtain a biological readout of the process: the entire site is required for the interaction and should not be further mutated. The risk G allele shows 1.5 fold increase in WNT responsitivity compared to the T allele.

RNA stability is tightly regulated, we expected that uncut RNAs are less stable than processed ones. The export in the cytoplasm of mRNAs is mediated by Transcription Exports (TERT); non-spliced/retained RNAs in the nucleus can be targeted for degradation. This is a side concept, but still linked inextricably to transcription and cell proliferation. By increasing transcription export also gene expression increases, and therefore cells are able to achieve a higher proliferation.

How can we study the impact of enhancers in cancer?

- chromatin marks
- reporter assay

Hypothesis: the rs6983267 risk allele may regulate MYC during tumorigenesis (Figure 2.13).

TCF4 and β -catenin were enriched with the rs6983267 risk allele.

Figure 2.14: the peak is placed right on rs6983267 risk allele in ChIP-seq (e). In Sanger seq. (f), we cannot discriminate if a G or T is present in the input, whereas in the pull down of TCF4 G has a clearly higher signal.

2.5. ENHancers

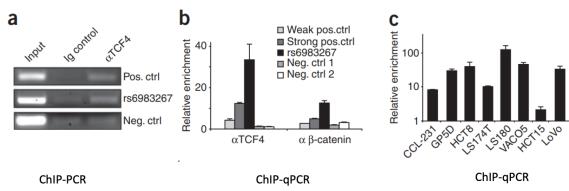


Figure 2.13: Tuupanen et al, Nat Genet, 2009

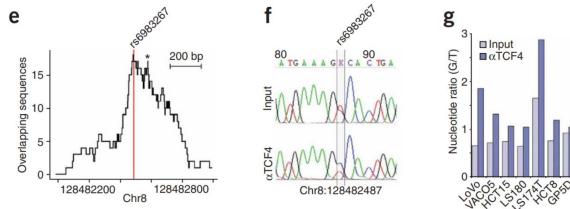


Figure 2.14: Tuupanen et al, Nat Genet, 2009

By analyzing TCF4 binding sites genome-wide, the risk allele was found to be the fourth most enriched site.

In order to asses whether MYC was indeed regulated by Wnt signalling pathway, the authors of the paper cloned a sequence including the risk allele upstream of a lacZ gene. The expression recapitulated an endogenous MYC (checked by in situ hybridisation).

Conclusions:

- A high risk allele SNP for CRC was identified resulting in a TF binding site within an enhancer
- The risk allele is more sensitive to TCF4 binding and WNT activation
- The risk allele regulates MYC expression (most likely) upon WNT signalling activation
- The presence of this allele may render cells hypersensitive to the WNT pathway resulting in deregulated (increases) expression of MYC with consequent pro-proliferative behaviour

2.5.5 CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability n colon cancer

Recently, it was observed that the conservation of the locus including the risk allele is particularly high in mammals (28 species). By looking at the Norther blot for the conserved sequence, there was a transcript of about 0.5 kb length which was called CCAT2, expressed in different cancer cell lines. Looking at RT-qPCR analysis from tumor biopsies vs normal: tumor expressed higher levels with respect to control. Interestingly, increased expression was observed for microsatellite regions, while the expression in antisense was almost undetectable. This suggests that the enhancer is transcribed unidirectionally, giving rise to CCAT2 transcript, which does not codify for proteins → long non-coding RNA.

CCAT2 expression is significantly higher in microsatellite stable (MSS) colon cancer tumours, compared to non-neoplastic mucosal samples and MSI-H tumours. In situ hybridization confirms increased expression of CCAT2 in tumor cells.

2.5. ENHancers

Retroviral-mediated expression of CCAT2 promotes tumor growth in HCT116 (MSI-H) cells. In particular, the endogenous is 1000 higher than HCT116 - control to check for overexpression.

When looking at the extent of proliferation, a lack of effect on 2D cell growth by CCAT2 was observed. **Xenograft study** (transduce with retroviral vector clones and inject in mice): CCAT2 increased subcutaneous tumor formation in a mouse xenograft model.

Modulation of the CCAT2 levels can affect the metastatic capacity of cells. In particular, a higher CCAT2 expression was detected in primary CRC tumors from patients with metastasis and higher CCAT2 RNA levels associate with shorter metastasis-free survival in breast cancer patients.

Is this related to MYC/enhancer interaction or only to CCAT2? The tumours showing increased MYC expression also had an increased CCAT2 expression. By depleting CCAT2 through siRNA, a decrease is also produced in MYC expression.

Gene known-down approaches:

- *small interfering RNAs*: tiny oligo RNA, transfected in cells, bp with target RNA and enable degradation
- *s h RNAs*: always short RNA bp with target, but expressed like microRNA in a plasmid. Pro: stabilized (cannot stabilize oligo), Cons: highjack DICER and other RNA enzymes
- *antisense oligonucleotide*: completely different chemistry, instead of phosphodiesteric bond between different bases, can have sulphur atoms making them resistant to endonucleases.

MYC is a downstream target of Wnt. Pull down for TCF a long non-coding RNA is bound to it (not *in vitro*). Instead of CHIP, extract RNA followed by PCR.

Conclusions:

1. An ~335-kb DNA loop brings the rs6983267 genomic region close to the MYC locus, and this physical association may contribute to the enhancer function of the SNP-containing region on MYC transcription.
2. The enhancer region is transcribed into a long noncoding RNA (CCAT2), and the SNP status affects CCAT2 expression.
3. The CCAT2 transcript up-regulates WNT activity and increases expression levels of WNT target genes (including MYC). This regulation by CCAT2, possibly through its physical interaction with TCF7L2, may lead to genomic instability and promote cell growth.
4. CCAT2 expression is regulated by transcriptional factors TCF7L2, indicating a positive feedback loop between CCAT2 and WNT signaling.

It is not the end of it! CCAT2 can interact with BOP1, regulating AURKB (Aurora B Kinase) and leading to unbalanced segregation of chromosomes during mitosis → chromosome instability. CCAT2 is also found in serum (can evade from cell through exosome), can be used as therapeutic target and diagnostic tool.

2.5. ENHancers

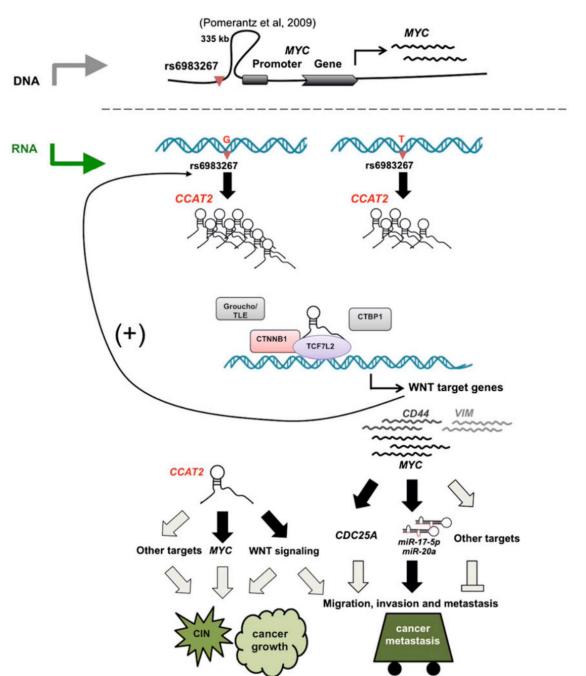


Figure 2.15: Ling et al, Cold Springer, 2022

Chapter 3

Chromosome domains and oncogene activation

3.1 Genome organization and transcription regulation

3.1.1 CTCF

An enhancer can work at very high distance with respect to the gene of interest; this means that enhancers have the possibility to regulate many different promoters, how can enhancer activity be directed to a specific target promoter?

Insulator sequences prevent the propagation of enhancers signals along the chromatin fiber through their enhancer blocking function mediated by proteins. Insulator genes are usually found at a region densely distributed with genes. They can act in long distance, do not need specific direction and the sequences are recognized by proteins. The main insulator binding protein is a factor called CCCTC-binding factor (CTCF).

CTCF has a peculiar DBD with 11 zinc fingers; each of them can recognize a short consensus motif. In principle, it could recognize huge regions and be very flexible in recognition capabilities, as combinatorial use of its 11 ZF allows binding to 50bp target sites with high sequence variations. CTCF regulates transcription both positively and negatively and can operate as a transcription factor. At first, it was recognized as a repressive TF: it has the ability to shape the DNA molecules, not merely binding on a site distorting it, and to recruit co-factors.

CTCF is the main *insulator-binding protein* in vertebrates. ChIP shows that CTCF binds thousands of sites in the human genome. We can find it mainly in intergenic regions and introns, as well as promoter regions. There is an extensive overlap of CTCF binding sites among cell lines (12000). The consensus sequence recognized by CTCF is the same among cell lines.

CTCF is enriched at the H3K27me3 domain boundaries (CTCF barrier sites). A larger genomic region was analyzed in different cell lines. Almost no overlap was found in the CTCF barrier sites between CD4T cells and HeLa cells, suggesting that CTCF barriers are cell-type-specific. In some cells the site is at the boundary, in others not.

3.1. GENOME ORGANIZATION AND TRANSCRIPTION REGULATION

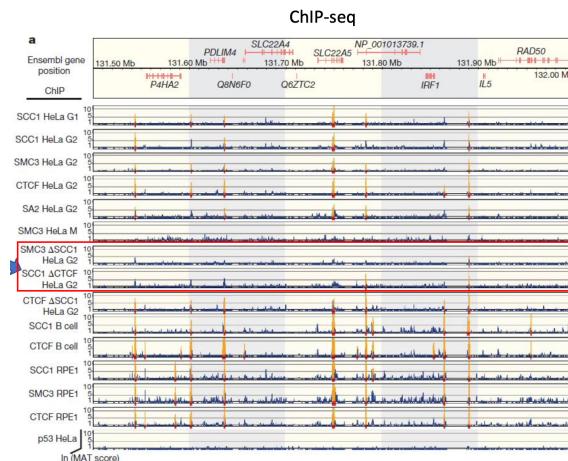


Figure 3.1: Yatskevich et al., Annu Rev Genet. 2019

We now know that heterochromatic marks spread up to insulator sequences, which can be recognized by protein and recruit directly or indirectly chromatin remodelling factors.

Main findings:

- CTCF can bind to the same locus in different cell types
- CTCF binding can separate active and repressed chromatin, functioning as barrier to heterochromatin spreading
- CTCF exerts its barrier ability in a cell type dependent manner

The function of CTCF appears to be regulated at least at two levels:

1. binding of CTCF to the target sites.
2. binding of interacting proteins, providing its functionality

3.1.2 Cohesin mediates transcriptional insulation by CCCTC-binding factor

CTCF was found to interact with cohesins (composed of SMC3, SMC1, SCC1 and SCC3 subunits), which are essential to maintain sister chromatids connected from S phase to metaphase. Cohesins are loaded on chromosomes during G1, but the function is exerted later on. For achieving anaphase, the ring is cut by securins - regulated by APC.

Cohesin is expressed in differentiated postmitotic cells. We expect that replicating cells express cohesin, as it is heavily required in cell cycle. Pulling down SMC3 in mouse brain we still observe a signal, it is probably linked to an additional function of the complex. During G2 we expect that cohesins are holding together sister chromatids, which instead are absent in G1; the binding site is the same, checked by ChIP-Seq → cohesin binds to the same sites independent on cohesion CTCF and cohesion share several binding sites (Figure 3.1).

3.1. GENOME ORGANIZATION AND TRANSCRIPTION REGULATION

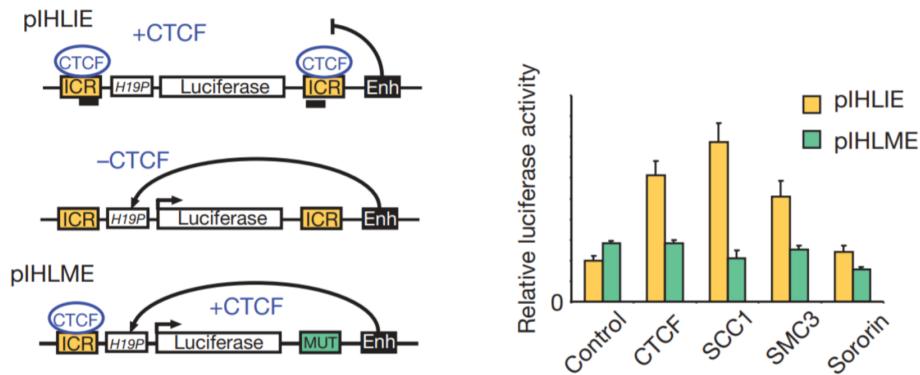


Figure 3.2: Wendt et al., *Nature*, 2008

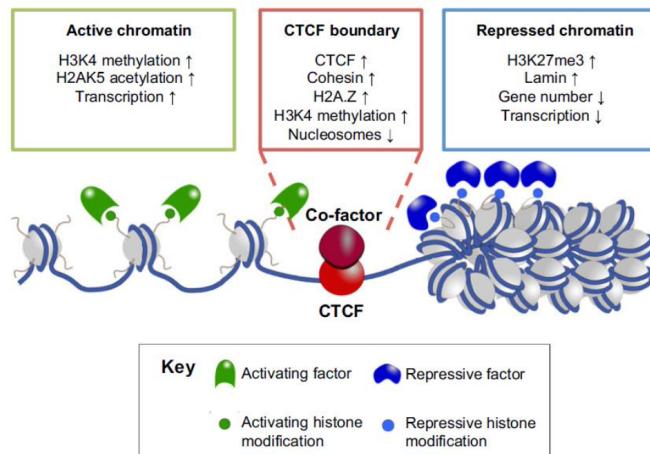


Figure 3.3: Herold et al., *Development*, 2012

By looking at CTCF KO the binding is lost, as well as SCC1 binding.

Cohesin and CTCF co-occupy thousands of genomic binding sites and share the same consensus sequence. In particular, CTCF KO impairs SCC1 binding.

Does cohesion contribute to CTCF enhancer blocking function?

Reporter assay with pIHLIE, which is known to be responsible for the insulator sequence. By removing CTCF, the luciferase activity increases. Nothing happens by impairing sister chromatids link through a drug, Sororin. The insulator function of the H19 ICR requires CTCF binding as well as cohesin binding to DNA, but not the establishment of cohesion.

Cohesin complexes co-localize with CTCF insulator protein and are required for it to block enhancer sequences. CTCF can slide on DNA and when it reaches an already bound CTCF it stops there → functional to repress enhance function.

Does cohesin have anything to do with transcription regulation? Upon CTCF insulator protein (orange) binding to its DNA consensus sequence, the cohesin ring may tether two DNA double

3.1. GENOME ORGANIZATION AND TRANSCRIPTION REGULATION

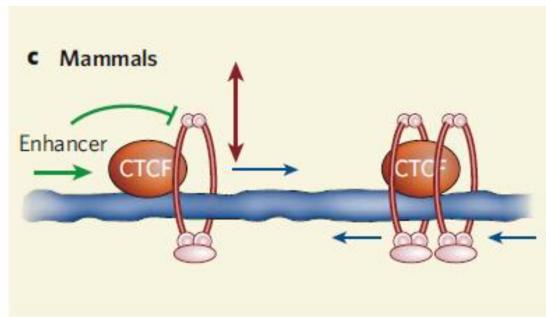


Figure 3.4: Parelho et al, Cell, 2008

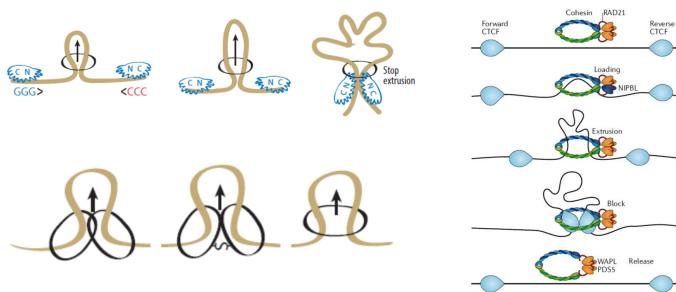


Figure 3.5: Merkenschlager and Nora, Ann Rev Gen Hum Genetics, 2016

strands (red-blue helices) by encircling them. Thus, **cohesion-CTCF complexes could shield genes from the effect of enhancer sequences**, regulating gene expression, by impacting genome organization.

3.1.3 CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation

CTCF binding motifs tend to interact with each other forming a *chromatin loop*. During interphase, DNA is not condensed as in a canonical chromosome, each chromosome forms territories with hierarchical distribution of 3D folding. Different territories can overlap with each other, depending on the extent of histone modificationa. We find compartments, territories and 3D folding of chromatin made by loops. Loops are possibly obtained thanks to the action of cohesins: in *loop extrusion*, a cohesin can slide on DNA for circling 2 ds filaments, until it reaches CTCF is a specific direction. The directionality of CTCF DNA sequences at the loop boundaries indicates the presence of a “tracking” mode for DNA loop formation instead of a simple 3D diffusion.

The hierarchical model of chromatin organization impacts gene regulation:

- linear chromosome map
- local 3D folding: local folding brings convergent pairs of CTCF sites in close spatial proximity fostering the contact between the enhancer and the target promoter

3.1. GENOME ORGANIZATION AND TRANSCRIPTION REGULATION

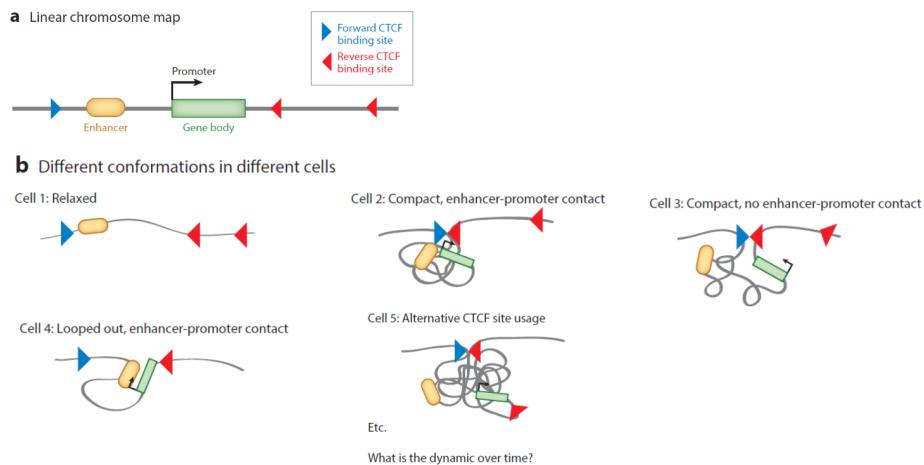


Figure 3.6: Merkenschlager and Nora, Ann Rev Gen Hum Genetics, 2016

- segmentation into TADs: topologically associated domains (TADs) are self associating chromosome segments with high frequency of DNA interaction within them. TADs folding packages enhancers and promoters from the same domain while insulating them from regulatory elements of neighbouring domains.
- compartmentalization of the chromosome territory: euchromatin and heterochromatin compartments. The same compartment can also include DNA from different territories/chromosomes. Compartment A and Compartment B can be subdivided in subcompartments with distinct patterns of histone modifications.

Most TADs are constrained within CTCF-cohesin boundaries and may consist of a single or multiple DNA loops called insulated neighborhoods (or loop domains, or sub-TADs).

The sequences that are within a loop are interacting with each other with higher frequency as compared with canonical interaction rates.

The 3D structure of the genome leads to TAD (topologically associated domains), characterized by a high frequency of DNA interaction within them. Compartment A (euchromatin) and B (heterochromatin) can be subdivided in subcompartments with distinct patterns of histone modifications.

Most of the times, TAD consists of one *insulated neighbourhood* (1 loop). We can also find other cohesins inside the same TAD, resulting in *nested* insulated neighbourhoods or *two insulated neighbourhoods*. 90% of enhancer promoter interactions occur within neighbourhood boundaries in human ESCs and T cells. However, promoters and enhancers do not exclusively interact with each other, as they interact also with other sequences in the same loop.

Chromatin conformation within TADs is highly variable from cell to cell and contacts between DNA loops are dynamics.

3.2. CHROMOSOME CONFORMATION CAPTURE (3C)

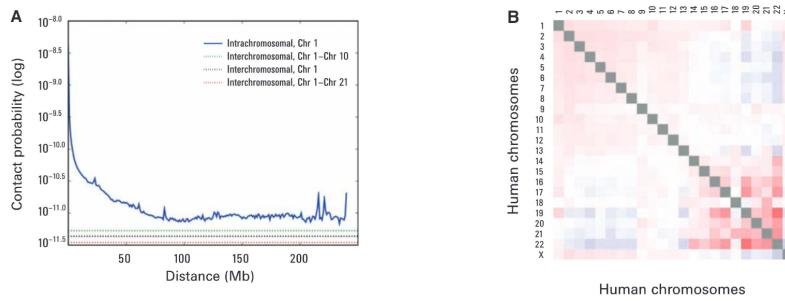


Figure 3.7: Lieberman-Aiden *et al.*, *Science* 2009

3.2 Chromosome Conformation Capture (3C)

This method is used to verify whether distant sequences are interacting in their 3D folding. 3D conformations at the regional, chromosome and whole genome levels can be inferred by calculating the number of ligation junctions between genomic loci.

It is required to have information on all sequences of interest. It is possible to apply inverse PCR: we circularize and amplify from known primers a piece of unknown sequence. In order to test a bigger region, we can use specific primers and verify which sequence is mapped where (computational analysis).

ChIA-PET: which protein is mediating the interaction among two sequences? Use antibody for the protein of interest e.g. CTCF, pull down cross linked, digest and sequence.

Lastly, in Hi-C approach there is a fill in of the extremities of the restriction sites with biotin and direct sequencing → higher throughput.

3.3 Hi-C approach

1. HindIII restriction enzyme
2. Fill extremities and mark with biotin
3. Ligate (NheI)
4. Purify and shear DNA; pull down biotin
5. Sequence using paired-end

A Hi-C map, or **contact matrix**, is a list of DNA-DNA contacts produced by a Hi-C experiment. Spatial proximity maps of the human genome generated with Hi-C at a resolution of 1 megabase. Each pixel represents all interactions between a 1-Mb locus and another 1-Mb locus.

Changing restriction enzyme, a similar pattern is observed. The closer two sequences are in linear form, the higher their contact probability (frequency of interaction).

The probability of contact decreases as a function of genomic distance on chromosome 1. Interchromosomal interactions are depleted relative to intrachromosomal interactions. The level of interchromosomal contact differs for different pairs of chromosomes. This suggests that the chromosomes form defined territories, with an overall frequency of interaction.

3.3. HI-C APPROACH

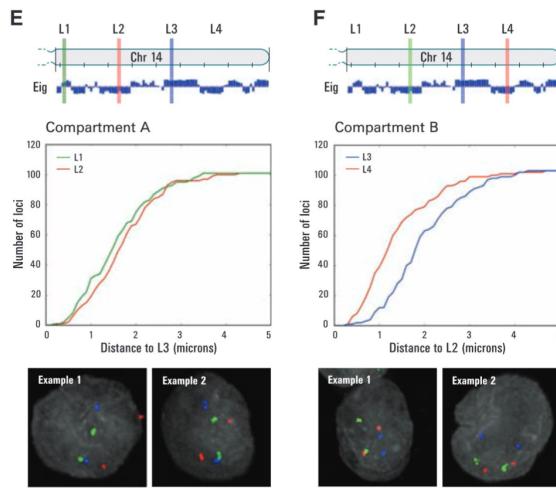


Figure 3.8: Lieberman-Aiden *et al.*, *Science* 2009

Depending on the chromosome, we can find preferential contacts. Example: 20-21-22 region. In low resolution, we observe conserved interaction for synthetic or orthologous sequences.

Each chromosome can be decomposed into two sets of loci (arbitrarily labeled A and B) such that contacts within each set are enriched and contacts between sets are depleted. Regions tend to be closer in space if they belong to the same compartment (A versus B).

Bimodal pattern in eigenvector → either A or B compartment.

3D-FISH to probe four loci (L1, L2, L3, and L4) on chromosome 14 that alternate between the two compartments (L1 and L3 in compartment A; L2 and L4 in compartment B).

- Compartment B showed a consistently higher interaction frequency at a given genomic distance than pairs of loci in compartment A, suggesting that it is more densely packed.
- Compartment A correlates strongly with the presence of genes, higher mRNA expression, and accessible chromatin, as measured by DNase I sensitivity.
- Open and closed chromatin occupies different compartments in the nucleus.
- Such partitioning of the human genome was obtained based on the contact frequency of millions of loci at 1 Mb resolution.

3.3.1 Optimization of Hi-C approach

Digestion *in situ*, permeabilize nuclei from cell and without lysing them introduce restriction enzyme cutting 4 nucleotide bases → 1kb resolution, proximity ligation performed in intact nuclei (reducing the frequency of spurious contacts).

We can divide a compartment in 6 subcompartments: A1,A2,B1,B2,B3 and B4. The genomic and epigenetic content is different: different density of active genes, CG richness, epigenetic marks,... If we increase the resolution within subcompartments, we find frequently interacting regions, which are called domains:

3.3. HI-C APPROACH

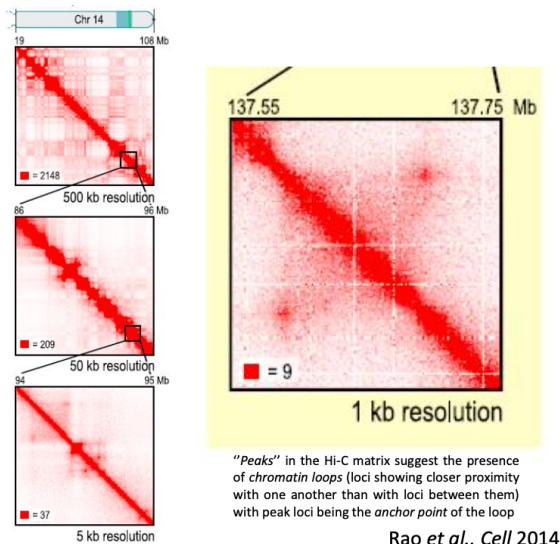


Figure 3.9

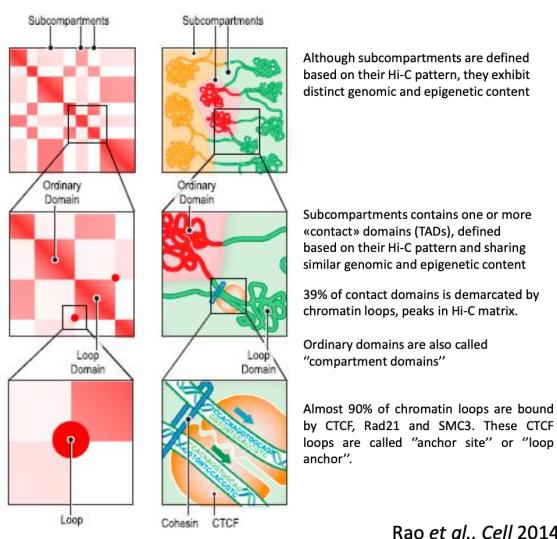


Figure 3.10

3.3. HI-C APPROACH

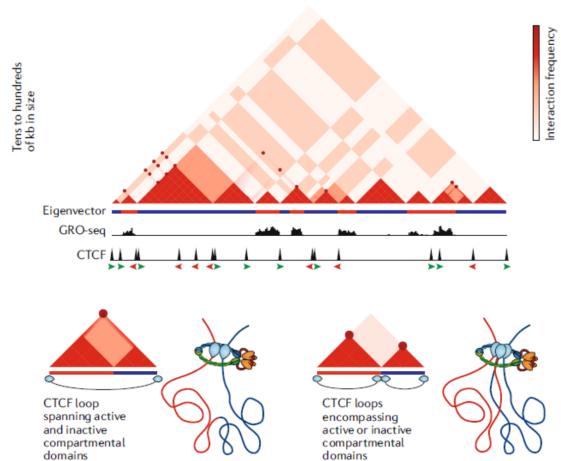


Figure 3.11: Rowley and Corces, *Nature Rev Genetics* 2018

- *loop domains* (90% bound by TFs, they are also called “anchor sites”)
- *ordinary domains or compartment domains*

Depending on the resolution, we can define subcompartments or domains TADs.

Chromosome compartments are formed by aggregation of multiple domains with similar biochemical or functional properties. In literature some studies indicate “sub-TADs” or “insulated neighborhoods” to refer to domains identified by high-resolution Hi-C as being smaller than the domains TADs, either ordinary or loop domains.

The frequency of interaction can also be depicted as a linear genomic coordinate. A Hi-C map can be integrated with ChIP-seq and RNA-seq data to assess the complexity of the domains identified by low resolution data.

The identification of TADs is highly dependent on the resolution of the Hi-C data and the scale at which the data are analysed

Some CTCF loops encompass active and inactive domains.

Given the high frequency of interaction, we can observe co-regulation favouring. Also, since different TADs do not contain frequently interacting genes, a TAD could feature enhancer blocking. Boundary function: ability of repressing the spreading of heterochromatin. Finally, repressing transcription from one TAD to another.

Conclusions:

- Genomes are partitioned into ‘contact domains’ (ordinary domains or loop domains), or TADs, with median length of 185 kb, that share similar chromatin states and tend to associate with each other forming subcompartments (some people refer TADs to the subcompartments)
- CTCF and the cohesin associate with loop domains and are found at over 86% of loop anchors.
- The pair of CTCF motifs present at the loop anchors occurs in a convergent orientation in >90% of cases

3.4. MEDIATOR AND COHESIN CONNECT GENE EXPRESSION AND CHROMATIN ARCHITECTURE

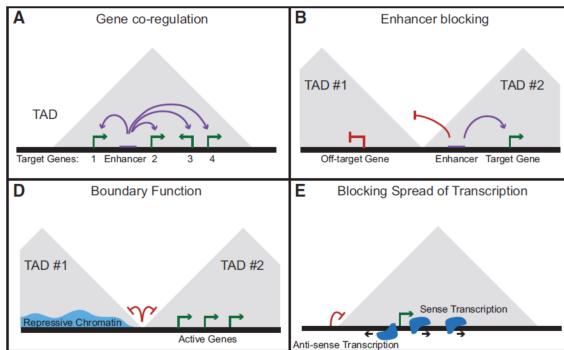


Figure 3.12: Dixon *et al.*, *Mol Cell.* 2016

- Loops frequently link promoters and enhancers, correlate with gene activation, and show conservation across cell types and species (up to 75% are conserved between cell lines, and 50% between mouse and human orthologous genomic regions)

3.4 Mediator and cohesin connect gene expression and chromatin architecture

Enhancers and genes generally interact within the context of the CTCF-CTCF loops which form insulated neighborhoods that constrain interactions between regulatory elements and genes. This structure helps explain why enhancers generally control only a limited number of genes despite having an ability to function in either orientation and at long distances.

Do enhancers and promoters randomly interact within the same loop domain? Probably not, since single CTCF loops can contain both active and inactive genes.

Enhancer-gene interactions occur in insulated neighborhoods by synergic activity of Mediator and Cohesin, and not CTCF.

Mediator and cohesin physically and functionally connect the enhancers and core promoters of active genes within DNA loop domains in murine embryonic stem cells. When the transcription activators bind mediator, the mediator complex undergoes a conformational change, and this activator-bound form of mediator binds cohesin and its loading factor Nipbl, which all contribute to gene activity and DNA looping. Mediator and cohesin co-occupy different promoters in different cells, thus generating cell-type-specific DNA loops linked to the gene expression program of each cell.

There are different binding sites for cohesin mediator and cohesin CTCF. Distinct overlapping pattern in interaction sites, depletion by shRNA results in a very similar gene expression.

Co-IP:

1. lysate and wash
2. add antibody-immobilized agarose resin
3. precipitate immune complexes and wash
4. analysis by Western blot (nuclear extract vs Co-IP with IgG control)

3.4. MEDIATOR AND COHESIN CONNECT GENE EXPRESSION AND CHROMATIN ARCHITECTURE

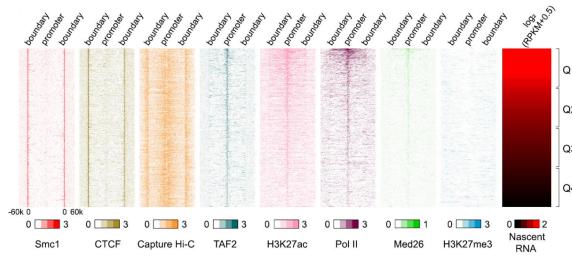


Figure 3.13: Sun et al., 2019

Adapt Co-IP with ChIP: protein in chromatin complex, instead of extracting RNA de-crosslink and run protein on a gel.

Co-IP results indicate that Cohesin and Mediator interact. Chromosome Conformation Capture (3C) analyses confirmed DNA looping between promoters and enhancers of active genes. Depleting cohesin, frequency of interaction is disrupted, similar as Mediator 12.

Mediator and cohesin physically and functionally connect the enhancers and core promoters of active genes within DNA loop domains in murine embryonic stem cells

When the transcription activators bind mediator, the mediator complex undergoes a conformational change, and this activator-bound form of mediator binds cohesin and its loading factor Nipbl, which all contribute to gene activity and DNA looping. Mediator and cohesin co-occupy different promoters in different cells, thus generating cell-type-specific DNA loops linked to the gene expression program of each cell.

3.4.1 Promoter-Enhancer Communication Occurs Primarily within Insulated neighbourhoods

Computational analyses of published CTCF and Smc1 ChIA-PET datasets enabled identification of insulated neighborhoods encompassing 9,407 protein-coding genes of which 3,929 were transcriptionally active in mouse ESCs.

The sites of co-binding of cohesin and CTCF define *DNA loops boundaries*. We can recognize promoters by TAF2 Hac, POL II (not much Med26, repressive).

Enrichment at the boundaries, functional meaning has not yet been defined. Group promoters in quartiles (Q1-Q4) according to efficiency of transcription. The extent of interaction in loops is not necessarily dependent on the frequency of interaction. Capture Hi-C showed that almost 80% of promoter-interacting loci are constrained within the insulator or at the boundaries; these interactions do not correlate with transcription levels.

Distribution of the indicated proteins binding and promoter Hi-C interactions in insulated neighbourhoods in different RNA expression quartiles.

Promoter sequences form loops predominantly within the insulated neighborhood, in particular with enhancers and boundaries. Are the enhancer function and gene regulation constrained within the same insulated neighborhood? Is it possible to interfere with the function of specific enhancers and test whether it has an effect on genes inside and outside the insulated neighborhood?

3.4. MEDIATOR AND COHESIN CONNECT GENE EXPRESSION AND CHROMATIN ARCHITECTURE

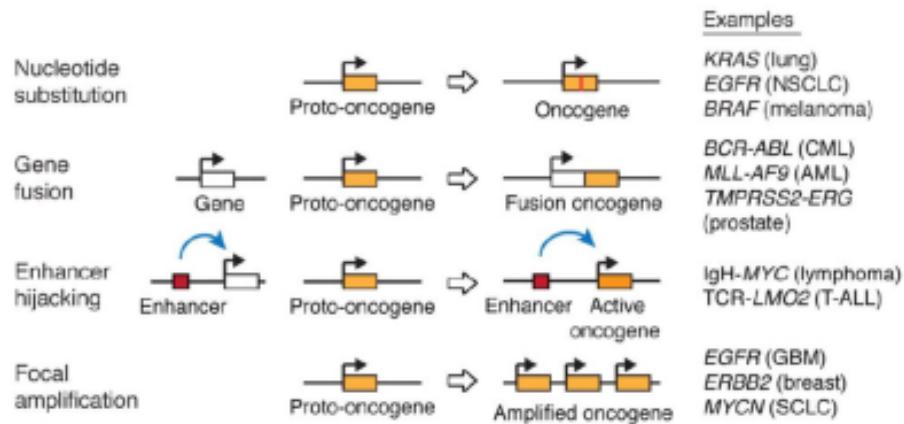


Figure 3.14: Hnisz et al., Cell, 2016

Can the Mediator complex be removed from specific genes? They generated a sequence with promoter and TF binding site, anchored to a magnetic bead. Test with Western blot the binding of the protein. The transcription factor estrogen-related receptor b (**Essrb**) directly interacts with the Mediator complex. Downregulating Essrb would impair Mediator binding to specific enhancers, deregulating expression of distinct target genes.

Indeed, Essrb knock down can be used to deplete Mediator, inactivate enhancers and downregulate target genes.

Now that a system to inactivate enhancers is in place, the authors can test whether enhancer function is constrained within insulated neighborhoods. 222 insulated neighborhoods contained genes downregulated more than 2 folds upon Essrb depletion. Transcription regulation is largely constrained within insulated neighborhoods, and it involves TF-mediator interactions. Is this regulation directly related to promoter- enhancer looping?

3.4.2 Depletion of Mediator causes loss of PIC assembly

Hierarchical structures of chromosomal domains, DNA loops can be clustered, single... Most enhancer-promoter interactions occur during in DNA loops, also called insulated neighborhoods. In the paper, they managed to isolate genes inside the DNA loops. Hypothesis: in DNA loop are present oncogene, which can be activate by different mechanisms, listed in Figure 3.14.

Is there a way for DNA loops to promote oncogene transcription?

In Jurkat cells , 9038 CTCF loops were identified with median length of 270 kb, containing 2-3 genes on average. They crossed adult and embryonic cell lines to discover TADs. Through other analyses they tried to find out presence of loops, enhancer and actively transcribed RNA.

They found out that most cohesin associated enhancer-promoter interactions occurs within CTCF-CTCF loops.

40/55 known oncogenes in the cell line map in insulated neighborhoods. E.g: LMO2 would not

3.4. MEDIATOR AND COHESIN CONNECT GENE EXPRESSION AND CHROMATIN ARCHITECTURE

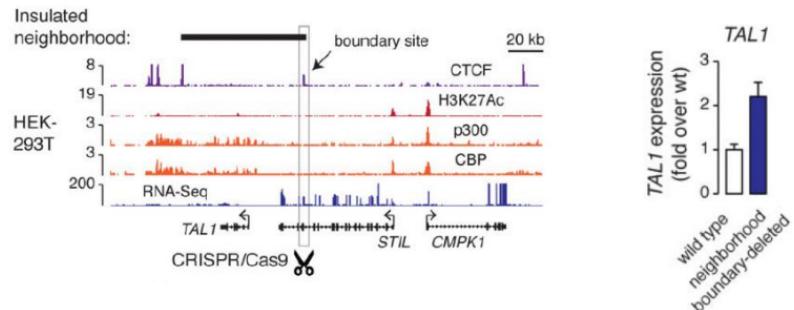


Figure 3.15: Hnisz et al., Cell, 2016

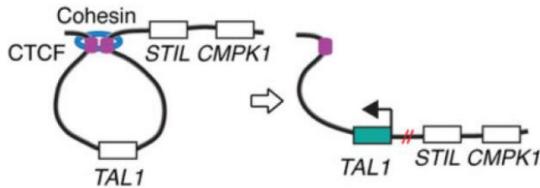


Figure 3.16: Hnisz et al., Cell, 2016

usually be transcribed because enclosed in the loop. But with the help of super-enhancers, proto oncogenes like TAL1 can be transcribed.

Both active oncogenes and silent proto-oncogenes are located within insulated neighborhoods in Jurkat cells. The main hypothesis states that some insulated neighborhoods function to prevent proto-oncogene activation, then some T-ALL tumor cells may have genetic alterations that perturb the CTCF boundaries of neighborhoods containing T-ALL oncogenes.

Example: deletion in promoter of TAL1 -> its transcription is initiated by STIL promoter, which is normally active in T-Cells. TAL1 is normally silenced after fetal stage. Using CRISPr/Cas9 they removed a little piece on the boundary. Disruption of insulated neighborhood boundaries is linked to proto-oncogene activation (Figure 3.15).

The silent state of the TAL1 proto-oncogene is dependent on the integrity of the insulated neighborhood. TAL1 encodes a transcription factor that is overexpressed in ~50% of T-ALL cases and is a key oncogenic driver of this cancer (Figure 3.16).

They tried to do the same on another locus, without the mediation of enhancer or promoters. Even in this case, the disruption of DNA loop allows for the transcription of proto oncogenes.

CTCF in constitutive neighborhoods (present in all cell lines) has a somatic mutation in this cancer, but not the ones in non-boundary CTCF sites (specific to the cell lines). It becomes clear that somatic mutations of insulated neighborhood boundaries occur in the genomes of many different cancers.

Main conclusions of the article:

3.5. SUPER-ENHancers

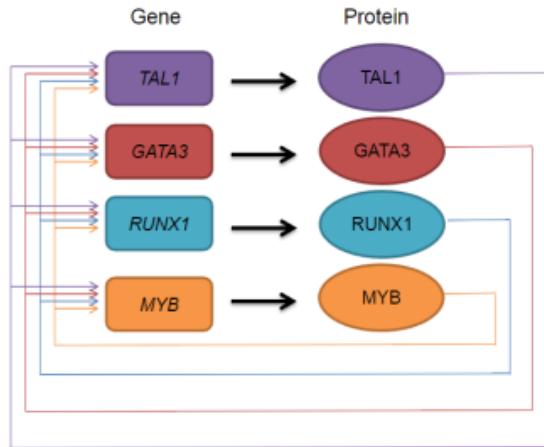


Figure 3.17

- Disruption of insulated neighborhood boundaries, marked by the presence of CTCF and cohesins, can cause oncogene activation in cancer cells
- Recurrent perturbations of boundary elements may impact expression of genes with roles in tumor biology

3.5 Super-enhancers

Super-enhancers regulate genes involved in cell identity: they are bound by tissue-specific TFs and pioneer transcription factors such as OCT4, SOX2, NANOG. They generally activate a self-regulatory circuit: most SE produce promoters, which in turn regulate their own SE and other gene products that they lead.

TAL1 forms transcriptional regulatory networks with other transcription factors in hematopoietic stem cells (Figure 3.17).

This circuits underline the tissue-specificity of most cancers. TAL1 requires a SE, a transcriptional re-wiring.

- The expression of TAL1 is downregulated during normal T-cell development, while GATA3 and RUNX1 are expressed in T cells
- Re-expression of TAL1 in T cells dictates a transcriptional circuitry inducing tumorigenesis
- TAL1, GATA3, RUNX1 and MYB gene loci are associated with super-enhancers in T-ALL cells

3.5. SUPER-ENHancers

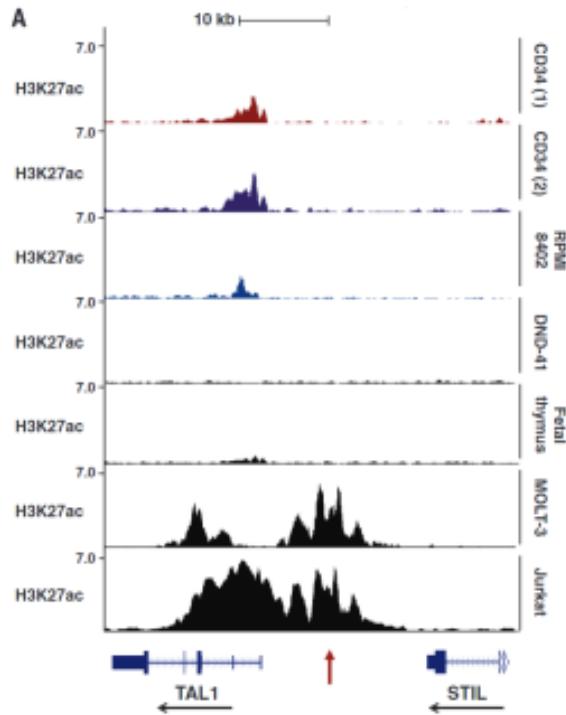


Figure 3.18

3.5.1 How is a SE generated during tumorigenesis?

Even SNPs can generate super-enhancers! TAL1 SE was not found in T cells, or hematopoietic stem and progenitor cells (HSPCs). 3C experiments identified a looping interaction within the SE (red arrow), not anymore an insulated neighborhood in this particular site (Figure 3.18).

When investigating the reason for this difference, they found out that some nucleotide addition (but not much). Small insertion leads to a TFBS in T-ALL cancer cell lines. Only transcription factors expressed in T-ALL cells are involved in the activation of the mutant enhancer.

BUT a mutation on the promoter of TAL1 is not enough. We need all the transcription circuit of the actors that co-regulate each-other. The Mediator binding pattern stretching for 20kb, overlapping with the H2K27ac mark (seen in previous ChIP), indicates the presence of a super-enhancer. Not observed in HPSCs or T-ALL cells devoid of the mutated TAL1 enhancer site: it is the de novo MYB site to promote TAL1 complex formation and induce TAL1 expression.

ChIP seq revealed the presence of a TAL1 complex, including MYB, at the TAL1 mutated enhancer. Targeted deletion of the TAL1 enhancer mutation collapses the TAL1 super-enhancer (Figure 3.19).

This results in a gain of function even in heterozygosity. A SE can be generated by a single transcription factor biding site and can extend to hundred of bases. They are very sensitive to mutations.

Finally, there's mainly two ways to activate proto-oncogenes (Figure 3.20). Both can be triggered by very short deletions/mutations that happen in regulatory regions.

3.5. SUPER-ENHancers

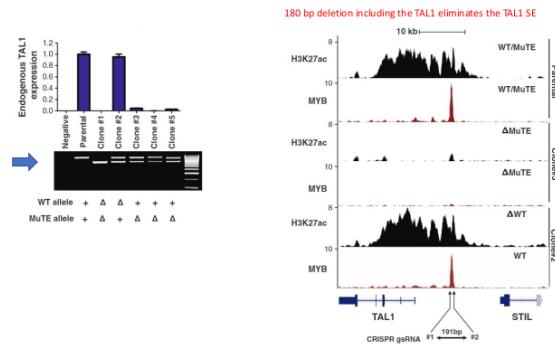


Figure 3.19

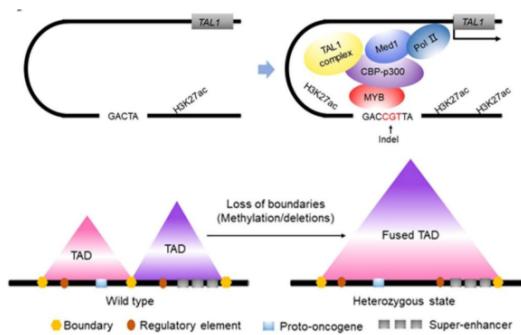


Figure 3.20

3.5. SUPER-ENHancers

Cancer	GISTIC peak	Enhancer	Candidate
UCEC	chr8:129,543,949–129,554,294	Super	MYC
HNSC	chr13:73,880,680–73,990,596	Super	KLF5
LUAD	chr8:129,166,547–129,190,290	Super	MYC
CRC	chr13:27,523,026–27,544,353	Super	USP12
LIHC	chr20:48,997,377–49,019,434	Super	PARD6B
ESCA	chr13:73,880,413–74,042,621	Super	KLF5

Figure 3.21

More recent genomic insertions form enhancers that misregulate oncogenes; this hypothesis was tested in different cancer cell lines.

3.5.2 How cancer cells take advantage of super-enhancers

In the study **Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers** scientists interrogated a dataset of 10k tumors and found recurrent focal amplifications. They found 25 peaks which correspond to super-enhancers. They also found that six different cancer cell types had some common recurrent focal amplification in the same spot (Figure 3.21).

- UCEC, uterine corpus endometrial carcinoma;
- HNSC, head and neck squamous cell carcinoma;
- LUAD, lung adenocarcinoma;
- CRC, colorectal carcinoma;
- LIHC, liver hepatocellular carcinoma;
- CESC, cervical squamous cell carcinoma;
- ESCA, esophageal carcinoma.

KLF5 is a putative oncogene that can be upregulated in tumors by super-enhancer amplification. For the **MYC** genes, they found data in two different cancer lines. Selectivity of amplification depending on the cell line. Tumors with amplification of MYC alone or MYC-LASE or MYC-ECSE alone express higher levels of MYC than tumors lacking either amplification (meaning, there's no need for the tumor to evolve a over-expression of MYC).

NFE2L2 and CEBPB transcription factors regulate the e3 enhancer: so to study it, we can down-regulate these two factors.

Conclusion: MYC regulation changes based on the tumor and tissue type. If we are in a cancer setting we can target a super-enhancer and we impact the expression of the target gene. However, there are several onco-genes in a cancer setting.

3.6. THE ROLE OF MYC IN CANCER CELLS

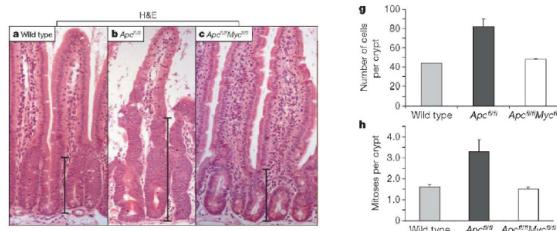


Figure 3.22: Deletion of both Myc and APC from small intestine cells

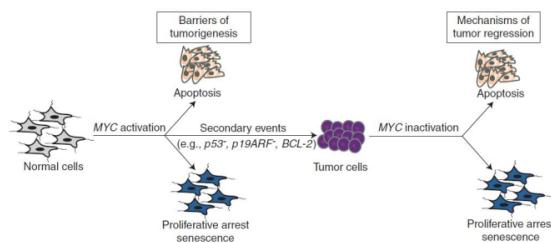


Figure 3.23: Gabai et al., CSHL 2014

3.6 The role of MYC in cancer cells

In cancer, MYC is always the most amplified gene. MYC is the downstream target of so many pro-proliferation pathways, these cells are not even under selective pressure to activate the pathways. It has a great impact on cells!

The inhibition of MYC and APC rescues or prevent tumorigenesis. By removing one single target gene we can have drastic results and MYC is capable of driving tumorigenesis on its own (Figure 3.22).

In most human tumour types, the expression of MYC proteins is deregulated and enhanced relative to the corresponding normal tissue, and high MYC expression often correlates with poor prognosis. **MYC induces cell proliferation and growth** in the absence of external growth factors. Conversely, inhibition of MYC almost invariably suppresses cell proliferation. Transgenic mouse studies provided the evidence that deregulated expression of MYC is sufficient to drive tumorigenesis in a number of transgenic mouse tissue.

3.6.1 Key aspects of MYC in tumorigenesis

Cells become addicted to a single oncogene! (and to transcription in general). The key aspects of MYC are recapitulated in Figure 3.23.

3.6.1.1 MYC transcription factor forms a heterodimer with MAX

MYC is quite an unstable protein, even its mRNA has a very short life. It has an one-of-a-kind binding domain, it is a very special protein.

3.6. THE ROLE OF MYC IN CANCER CELLS

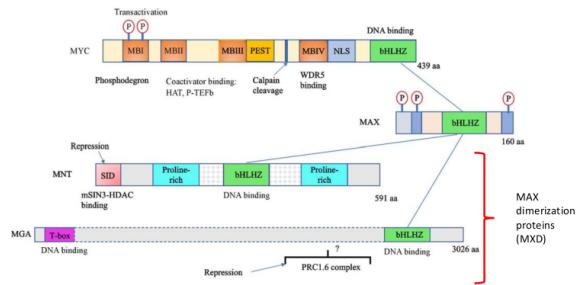


Figure 3.24: MYC is a transcription factor forming a heterodimer with MAX

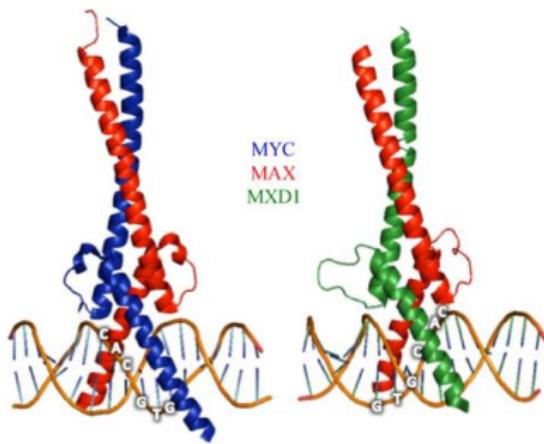


Figure 3.25: Carroll et al., Front.Med. 2018

MYC, MAX and MXDI all recognise the same site (Figure 3.24). MXD and MYC act as antagonists. MXD proteins can recruit histone deacetylases and repress transcription while MYC proteins recruit complexes that promote transcription (not always the case...). MYC stimulates cell growth and proliferation, while MXD inhibits cell growth and proliferation and blocks MYC effects. MYC-MAX bind to thousands of genes whose identity depends on the cellular transcriptome and the abundance of MYC protein (Figure 3.25).

Studying MYC is very difficult. What we know is that if we do a chIP for MYC we have an extreme range of possible results, result is basically random. For RNA-seq: regulation of genes from MYC is also very subtle, with down-up regulated genes changing like 20%, but also transcribe for TF that regulate their own site.

Trunk sustains the outcome of MYC, but the net outcome of MYC interaction and transcription we have epigenetic regulation, pre initiation complex assembly, pause-release behaviour....

MYC is not a pioneer TF! So important interaction with proteins recruited on the chromatin. With open chromatin state, we have un-specific transcription. In addition to regulate its own target genes, MYC can ‘amplify’ the transcriptional program already active in a cell, leading to a signature which is cell type-specific. MYC increases the expression of (already poised) thousands of genes (Figure 3.28).

3.6. THE ROLE OF MYC IN CANCER CELLS

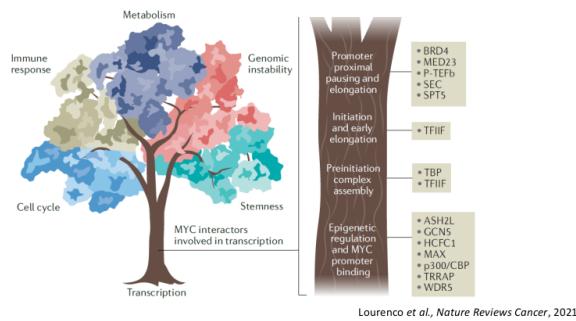


Figure 3.26: Lourenco et al., *Nature Reviews Cancer*, 2021

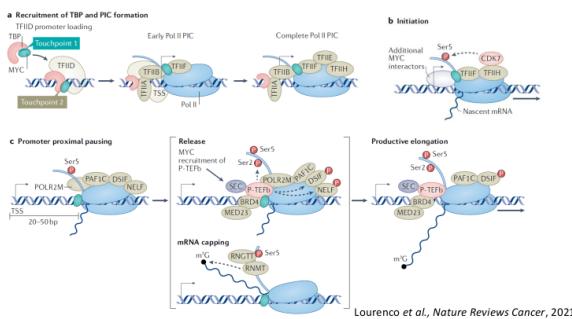


Figure 3.27: Lourenco et al., *Nature Reviews Cancer*, 2021

It depends a lot on the level of MYC2: since there are low affinity and high affinity binding sites, the more the MYC the higher the probability of MYC binding the low-affinity regions. Genes can be either down or up regulated.

BUT tumorigenesis is not only about proliferation: we need proteins, membranes,... obviously MYC is also involved in anti-apoptosis behaviour, energy metabolism etc.

3.6.2 By which major mechanism or pathway do MYC proteins promote tumorigenesis?

The final outcome of MYC gene regulation depends on the network of MYC interactions, in particular MYC-MAX and MAX-MXD and by the transcriptional program of the cells. MYC is a mild transcriptional modulator, inducing moderate changes in mRNA levels. It is able to regulate the expression of several other transcription factors and cofactors, including E2F proteins, Polycomb group repressive complex 2 (PRC2) components, along with microRNAs (miRNAs), thus orchestrating complex networks of gene regulation. MYC dependent gene regulation has been shown to be dependent on: cell type, experimental conditions (stress), levels of MYC.

In conclusion, MYC activation contributes to several hallmarks of cancer:

1. autonomous proliferation and growth
2. sustained DNA replication

3.6. THE ROLE OF MYC IN CANCER CELLS

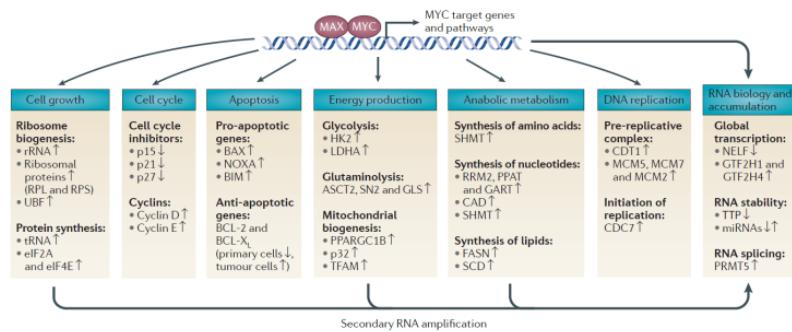


Figure 3.28: Kress et al, NatureRevCancer-2015

3. increased protein biogenesis
4. global changes in cellular metabolism
5. activation of the angiogenic switch
6. suppression of the response to autocrine and paracrine regulatory programs
7. restraint of host immune responses
8. The transcription rewiring induced by oncogenic Myc levels also renders cancer cells addicted to this particular oncogene

Chapter 4

Targeting transcription in cancer

In order to perform transcription targeting in cancer, we first need to understand which genes are driving tumorigenesis.

4.1 Age-related remodelling of oesophageal epithelia by mutated cancer drivers

Biopsies from cancer (red) and non-cancer (green) sites from 58 y.o. male with risk for esophagus cancer (smoking and drinking heavily). By looking at the mutational burden, cancer samples has a higher MCF (mutant cell fraction). The normal biopsies showed several mutations too. If we look closer, mutations are hitting driver genes e.g. NOTCH, TP53. Cancer samples' mutations are overlapping to a good extent and are present at a higher degree. In normal tissues we have *independent mutations*, which suggests that independent clones are present in a normal epithelium.

Divide patient samples in low and high risk, in normal tissues. The number of mutations per sample increases with risk and over time. By analyzing the mutation, 4 mutational signatures were identified: e.g. APOBEC is responsible for globulin production, should be restricted to lymphocytes, C→T mutation (Figure 4.2).

Normal tissues can accumulate mutations to an extent dependent on lifestyle and age.

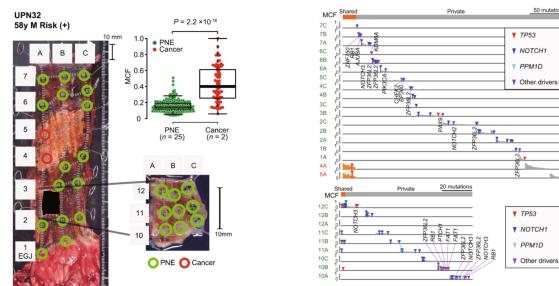


Figure 4.1: Yokoyama et al., Nature 2019

4.1. AGE-RELATED REMODELLING OF OESOPHAGEAL EPITHELIUM BY MUTATED CANCER DRIVERS

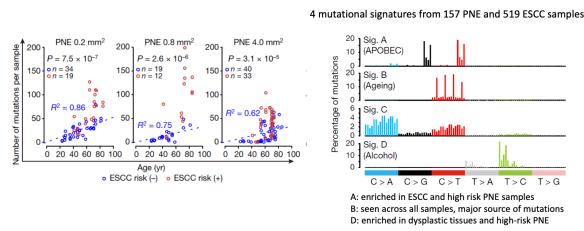


Figure 4.2: Yokoyama et al., Nature 2019

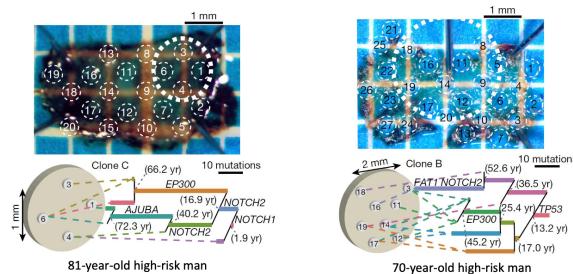


Figure 4.3: Yokoyama et al., Nature 2019

Single cell sequencing analysis: the mutation number in colonies from low-risk individuals linearly increases with age with an annual increase of 41.5 mutations per genome per year.

Driver genes are mutated in PNE although with different frequencies as compared with cancer samples and there are substantial differences in the major drivers between PNE and ESCC. Most of the PNE samples from high-risk individuals (62 out of 64, 97%) contained one or more driver mutations. The positivity of driver mutations was 74% (69 out of 93) of all samples from low-risk individuals.

Spatial distribution and mutational composition of clones within the same biopsy:

- Young individuals: small number of shared mutations with small MCF. Limited number of driver-mutant clones. During development to early adult life stem cells and their progenies spread giving rise to **somatic mosaicism** more than positive selection.
- Elderly individuals: shared mutations and MCF markedly increase. Driver-mutant clones increase in number and size, a trend which is enhanced with lifestyle ESCC risk. Positive selection may be ongoing.

In young individuals we observe a mosaicism of small mutations, while advancing with age we can acquire more mutations which might lead to cell proliferation increase and higher probability of new mutations.

Some mutations can occur early on in life and they accumulate on specific proliferative pathway they might lead to cancer or not.

80-year-old high-risk man: Divergent clones populating a small epithelial region originated from a single stem cell with a NOTCH1 mutation almost 80 years previously. Positive selection lead to the fixation of 4 driver mutations and subclones formation. 70-year-old high-risk man: Clonal evolution

4.2. THE RISK OF SUFFERING ANY CANCER INCREASES DURING AGING

initiated in a stem cell with a TP53 mutation at the age of about 13 years which persisted without evolving to cancer. Meanwhile other driver mutations were acquired.

Conclusions

Mutations inevitably occur in normal cells and accumulate during aging, their frequency depends on lifestyle risks. Such mutations may be positively selected (driver mutations), favoring changes that are beneficial to the individual cells, generating clones. These clones must acquire a number of specific driver mutations to transform generating cancer cells. Normal cells can become cancerous when the “right” combination of driver genes has been mutated and/or deregulated. Somatic evolution is not a linear road towards cancer, the clone size, the specific mutations and the tissue context may slow down the tumorigenic process or boost it. Tissues differ substantially in their susceptibility to specific oncogenic events and barriers to tumor formation are tissue specific due to a number of variables, including:

- Tissue-specific oncogenic function of a cancer drivers
- Cell-extrinsic factors such as cell-cell signaling
- The cell of origin and its differentiation status

4.2 The risk of suffering any cancer increases during aging

With age, the spreading of mutant clones within tissues may progressively compromise the tissue contributing to aging, cancer and other diseases. The same pattern was observed in colorectal tissue and especially skin, which has a high burden of somatic mutations.

Standard chemotherapy relies on targeting proliferating cells with alkylating agents. We can also apply precision therapy to certain oncogenes, but this approach should take into account that multiple mutations are present. Over the years, it was observed that if the expression of a target oncogene is repressed, we observe growth inhibition and tumor regression.

Cancer cells can become addicted to the expression of specific oncogenes.

Important bias: we are inducing tumorigenesis by inducing overexpression of a single gene, which is not what happens physiologically. However, we know have in clinics drugs targeting different proteins and able to induce tumor regression. One of the first compounds developed was Imatinib, meant to target a kinase fused in CML. Roughly 53 out of 54 patients were cured in phase I clinical trial. All the targets hit by target therapy drugs are enzymes, most of them kinases.

Cancer cells tend to lose cellular functions which are not essential to cell viability or do not increase cellular fitness. This is due to the *genetic drift* determined by the mutational burden, epigenetic modifications, and tumor microenvironment. The silencing of redundant functions and pathways can render cancer cells more susceptible to perturbations. If we hit a particular pathway in a cell, we can impair a function which was previously regulated by multiple pathways → synthetic lethality.

If the domain changes e.g. accumulation of mutations on oncogene, cancer cells can develop resistance to target therapies. The same happens with new mutations on alternative pathways.

4.3. TARGETED THERAPEUTICS GENERALLY LEAD TO RESISTANCE

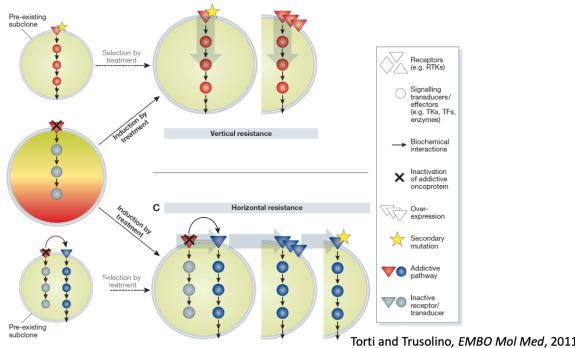


Figure 4.4: Screen Shot 2022-11-04 at 11.25.54.png

4.3 Targeted therapeutics generally lead to resistance

Transcriptional regulators are essential effectors of the transcriptional program imposed by oncogenic drivers, also called transcriptional drivers.

RTK growth factor, frequently hit by mutations in cancer. Resistance can arise through the persistence of downstream pathway or parallel signaling transduction. This occurs as pathways converge to the same effectors, which are either co-factors regulating transcription or transcription factors. E.g. MYC or beta-catenin, which are also called *transcriptional drivers*.

Example: MYC inhibition eradicates K-RAS (driver oncogene) driven lung cancers in mice.

The consequences of oncogene inactivation for the reversal of tumorigenesis depend on the type of tumor and the genetic context. E.g. if we remove MYC from MYC driven lymphoma, we observe apoptosis of lymphoma cells. In hepatocellular carcinoma, the re-expression of MYC leads to restoring cancer. Cancer cells can escape dependence on oncogenes by acquiring other genetic events.

4.3.1 Mutations in cancer

Cancer is caused primarily by genetic mutations and is initiated and maintained by recurrent driver mutations. Cancer cells genome undergoes a constant genetic drift. Transcription deregulation is a hallmark of cancer, as cancer cells inevitably undergo transcription rewiring. Dysregulated transcriptional programs fuel tumorigenesis. Cancers can remain addicted to the oncogenes that have driven tumorigenesis and/or to the transcriptional drivers (definitive cancer drivers). Dysregulated transcription also creates transcriptional dependencies which are not typically identified by cancer genome sequencing. **Transcriptional addiction** in cancer can be harnessed for therapeutic intervention.

Resistance to targeted therapy can occur by activating *alternative signaling molecules* and pathways converging on the same transcriptional regulators or by *acquiring mutations* on the targeted gene.

Transcriptional regulators act as effectors of the oncogenic drivers. Cancer cells can become as “addicted” to these effectors as they are to the oncogenic drivers.

The complexity of TFs networks renders resistance mechanisms difficult to be developed.

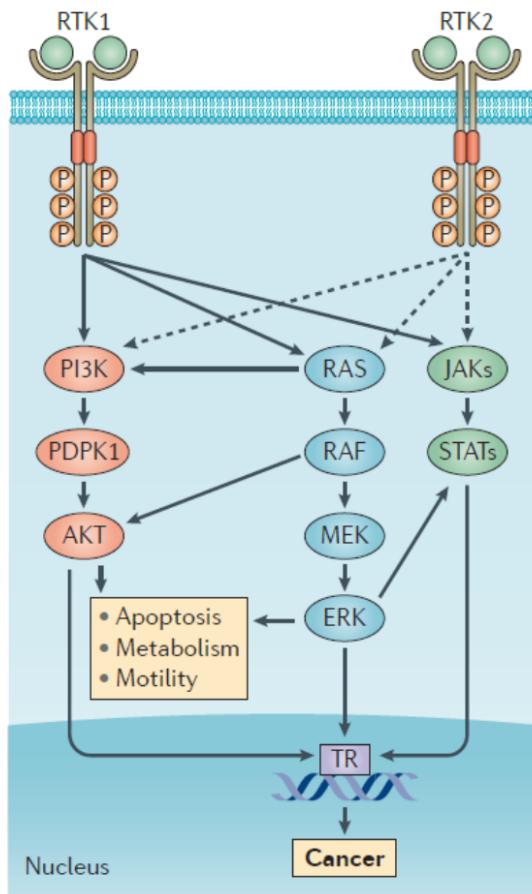


Figure 4.5: Gonda and Ramsay, *Nature Review Mol Cell Biology*, 2015

Transcription factors interact in complex networks, also involving other transcription regulators. It is unlikely that any TF, or non-TFs transcriptional regulators such as CBP-p300, BRD4, can be completely replaced by another. It is more difficult to bypass this particular targeting.

Targeting transcription regulators will have a lower likelihood of emergence of resistance than targeting intracellular signaling pathways, also because multiple pathways can converge on the same transcription regulator. Inhibiting a specific transcriptional program may impact more than one oncogenic signal. The main issue is to identify transcriptional drivers, which might not be affected by mutations; it is required to perform wet lab experiments. Furthermore, target mutations can still emerge as resistance mechanism.

4.4 Target transcription in cancer

There exist numerous ways to target transcription in cancer:

4.4. TARGET TRANSCRIPTION IN CANCER

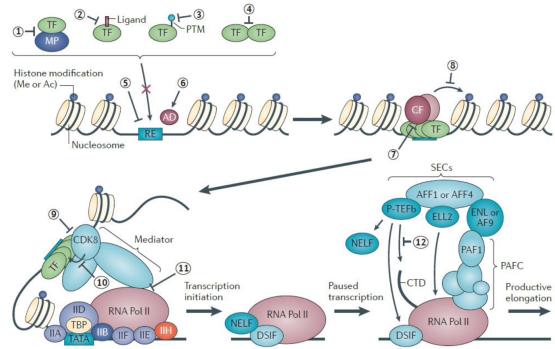


Figure 4.6: Bywater *et al.*, *Nature Review Cancer*- 2013

- interaction with co-factor
- interaction with ligand → most effective mechanism, clinical trials
- PTM
- TF-TF interaction
- impairing recruitment to responsive elements

The mechanism of action can be PROTAC or monomeric degradation: they bind their own target and promote degradation. The **monomeric degrader** can alter 3D structure to make accessible to proteosomal degradation. **PROTAC** has a domain for E3 ligase, which leads to the recruitment of protease. In principle these molecules can interact with any region of the protein e.g. epitope, AA sequence (not necessarily domain). Once the protein is degraded, the molecule is still around and can be recycled to degrade other proteins (1:1 stoichiometry is not required).

Bywater *et al.*, *Nature Review Cancer*- 2013

The Mediator and other complexes are controlled by kinases, so we can foster their inhibition.

THZ1 interacts with the ATP binding pocket of CDK-7, covalent bond with Cys321. In the study the THZ1-R,R was used as a control, as it should not achieve Cys binding.

By pulling down THZ1, we can visualize CDK7 activity: by increasing the amount of THZ1, the competition with bio-THZ1 is overcome and CDK7 expression is inhibited. RNAPII phosphorylation is impaired in vitro with high amounts of THZ1. If the experiment is performed with mutated C312S, phosphorylation of Pol II is not inhibited anymore.

Washout experiment: treat compound, wash and let cells grow without the compound. In this case CDK7 inhibition persists, since it is irreversible thanks to covalent binding.

THZ1 treatment shows broad antiproliferative activity in cancer cell lines by modulation of transcription.

High throughput screening of cell lines in micro plates to test cell proliferation shows that T-ALL cells are particularly sensitive to small perturbations in transcription and CDK7 kinase function. Instead BJ fibroblasts and RPE-1 were not affected.

Bioluminescent xenografted mouse model confirmed efficacy of THZ1 in blocking tumorigenesis of the human T-ALL cell-line KOPTK1. No toxicity of the compound was observed in mice. High doses of THZ1 were required to impair proliferation of non transformed cells.

4.5. BRD4

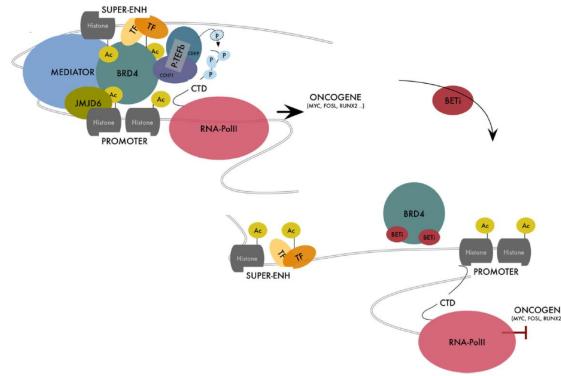


Figure 4.7: Screen Shot 2022-11-13 at 19-35-28.png

Why do cancer cells and in particular T-ALL cells are more sensitive to CDK7 inhibition than non cancer cells? By applying 250 nM of THZ1 we observe the reduction of transcription of all genes. If we lower the concentration at 50 nM, only a subset of genes will be impacted e.g. TAL1, GATA3, RUNX1 → super enhancer region, these factors autoregulate their own gene expression while simultaneously regulating many other genes.

RUNX1 forms a core regulatory circuitry with TAL1 and GATA3 transcription factors that have prominent roles in leukaemia biology

Takehome message:

Targeting transcription is a promising anticancer approach which can be achieved by CDK7 inhibition. Transcription rewiring of cancer cells can be triggered by acquisition of super-enhancers driving oncogene expression and tumorigenesis. Super-enhancer driven oncogenes lead to transcription addiction generating transcriptional dependencies which can be used for therapeutic intervention. Cancer cells require continuous active transcription which renders them more sensitive to transcription targeting therapeutics. In particular oncogenic drivers like MYC and RUNX1 have short mRNA and protein half-lives and depends on continuous transcription. Genomic sequences may not suffice to identify driver oncogenes and cancer vulnerabilities.

4.5 BRD4

BRD4 is a member of the Bromodomain and Extraterminal (BET) family along with BRD2, BRD3 and BRDT. It is a histone acetyltransferase that evicts nucleosomes from chromatin. In particular, BET inhibitors impair super-enhancer driven oncogene expression.

4.5.1 Selective inhibition of Tumor Oncogenes by Disruption of Super-Enhancers

Super-enhancers associated genes were expressed at higher levels than enhancer-driven genes and specifically in mieloma cell line. The majority of super enhancer associated genes were previously

4.6. YAP/TAZ

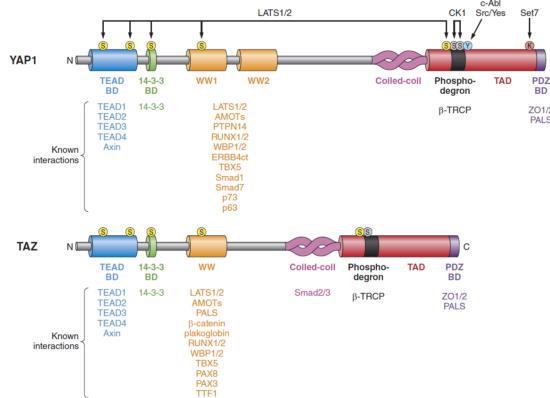


Figure 4.8: Screen Shot 2022-11-13 at 19-40-17.png

shown to be involved in mieloma tumorigenesis e.g. MYC, CCND2. Treatment of MM1 cells with Brd4 inhibitor JQ1 resulted in reduced levels of BRD4 at enhancers and promoters.

JQ1 treatment lead to more pronounced reduction of BRD4 enrichment at super- enhancers showing almost complete loss of BRD4; in particular, super-enhancers are more sensitive to BRD4 inhibition than regular enhancers. MYC is among the most rapidly depleted genes upon BRD4 inhibition.

Genes associated with super-enhancers showed greater decrease in RNAPol II at gene body than the ones associated with enhancers.

- Cancer cells can acquire super-enhancers (SE) as mechanism to drive oncogene expression
- SE are characterized by disproportionately high levels of BRD4 and Mediator, interacting with each other and with multiple cofactors
- SE are highly reliant on cooperatively interacting factors and lose activity more rapidly (than enhancers) when the levels of SE-bound factors are reduced
- BRD4 inhibition leads to preferential disruption of SE driving critical oncogenes in cancer cells

The molecular details underlining cancer cells addiction to BRD4 are being unveiled.

4.6 YAP/TAZ

The transcriptional coactivators YAP (Yes-associated protein)/ TAZ (transcriptional coactivator with PDZ-binding motif) and their interactors.

The **HIPPO signalling** regulates YAP/TAZ and epithelial architecture and cell polarity are inhibitors of YAP/TAZ → EMT triggers inactivation of the Hippo cascade and YAP/TAZ activation.

Mechanical signals are ubiquitous, targeting every cell in every tissue. Mechanotransduction converge on YAP/TAZ in multiple cellular contexts.

YAP and TAZ are nuclear effectors of mechanical signals exerted by ECM and cell shape:

4.6. YAP/TAZ

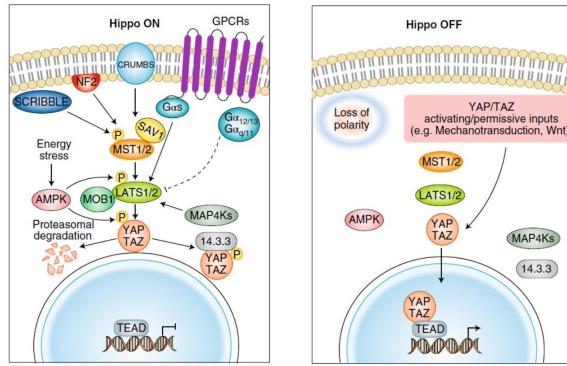


Figure 4.9: Screen Shot 2022-11-13 at 19-41-35.png

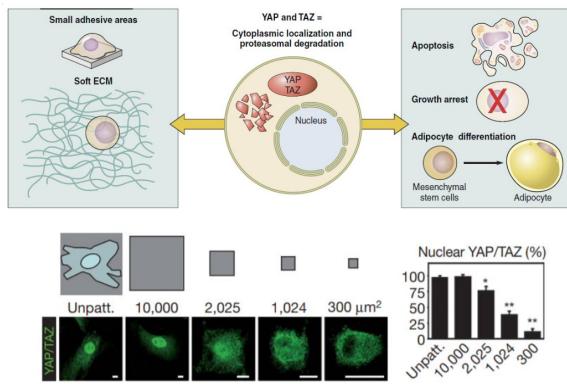


Figure 4.10: Dupont et al., Nature 2011

4.7. TEAD

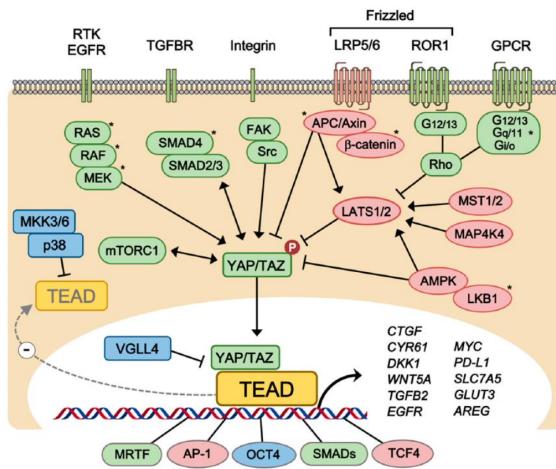


Figure 4.11: Screen Shot 2022-11-13 at 19-48-01.png

The process of tumorigenesis is accompanied by collagen crosslinking, ECM stiffening, and increased focal adhesions. YAP/TAZ activity is controlled by cell shape and polarity through the cytoskeletal structure which also senses the topology and the rigidity of the extracellular matrix.

Cells probe the physical features of the micro environment through integrins and adhesive proteins responding to extracellular forces by adjusting their tensile state through the cytoskeleton activity. Physical and mechanical cues transmitted by the extracellular environment (i.e.: stiffness of the ECM) reach the nucleus at least in part by YAP/TAZ activation.

YAP/TAZ interact with the TEAD transcription factors to regulate gene expression.

4.7 TEAD

TEAD is regulated by Hippo, Wnt, TGF-beta and EGFR pathway and is responsible for controlling drug resistance, metastasis, EMT and cancer stem cells. TEAD-driven transcriptional targets include well-established genes that are involved in cell growth, proliferation, and tissue homeostasis. YAP/TAZ drive several key attributes of cancer cells.

4.7.1 YAP/TAZ/TEAD

Mapping YAP/TAZ/TEAD associations with chromatin in breast cancer cells reveals that these regulators associate with active enhancers. Hi-C analyses predict about 3000 genes regulated by YAP/TAZ/TEAD. 3C analyses confirmed YAP/TAZ/TEAD binding to MYC and TOP2A enhancers.

YAP/TAZ are required for activation of MYC and TOP2A enhancers.

YAP/TAZ-depleted cells stop proliferating and accumulate in G1.

TEAD depletion phenocopies YAP/TAZ depletion → TEAD is determinant for YAP/TAZ induced proliferation. Most YAP/TAZ motifs contained both TEAD and AP-1 motif.

4.8. NUCLEAR RECEPTORS

Requirement of AP-1 for YAP/TAZ/TEAD induced transcription program tumorigenesis.

YAP/TAZ interact with the TEAD transcription factors to regulate gene expression. AP-1 enhances YAP/TAZ/TEAD transcription program and tumorigenesis

BRD4 inhibition impacts the YAP/TAZ transcriptional program in the MDA-MB-231 breast cancer cell line → **Brd4 is required for YAP/TAZ-mediated transcriptional regulation.** In addition YAP/TAZ are required for BRD4 recruitment to chromatin.

Conclusions:

- YAP/TAZ and BRD4 associate physically and functionally in breast cancer cell lines
- YAP/TAZ-BRD4 complex confers a transcriptional advantage to a broad YAP/TAZ targets the expression of these genes can be targeted by BET inhibitors
- Super-enhancers consist of YAP/TAZ-occupied enhancers showing strong enrichment of BRD4, high expression levels of regulated genes and higher sensitivity to BET inhibitors than average
- Oncogenic effect of BRD4 in association to YAP/TAZ offers a new perspective: to stratify patients which are more likely to benefit from BET inhibitors, alone or in combination with other drugs
- Development of new therapeutic approach around YAP/TAZ-BET interaction surfaces

→ YAP/TAZ represent ideal candidates to mediate cancer-specific transcriptional addictions and potential drug targets

Recap

[...] copy from slides

The inhibition of BRD4 could represent an important mechanism to target.

If we analyze tumorigenic mechanisms in vivo and in vitro we will observe different results: processes dictated by the microenvironment will significantly impact on the tumorigenic potential of cancer cells.

Binary pan-cancer: YAP/TAZ are upregulated oncogenes. Cancers that do not express YAP/TAZ can silence its expression, as it will lead to cell death and differentiation. The effect of YAP/TAZ all depends on the transcriptional program. It is possible to impair YAP/TAZ through targeted therapies e.g. Tyr kinase inhibitors or Ser/Thr kinase inhibitors; however we must be careful, as resistance can arise. When the transcriptional program is inhibited, tumor suppression is observed. In the case of YAP^{off} cancers we observe the expression of MYC, which could be targeted. If we manage to re-express YAP/TAZ, they will bind to TEAD and bHLH activating genes promoting adhesion and cell differentiation, always resulting in tumor suppression.

4.8 Nuclear receptors

Nuclear receptors are transcription factors which mediate transcription regulation upon receptor-specific ligand binding. Usually NR are found bound to DNA in the nucleus or to other proteins in the cytoplasm. In most of the cases they are hetero or homodimers. Nuclear receptors share a common architecture - very conserved modular structure composed by 2 transactivator domains, DNA binding domain and C terminus ligand binding domain (11 helices forming a pocket, 12th helix

4.8. NUCLEAR RECEPTORS

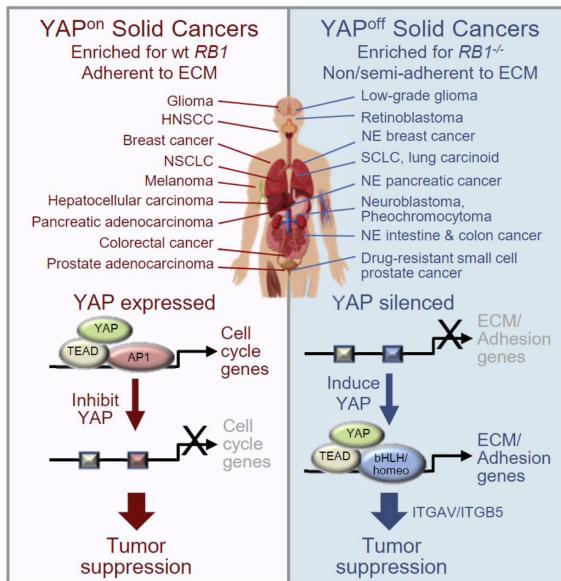


Figure 4.12: Screen Shot 2022-11-18 at 11-11-12.png

outside acting as a pocket closure) - and functional behavior. NR are involved in many processes and in different forms of cancer e.g. estrogen receptor alpha ((ER α)).

ER α is associated with proliferation and development and cells for the formation of tissues and organs in women. ER β has the opposite function, it is repressing proliferation. ER α is expressed in mammary glands and uterus., ER β colon or immune system cells. It is still unclear how the two behave so differently, probably specific TF interaction.

4.8.1 Estrogen receptor α

Estradiol induces dimerization of ER and binding of the dimer to ER response elements (EREs). ER α binds to HSPs and when estradiol is met we observe the conformational change leading to ERE binding. For achieving a complete activation of the NR, PTMs are required e.g. Cdk7 targets S118 for phosphorylation.

The main antagonists of ER are known as **SERM** and **SERD**. The first target therapy for ER, *Tamoxifen*, was developed in the late 60s. Tamoxifen belongs to SERM, modulator interacting with ER, impairing the ability to interact with the binding site. Once ER binds it cannot recruit co-regulators, its transcriptional action is shut down. Short term Tamoxifen treatment does not show high survival rates, but longer times expose patients to side effects e.g. acts as agonist for ER in endometrial cells. Other drugs from SERM family act similarly e.g. Raloxifene and Bazedoxifene. In addition to SERM, we have ER degraders from SERD family → through degradation PTMs e.g. sumoylation. Fulvestrant is a steroid derivative molecule (quite different from Tamoxifen), not water soluble so needs to be injected in muscles.

The two major strategies for therapeutic targeting of hormonal signaling in breast cancer are *direct antagonism* of the ER and *estrogen deprivation*. Aromatase inhibitors can avoid estradiol production

4.8. NUCLEAR RECEPTORS

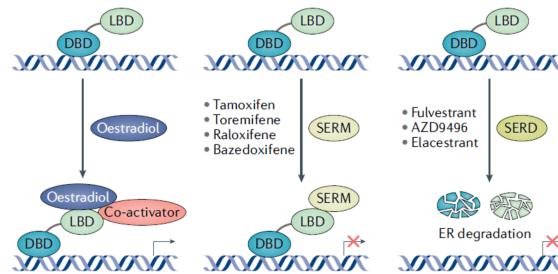


Figure 4.13: Screen Shot 2022-11-18 at 11-06-54.png

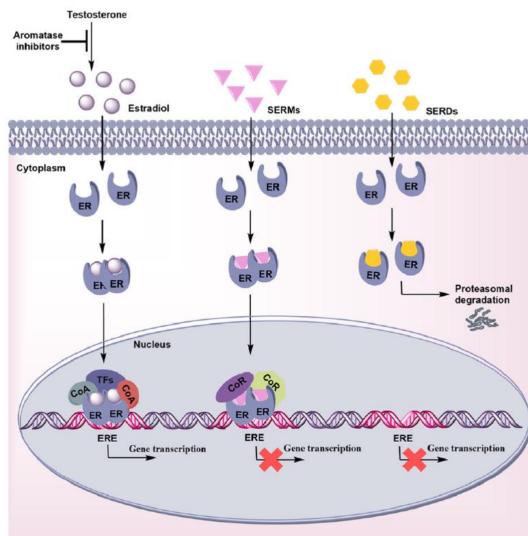


Figure 4.14: Screen Shot 2022-11-18 at 11-14-09.png

in order to deactivate ER activation.

4.8.2 ER α in breast cancer

- Nearly 75% of breast cancers are driven by ER α -mediated transcriptional activity
- Estrogen deprivation and direct antagonism of ER represent the two major strategies for therapeutic targeting of ER+ breast cancers
- More than 20% of patients develop resistance to anti-estrogens and relapse with metastatic disease

4.8.2.1 Mechanisms provoking breast cancer drug resistance

Clinical sequencing of 11 metastatic ER-positive breast cancer cases before and after therapy showed that **ESR1 mutations** were not present at an earlier stage, indicating that they were acquired after

4.8. NUCLEAR RECEPTORS

endocrine therapy. Tumors had survived estrogen targeting or deprivation treatments by acquiring ESR1 mutations.

ESR1 alterations are focused on H12 of the LBD. Analyses of 390 ER-positive breast cancers (primary tumors, before hormonal treatment) from the TCGA revealed no LBD-disrupting mutations of ESR1.

ESR1 with acquired mutations encode constitutively active proteins functioning in the absence of ligand → resistance mechanism. The mutated ESR1 variants are active in the absence of estrogen and continue to be responsive to direct ER antagonists. These mutations might not have been arisen under selective pressure of anti-estrogen treatment but rather in the context of an estrogen deprivation setting, such as treatment with aromatase inhibitors and/or oophorectomy. 67.4% of the ESR1-mutant metastatic patients had prior exposure to an aromatase inhibitor.

The presence of a somatic mutation in the LBD of ESR1 does not necessarily imply constitutive activity. The most frequent mutation is D538G. Most mutations were activating ER in the absence of estradiol. ER mutants show increased S118 and S167 steady state phosphorylation.

Fulvestrant (ICI) was able to inhibit the activity of all of the mutants which however showed significant differences in sensitivity to the drug. Y537S mutants displayed 70-fold higher IC99 than WT ER, while E380Q and S463P showed 2-fold higher IC99. Fulvestrant fully inhibited the growth of the WT-, E380Q-, and S463P expressing tumors while nearly completely inhibiting the growth of D538G tumors. The Y537S-expressing tumors, however, continued to grow in the presence of fulvestrant, albeit more slowly than in untreated controls, AZD9496 was able to completely inhibit their growth.

These findings are consistent with results from a subsequent clinical trial identifying the ER α Y537S as an acquired mutation promoting resistance to fulvestrant treatment. Patients previously progressing on endocrine therapy were enrolled for either palbociclib plus fulvestrant or placebo plus fulvestrant treatment, demonstrating an improvement in median PFS from 4.6 to 11.2 months with the addition of palbociclib to fulvestrant. Acquisition of new PIK3CA and ESR1 mutations, in particular the ESR1 Y537S mutation, in both treatment arms implicates these changes in the development of parallel mechanisms of resistance and suggest potential new avenues for treatment.

4.8.2.2 Summary

- *ESR1 gene mutations* are found in about **14% of breast cancer metastasis**
- Most ER mutations promote an activated conformation in the absence of ligand which remains permissive for ligand binding, thus a direct antagonism may be a strategy for these mutants
- Mutations behave differently in terms of ER activity and sensitivity to drugs
- Mutated-ER proteins can be inhibited by *fulvestrant* and *AZD9496*, Y537Y mutation is more resistant to fulvestrant
- Mutation of Y537 site to cysteine (C), aspartic acid (D), and asparagine (N) caused receptor activation, but to a lesser degree than did the S mutant. Hence, the level of ER activation depends on both the site of mutation and the nature of the mutant residue
- Tumor genotyping of *ESR1* mutant breast cancers also revealed recurrent alterations in the PI3K/AKT pathway, cyclin D1, and FGF receptors which will likely influence the tumor addiction to ER. Combinations of antiestrogens with inhibitors of PI3K, AKT, CDK4/6, and FGFRs or other chemotherapeutic agents can represent the best treatment

4.8. NUCLEAR RECEPTORS

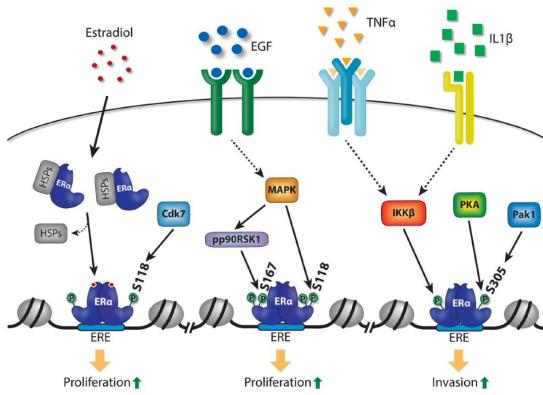


Figure 4.15: Screen Shot 2022-11-18 at 11-59-02.png

4.8.2.3 ER α transcriptional program

Multiple growth factor and cytokine signaling pathways can induce phosphorylation of ER activating the receptor in the absence of estradiol, thereby promoting cell proliferation. The independence from ER is due to over activation in downstream pathway effectors.

ER α ChIP-seq analyses of tumor samples revealed 484 ER common binding regions. ER α -chromatin binding signal intensity is higher in tumors that progress towards a poorer prognosis and ultimately metastasize. The genes within 20 kilobases from the 484 ER-binding events exhibited elevated expression in the ER+ tumours, as compared to all other genes and were higher in ER+ tumours relative to ER- tumours. ER α ChIP seq revealed differential ER-binding events between patients with good outcome and patients with poor outcome or metastases.

FOXA1 motifs were enriched in ER α binding regions of tam-resistant cells while **GATA** motifs were enriched in ER-binding events of tam-responsive cells. What is driving ER reprogramming?

Mitogen treatment of MCF7 cells (EGF, IL-6, TNF-a and IGF-I) reprogrammed ER-binding events. Distinct ER-binding profiles are associated with clinical outcome of breast cancer patients. These differential ER-binding profiles are for the most part mediated by FOXA1. Upregulation of growth factor pathways (along with ESR1 mutations and changes in cofactors levels) can influence ER binding reprogramming and the consequent change in gene expression profile fuels tumorigenesis and resistance to treatment of breast cancer cells. Since ER binding to DNA is for a good part dependent on FOXA1, targeting FOXA1, instead of ER might provide an opportunity for blocking ER transcriptional activity.

Half of all ER-binding regions overlap with a FoxA1-binding region. FOXA1 can open chromatin to ER. In addition, numerous ER-binding partners regulate ER transcriptional activity.

It was observed that transcriptional reprogramming is present in Y537S and D538G mutants. Y537S and D538G mutants promote the transcription of a unique set of genes not induced by WT ER upon estrogen stimulation.

The FOXA1 motif was not significantly enriched in the mutant-selective binding sites, suggesting that FOXA1 may be less essential for mutant-specific ER DNA binding. E2-independent ER recruitment

4.8. NUCLEAR RECEPTORS

in the presence of the Y537S and D538G mutations, with a redistribution of 39% and 49% of the ER binding events for the Y537S and D538G mutations, respectively. CDK7 silencing impacted proliferation of both WT-ER and Y537S expressing cells.

In principle, if we find ER enhancer RNA promoting specific cancer genes, we can target it and selectively inhibiting it.

Chapter 5

Transcription and genome instability

5.1 R-loops

RNA is quickly displaced as soon as it emerges from Pol II; during transcription RNA-DNA hybrid structure is required, but it is *transient*. It has been observed that RNA can re-hybridize with the template DNA forming **RNA:DNA hybrid** and **R-loop structures**. R-loop formation results from a competition between the nascent RNA and the non-template DNA strand to hybridize with the template strand.

High G density in the non-template DNA strand promotes R-loops formation. RNA:DNA hybrids rich in RNA-G/DNA-C ratio are more stable than DNA:DNA duplex of the same sequence.

Transcription unwinding results in topological stress and generation of DNA supercoils.

- Positive toroidal supercoils: dsDNA is tightly packed
- Negative toroidal supercoils: loose DNA, predisposition to RNA insertion.

Superhelical stress can favor, and be mitigated by, R-loops formation.

In highly transcribed genes we witness histone turnover, which is favored by negative supercoiling. R-loops formation inhibit nucleosomes redeposition weakening surrounding nucleosome-DNA contacts.

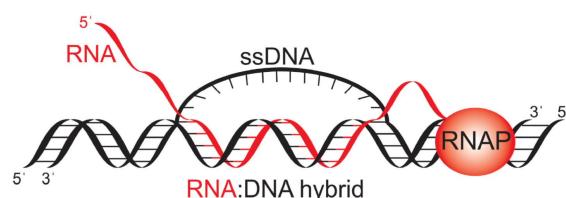


Figure 5.1: Hamperl and Cimprich, DNA Repair, 2014

5.1. R-LOOPS

G-rich sequences and negative supercoiling can promote the formation of R-loops

In addition, to win competition with the non-template strand, the stability of the sequence is important → G-quadruplexes highly increase stability.

G-quadruplexes: 4 guanines interacting by hydrogen bonds (Hoogsteen) on the same plane (planar quartet). We can find parallel G4 or antiparallel G4 (based on the orientation of guanines).

R-loops are typically formed co-transcriptionally (*cis*) but formation in *trans* has been reported and expected e.g. Cas9 pathway or ncRNA with unwinding promotion.

5.1.1 R-loop degradation

TREX complex: mediates transcript export for splicing and translation. If we remove THO subunit, we impair trex and formation of R-loops → limiting the amount of naked RNA in the cells prevents R-loops accumulation.

R-loops are also actively eliminated by specific enzymes:

- RNase H1/2 recognize and degrade R-loops
- RNA:DNA helicases e.g. Sen1/SETX

5.1.1.1 R-loop recognition and distribution

DRIP-seq: The monoclonal S9.6 antibody revealed thousands of R-loop hotspots in human genome. R-loops can be found in:

- highly transcribed genes (as rate of transcription can influence loop formation for negative torsional stress)
- telomeres: G-rich non-template strand, C-rich template strand
- ORF: proposed function at termination sites for termination factor recruitment e.g. chromatin remodeling.

! Pull down is performed after sonication, mechanic stress on DNA → bias. A good part of the signal could be coming from dsDNA. Over the years, improvements to the method were done e.g. bisulphide for T→U, only finds RNA.

Endogenous R-loop structures are also detected by **R-ChIP**: a catalytically dead RNASEH enzyme can be expressed in cells, able to recognize but not degrade. We visualize R-loops endogenously. Bias: RNASE can be recruited on paused sites, the technique can only applied in cell lines.

R-ChIP and DRIP-seq experiments reveal R-loops enrichment at promoter regions and TSS. Approximately 60% human promoters are associated with a CGI - CGI methylation generally associates with gene silencing.

R-loops form in human CpG island promoters with G-rich sequence at the non-template strand (GC skew) regulate DNA methylation.

Example of involvement of R-loops in transcription regulation.

5.1. R-LOOPS

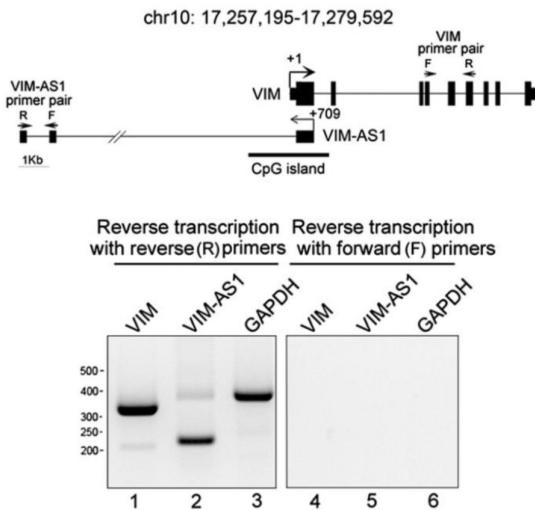


Figure 5.2: Screen Shot 2022-11-23 at 10-18-00.png

5.1.2 VIM-antisense1

VIM is an intermediate filament in mesenchymal cells (EMT process). There is a presence of a CpG island. Usually with antisense the transcription of the sense collides due to sterical hindrance (overlapping). In this case, the antisense is polyadenylated and nuclear, while the sense is cytoplasmic. The expression of the antisense is associated with down-regulation of the sense gene. Depletion of VIM-AS1 causes downregulation of the VIM gene expression caused by methylation.

The two transcripts can be visualized through FISH with different fluorophores. By combining this technique with the inhibition of transcription VIM disappears, while VIM-AS1 is more stable. This could be due to the fact that it is engaged with a R-loop structure. The genomic region between the two TSSs shows GC skew sequences.

R-loop formation is dependent on the antisense transcript. RNaseH1 overexpression inhibits VIM and VIM-AS1 expression. R-loop structures promote VIM expression by impairing nucleosome occupancy and favoring the binding of transcription factors.

Expression of antisense and R-loop associate with open chromatin

By blocking R-loops or inhibiting VIM-AS1 at FNkB TF binding sites accessibility is reduced and the recruitment of the TF at the locus is impaired.

5.1.2.1 Main findings

- loop formation by antisense RNAs at CpG island containing promoters can impair DNA methylation at these promoter sequences
- decrease nucleosome occupancy and favor binding of the transcription factor NF-kB to the promoter of Vimentin gene promoting transcription of the sense gene

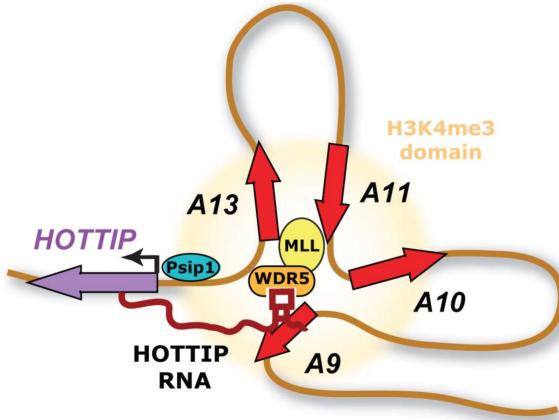


Figure 5.3: Screen Shot 2022-11-25 at 11-49-03.png

R-loops can regulate gene expression by multiple mechanisms: repressing chromatin modifiers, activating chromatin modifiers and chromatin regulating complexes. They arise naturally and have multiple physiological effects e.g. DNA repair, replication and gene expression, and are therefore tightly regulated. R-loops are generally transient and regulated by enzymatic activity; under certain conditions, R-loops become de-regulated and accumulate in cells.

5.1.3 HOTTIP-dependent R-loop formation regulates CTCF boundary activity and TAD integrity in leukemia

In vertebrates the Hox genes are located contiguously in clusters. Hox genes are expressed in a tightly regulated spatio-temporal manner during embryogenesis. They possess the homeobox domain and are divided into 4 clusters. The spatio-temporal expression is also observed in human primary fibroblast from different sites. 5C chromosome interaction highlights the presence of the *posterior domain*. Differentiated cells (distal cells from limbs) have a distinct pattern with respect to proximal cells (from lung): in distal Pol II signal and H3K4me is observed in the first HOTTIP region, while the opposite patterns are present in proximal cells (HOTAIRM1 region at 3' → lncRNA interacting with chromatin remodelling complex).

HOTTIP stands for HOXA transcript at the distal tip. HOTTIP lncRNA stimulates the transcription of HoxA genes by enforcing H3K4me3 chromatin modification.

Chromosomal looping brings HOTTIP RNA in close proximity to the HOXA genes. HOTTIP lncRNA binds to and targets WDR5-MLL complexes to the HOXA locus, leading to transcription activation. The mutual interdependence between HOTTIP RNA and WDR5-MLL creates a positive feedback loop that maintains the ON state of the locus.

HOX genes are mutated or deregulated in different cancers and play active roles in tumorigenesis. HOXA genes are expressed in hematopoietic stem cells and progenitor cells while downregulated during differentiation. Abnormal HOXA gene activation is a common feature of acute myeloid leukemia (AML). HOXA9 and HOXA10 genes are frequently aberrantly activated in AML patients. Dysregulation of HOXA genes (e.g., HOXA9) is a driving mechanism for hematopoietic deregulation and leukemogenesis. Overexpression of HOXA9 is a poor prognostic marker in leukemia patients while

5.1. R-LOOPS

its downregulation is a favorable predictor of AML patient outcome. The mechanisms regulating HOXA genes expression in AML patients is under investigation.

A transition from repressive to active chromatin is detected within the HOXA locus in AML patients. Among A7 and A9 we observe a sort of boundary, mainly regulated by insulator sequences → the CTCF binding site located between HOXA7 and HOXA9 genes may regulate the aberrant activation of the HOXA9-13 genes in AML cells.

Deletion of the CTCF binding site between HOXA7/9 genes (CBS7/9+/-) alters HOXA gene expression in AML cell lines. The homozygous deletion was incompatible with cell survival, so heterozygous was used. ChIP-seq analyses of CBS7/9+/- cells show altered chromatin structure in the HOXA9-13 domain but not in the HOXA1-7 locus, consistent with a loss of boundary function. 4C-seq data indicate decreased HOXA9 interaction with proximal genomic sites.

CBS7/9+/- is important to regulate the chromatin structure and the expression of HOXA9-13 genes in AML cells

Dysregulation of CBS7/9 boundary inhibits leukemic cell proliferation and prolongs survival time of transplanted NSG mice. In addition to the local effect on chromatin marks and structure, a significant number of dysregulated genes was observed e.g. oncogenic pathways RUNX1, SOX involved in myeloid activation.

Attenuation of the chromatin boundary disrupts the active chromatin domain and perturbs oncogenic gene expression in AML, in part by disrupting the HOXA9 oncogenic pathway. The CBS7/9 boundary located at the edge of the TAD encompassing the posterior HOXA genes establishes and maintains aberrant chromatin signatures and expression of the posterior HOXA genes to facilitate myeloid leukemogenesis. The elimination of CTCF fosters the formation of aberrant protein.

Can a normal CTCF boundary be hijacked to control oncogenic chromatin domain and transcription profiles for leukemic transformation and progression?

HOTTIP^{-/-} perturbs HOXA gene-mediated oncogenic transcription program, we observe oncogene downregulation. The KO is able to recapitulate the effect of CBS7/9 on HOXA. In wt AML cells we can identify the anterior and posterior domain, in KO cells we observe a change in the posterior domain.

HOTTIP lncRNA is aberrantly expressed in a subset of AML patients and cells. NPM1-mutated (NPM1C+) or MLL-rearranged (MLLr+) AML cases (n = 76) exhibited elevated levels of HOTTIP expression. Survival rate was inversely correlated with HOTTIP expression.

Activation of HOTTIP rescues the HOXA gene chromatin defects in the CBS7/9+/- AML cells.

HOTTIP transgenic expression in hematopoiesis leads to AML-like disease and promotes hematopoietic transcription programs.

Conclusions

- HOTTIP expression alters HOXA genes containing TADs and HOXA genes expression
- HOTTIP KO affects leukemic transcription program while HOTTIP overexpression leads to AML-like disease
- HOTTIP is aberrantly expressed in AML patients

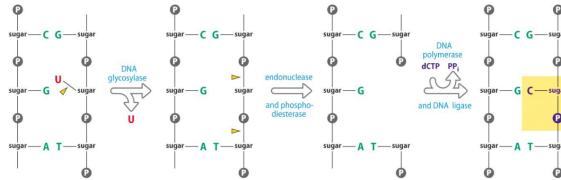


Figure 5.4: Base excision repair

5.1.4 HOTTIP lncRNA Promotes Hematopoietic Stem Cells Self-Renewal Leading to AML-like Disease in Mice

CHIRP: reminds us of ChIP for crosslinking and sonication. Instead of using antibodies, they use specific probes for RNA of interest with tilling oligos (bp with different regions of the RNA, increased yield of the pull down). At the end of the protocol we obtain RNA.

HOTTIP interactome isolated from AML cells contains CTCF/cohesin complex and R-loop-associated proteins.

HOTTIP KO resulted in a significant reduction of R loops and HOTTIP binding sites, especially when cobound to CTCF.

No signal from GRO-Seq → HOTTIP KO disrupted CTCF and cohesin TAD boundaries containing AML oncogenes such as β -catenin (CTNNB1). At these boundary sites HOTTIP RNAs form R-loops in trans. HOTTIP-mediated R-loop formation directly contributes to CTCF boundary activity.

R-loops disruption at the CBS-u2 boundary site impairs proliferation of AML cells.

5.2 DNA damage

Single stranded DNAs are more susceptible to chemical modifications, strand breaks, mutagen sensitivity and secondary structure than dsDNAs. Also dsDNA is undergoing spontaneous chemical modifications. Basically, any covalent bond in DNA molecule can be attacked by a water molecule. It has been estimated that there are thousands of spontaneous lesions in our cells e.g. depurination ($G \rightarrow$ sugar phosphate) or deamination ($C \rightarrow U$). Spurious methylation can occur in bases different from cytosine; when a RNA pol II passes through modified bases, we can observe a stall or the insertion of an erroneous base, leading to mutations. The addition of chemical groups occurs when bases react with oxygen species e.g. guanine + adduct radical gives 8-hydroxyguanine.

Base excision repair (BER) pathway eliminates modified bases or depurinated nucleotides.

The chemical modification rate in oligonucleotides depends on cation concentration, pH, and other experimental conditions. Roughly 5% of the genome can be occupied by R-loops.

5.3 R loop modified bases

Modified bases within R-loops that are not properly repaired form nicks and breaks (this step requires dsDNA) → replication and transcription stops. R-loops accumulation associates with genome

5.3. R LOOP MODIFIED BASES

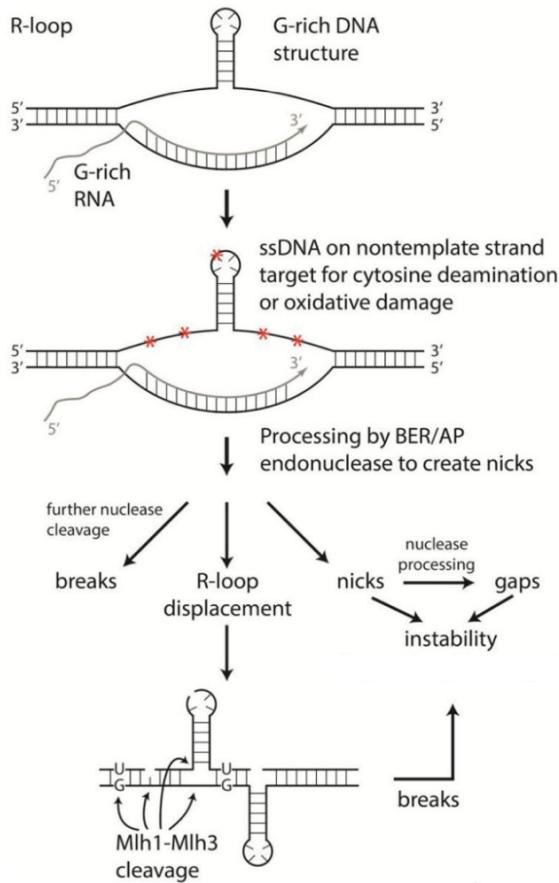


Figure 5.5: R loop repair

instability due to the spontaneous base modifications occurring at ssDNA and the ensuing processing events leading to nicks and breaks.

Thanks to R-loop displacement, BER can excite modified bases; this provokes a misaligned dsDNA, leading to breakages - especially if loops are hit by modifications - and genome instability.

R-loops can lead to nicks and breaks (DSBs) also due to the activity of the AID/APOBEC cytidine deaminases (catalyzed modifications). **Activation-induced cytidine deaminase** (AID) promotes somatic hypermutation and class switch recombination of immunoglobulin (Ig) genes in germinal center (GC) B cells. AID off-target activity has been implicated in malignant transformation of GC-derived B cell lymphomas.

C→U is meant to be, as cells want to induce mutation in a random manner to achieve hypervariability. In particular, several cancers are characterized by APOBEC signatures, which are a key source of mutation e.g. chromosome instability in early breast and lung cancer evolution. Not only *spontaneous* base modifications but also cytosine deamination catalyzed by **APOBECs** can lead to nicks and breaks within R-loops structures leading to genome instability.

5.3. R LOOP MODIFIED BASES

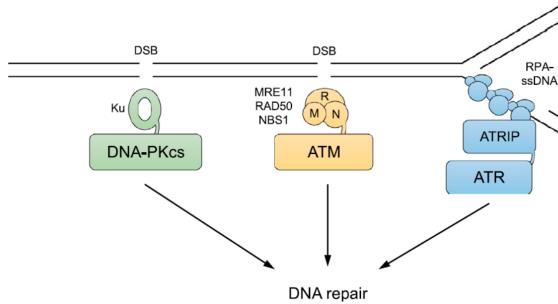


Figure 5.6: Apical kinases

5.3.1 Apical kinases

Cells have evolved the **DNA damage response (DDR)** in order to combat threats posed by DNA damage. Breaks are sensed by apical kinases: **ATM**, **ATR** and **DNA-PKcs**. Phosphorylated serine or threonine residues followed by glutamine (S/T-Q) S/T-Q sites are present in ATM, ATR and DNA-PKcs for autophosphorylation. They all contain a kinase domain at the C-terminal. Cancer cells and immunodeficient cells (mutated in these genes) are more susceptible to radiation, since the recognition of ds breaks is impaired → radiotherapy would result in important side effects.

ATM, ATR and DNA-PKcs require specific co-factors for their recruitment to damaged DNA. ATM is linked to a trimeric complex (MRN), where MRE11 is the endo/exo nuclease, RAD50 recognizes dsDNA and NBS1 is full of PPinteracting domain (ATM recruitment). ATR is recruited to dsbreaks when they are processed, so ssDNA bound by RPA (aspecific marker), bound by ATRIP co-factor.

They are also involved in cell cycle control, DNA replication, transcriptional regulation and RNA metabolism. It orchestrates real cell response.

DNA-PKcs promotes NHEJ of DSBs: DNA-PKcs phosphorylation allows the recruitment of downstream NHEJ core factors, leading to DNA-end ligation by LIG4. This pathway is fast and can be used in all cell cycle phases.

ATM activation promotes a signaling cascade on damaged chromatin. MRN recruits ATM, which can phosphorylate hundreds of different targets, which are themselves kinases → kinase cascade. The first event is the phosphorylation of H2AX, leading to gamma H2AX marker for DNA damage. MDC1 is recruited and phosphorylation, additional MRN recruitment and feed-forward mechanism. Gamma H2AX can spread megabases, self sustaining cycle enhancing the reaction to ds break. TIP60 activates ATM through acetylation.

53BP1 is an effector of DNA damage response (reaction of the cell to ds break, enabling sensing) and DNA repair mechanism (fix the break) → NHEJ activation. Among the ATM targets we have BRCA1 (scaffold molecule and ubiquitinligase) and CtIP (endonuclease), which are phosphorylated → signal for the cell to activate homologous recombination (requires the sister chromatid, S phase only).

BRCA1 and 53BP1 determine the choice of repair mechanism between NHEJ and HR. The decision between NHEJ and HR is also determined by several factors including cell cycle phase, chromatin state, genetics. NHEJ is “error prone” (but not in the classic pathway, alternative one), while homologous recombination is error free.

5.3. R LOOP MODIFIED BASES

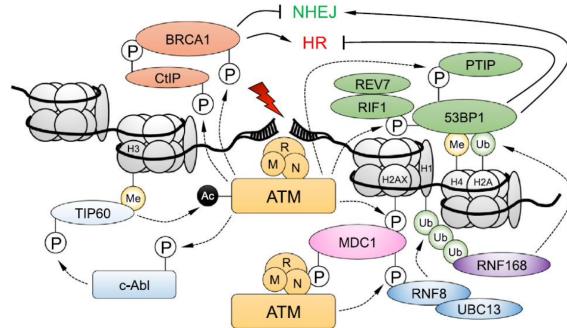


Figure 5.7: Blackford and Jackson, *Mol Cell*, 2017

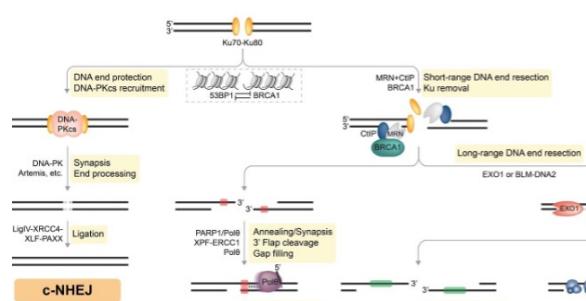


Figure 5.8: Trenner and Satori, *Frontiers in Oncology* 2019

5.3. R LOOP MODIFIED BASES

DNA damage response integrates regulation of the cell cycle, which is blocked in order to prevent mitosis. It is necessary to allow repair mechanism to act before resuming the cell cycle. Cell cycle is regulated by CHK2 (kinase phosphorylating CDK29) and p53.

The key signal for the cell in the context of repair comes from DNA ends; telomere mask chromosome ends from being recognized as double-strand breaks.

DNA damage response (DDR) involves DNA lesion recognition followed by a signaling cascade to promote DNA repair. The effectors lead to transient checkpoint, cellular senescence and apoptosis.

The roles of **PARP1** (poly(ADP-ribose) polymerase 1) in detection and repair of DNA double-strand breaks are several. It is recruited to ds break and adds poly(ADP-ribose) chains. The recruitment of covalently and non-covalently modified proteins to site of DNA damage allows for repair of ssDNA nicks and breaks, sd breaks and chromatin modifications. PARP is also associated with ATM itself and is involved in HR, cNHEJ and aNHEJ. Although it is not the only activating mechanism of BRCA, it supports it to the break and lead to strand invasion and resolution.

Summarizing, defects in DSB repair e.g. loss of Xrcc4, LIG4, BRCA1 lead to chromosome instability and accumulation of mutation → defective checkpoints, cancer.

5.3.2 Topoisomerases

Programmed ssDNA and dsDNA breaks occur during transcription. As we have mentioned, positive and negative torsional stress generate forces opposing the direction of pol II. Topoisomerase cleavage complexes (TOPcc) assemble at sites of topological stress. They are tightly regulated to minimize deleterious cleavage complexes - their activity is impeded by nucleosomes. They are recruited to chromatin via interaction with chromatin remodelling complexes (SWI/SNF), histone chaperones (FACT) and helicase enzymes (WRN).

TOP1 triggers SSBs or nicks, TOP2 induces DSBs. Both perform breakage in a controlled manner and once torsion is released they can exert ligation activity. During transcription TOP1 and TOP2 enzymes relieve positive supercoils while TOP1 and TOP3 counteract negative supercoils.

5.3.2.1 A topoisomerase IIb-mediated dsDNA break required for Regulated Transcription

Biotin-dUTP labeling by terminal deoxy .. → only if a break is present. Amplification of the promoter region and not of the ORF, signaling the presence of a break in the promoter leading to the recruitment of PARP. However the break is transient, after 10 minutes it disappears.

DSBs are induced transiently on the pS2 promoter upon estrogen treatment. Transcriptional activation associates with transient and local formation of DNA double strand breaks. TopoII β -mediated DSBs at the pS2 promoter is required for pS2 transcription. If we block topoisomerases, machinery for transcription is not recruited. The presence of DNAPK persists for a while after the resolution of the break, this does only mean that the signal is spreading.

TopoIIb/PARP-1 complex in nuclear receptor-mediated gene regulation TopoII mediated DSB and in PARP-1 activity serves as mechanism for gene transcription upon ligand or signal-dependent stimulation. PARP1 activity promotes removal of histone H1 from the ERE-containing nucleosome supporting transcription initiation.

5.3. R LOOP MODIFIED BASES

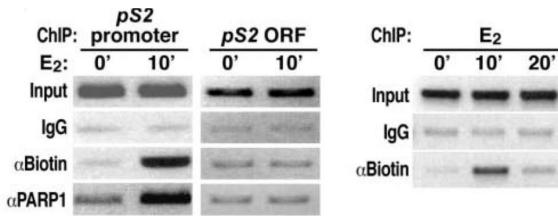


Figure 5.9: Screen Shot 2022-11-30 at 10-03-17.png

5.3.2.2 BRCA1

Sometimes, pathological topoisomerase 2 sites are observed. OP2 is required for ER-mediated transcription. Occasional estrogen-induced pathological TOP2 occur with DSBs covalently associated with the enzyme. BRCA1 participates in resolving these structures promoting genome integrity. In the absence of BRCA1 estrogen exerts a genotoxic effect with accumulation of DSBs during cell divisions.

The roles of BRCA1 in the maintenance of genome integrity and regulation of transcription may be the key of its tissue specific tumor suppressor function. BRCA1 has an ubiquitinylase domain (RING) and a huge number of interacting proteins. BRCA1 drives a transcriptional program supporting differentiation of luminal progenitors to mature luminal cells. BRCA1 interaction inhibits ER α activity directly and by promoting mono-ubiquitination of ER α . BRCA1 mutant progenitors are aberrantly proliferative and defective in differentiation giving rise to tumors of luminal origin. Cells with LOH in BRCA1 usually show basal-like tumor activity.

5.3.2.3 Stark et al.

24h estrogen treatment activates DNA damage response pathways and associates with DSBs in MCF7 cells. **Comet assay:** cells are lysate, nuclei are inserted in an agarose pad and placed under an electrophoretic chamber (no fragmentation): if there are ds breaks it will move a little bit, we will observe a “comet” signal. In 24 hours cells can cycle and perhaps this is not the same break as topoisomerase II. Long term estrogen treatment results in greater DDR activation than short term treatments.

DNA damage induced by long term estrogen treatment is replication dependent, the more the cell proliferate, the more the DNA damage response is activated. If a Cdc7 inhibitor, DNA damage response activation is impacted. Flavopiridol (inhibits Cdk9) results in Pol II elongation impairment, but does not affect replication.

Estrogen induces R-loops formation at ER responsive genes. The induction of R loop was depending on ER target genes. Gene responsive to estrogen are enriched in genomic rearrangements in breast tumors. Somatic mutation data from whole genome seq of 560 breast tumors determining whether the mutation sites are enriched in estrogen responsive loci.

Estrogen induced R-loops colocalize with DNA damage markers on chromatin. Proximity ligation assays suggest that estrogen induced R-loops occur on chromatin marked by DNA damage.

5.3. R LOOP MODIFIED BASES

RNaseH expression reduces estrogen induced DNA damage. RNaseH expression reduces estrogen induced DSBs. R-loops may be involved in DNA damage induction DSBs formation in breast cancer cells upon estrogen treatment.

Summary

- Estrogen treatment induces replication and transcription dependent DNA damage
- DNA damage and DSBs induced by estrogen stimulation are linked to R-loop formation at ER responsive genes
- Breast cancer rearrangements are enriched at estrogen responsive loci where R-loops are detected upon estrogen treatment
- Many DSBs that accumulate upon estrogen treatment are R-loops dependent
- Estrogen stimulation leads to genome instability through R-loops formation and in a DNA replication dependent manner
- Also highlights an «alternative» mechanism by which a transcriptional program plays a role in genome stability in cancer

Chapter 6

Role of transcription in DNA damage response and repair

The source of activation of DDR pathway is replicative stress i.e. dysregulation of replication machinery and involvement of transcription, which can collide with DNA replication processes, in particular when R-loops are present. Cells need to react quickly to this source of instability, as DNA damage can lead to single or double DNA breakage.

Until recently, only proteins were presumed to be involved in this signalling cascade.

In the figure we visualize human normal fibroblasts (HNF) hit by an ionizing radiation - reproducing X-ray diagnostic procedure. This leads to DNA break and activation of DDR. The experiment was performed with siDICER (targeting and degrading dsDNA) → impaired activity apart from γ H2AX, which is phosphorylated quickly. siDROSHA also shows a similar pattern.

6.1 Biogenesis of canonical miRNA

These enzymes are involved in the maturation process of miRNAs. The pri-miRNA is shaped in a loop structure, Drosophila binds and then pre-miRNA is bound to exportin for the export in the cytoplasm. Once outside, Dicer recognized the stem loop and cuts RNA right at the loop to obtain 20-25 nt sequence.

DICER and DROSHA inactivation impairs DDR loci formation in irradiated cells. The same occurs in mutant Dicer (DICERexon5) DDR positive cells.

RFP 5' UTR for endogenously expressed RNA. Under normal condition the fluorescence is not seen; if we keep the cells under the x-ray, we observe miR126 staining. If Dicer is depleted, siDICER will exhibit increased signal. No 53BP1 as the pathway is impaired. Lastly, they deplete proteins involved in transcription inhibition. 53Bp1 is detected if cells are irradiated.

This suggests that the effect of DICER and DROSHA is independent from the canonical miRNA-mediated translational repression mechanisms. The knockdown impairs G1/S checkpoint; OIS (oncogene induced senescence) cells further supports that these enzymes are important for DDR.

6.1. BIOGENESIS OF CANONICAL MIRNA

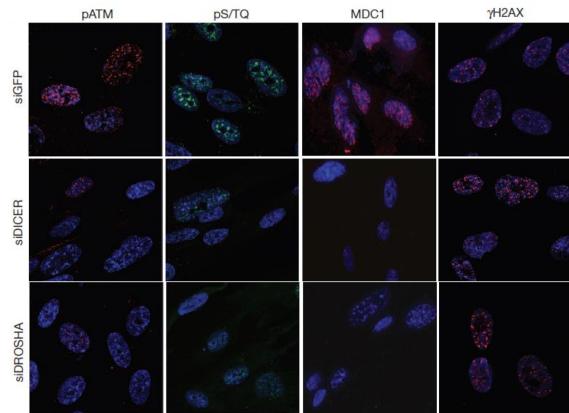


Figure 6.1: Francia *et al.*, *Nature* 2012

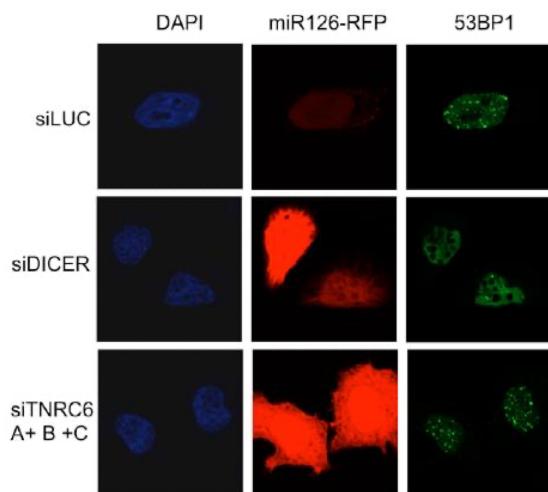


Figure 6.2: Francia *et al.*, *Nature* 2012

6.2. DICER AND DROSH REGULATION OF DDR AND CHECKPOINT ENFORCEMENT

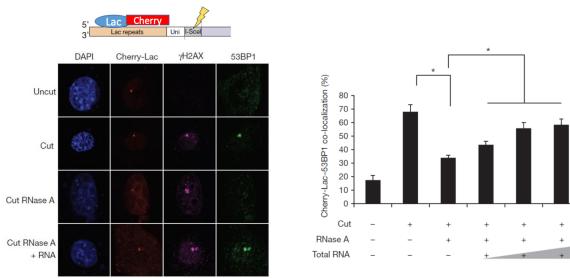


Figure 6.3: Francia *et al.*, *Nature* 2012

6.2 DICER and DROSH regulation of DDR and checkpoint enforcement

Cells treated with mild detergent promoting permeabilization of the membrane; add RNase to degrade RNAs, marker of DDR will not be detected. When irradiating cells, this leads to DDR inhibition. If before fixing cells introduce RNA after degradation, this enables the rescue of the effect. DDR activation is controlled by small RNA species of 20-35 nt.

Since irradiation leads to ds break in a random fashion, to gain more insight in the nature of RNAs and the nature of the breaks they used NIH 2/4 cells, which have a construct containing Lac repeats and IScell1 restriction enzyme binding site. IS1 cuts 18 nt binding sites - the binding site does not contain this site. A DDR focus generated on a defined DSB can disassemble and reassemble in an RNA-dependent manner.

Site-specific DDR focus formation is RNase A sensitive. Only RNA from the cut NIH2/4 cells can reassemble DDR focus formation. Is the RNA generated from the LacO-SceI cassette upon cut responsible to DDR foci formation?

Francia *et al.*, *Nature* 2012

Deep-sequencing of libraries generated from short <200 nucleotides nuclear RNAs revealed short transcripts arising from the exogenous locus (LacO-SceI cassette).

Chemically synthesized small RNAs are sufficient to restore DDR focus formation in RNase A-treated cells in a sequence-specific manner

6.3 DDRNAs

DDRNAs are DICER- and DROSHA-dependent products with the sequence of the damaged site. DDRNAs act differently from canonical miRNAs. For instance, microinjected DDRNAs (with complementary sequence to I-SceI binding site) localize to the DNA damage site. Lac+Tet are used as are repetitive elements, but in principle any sequence is fine.

Sequence-specific localization of DDRNAs at DNA damage sites is functional and transcription-dependent. The RNAs co-localize by the ds-breaks complementary to their sequence and manage to do so even in the absence of Drosha and DICER → transcription-dependent manner, PolII transcription at DSBs is required for DDRNAs localization to DSBs and DDR foci formation.

6.3. DDRNAs

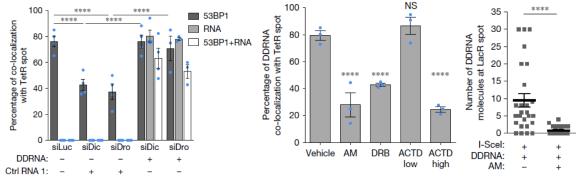


Figure 6.4: Michelini *et al.*, *Nature Cell Biology* 2017

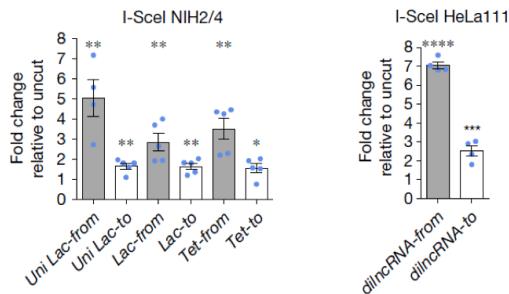
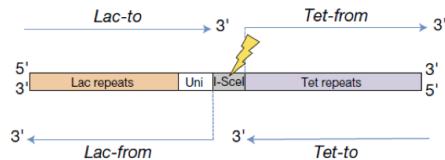


Figure 6.5: Michelini *et al.*, *Nature Cell Biology* 2017

DDRNAs localize to their homologous damaged site where they stimulate DDR focus formation in an RNA polII-dependent manner.

The investigation of transcription around DSB by RNA-FISH and RT-qPCR analysis allows us to understand the direction in which transcription is occurring. Once there is a cut, mainly *from* (outward) transcripts are visualized, but also some *to* (inward) transcript. RT-qPCR requires an amplicon of ?

dilncRNAs (damage induced lncRNAs) are transcribed by polII from DSBs, localize at DSBs where they are processed by Drosha and Dicer to generate DDRNAs. dilncRNA and DDRNAs induction are early DSB signals that together with γ H2AX may nucleate DDR focus formation.

ChIP revealed that RNAPII, RNAPII pSer5 and RNAPII pSer2 are enriched at the DSB following cut induction. It was observed that the transcription starts exactly from the extremity of the DSB, somehow Pol II must find the exact site → we said that it must be recruited at the promoter, how the hell is it activating transcription at this site?

The MRN complex might lead to the recruitment of RNA PII. Indeed, the MRN complex binds to RNAPII following irradiation (DNA damage) and is necessary for RNAPII transcription at DSBs.

RNAPII transcription is further necessary for DDR focus formation and DNA repair.

6.4. RNA POL II RECRUITMENT AT DSBS

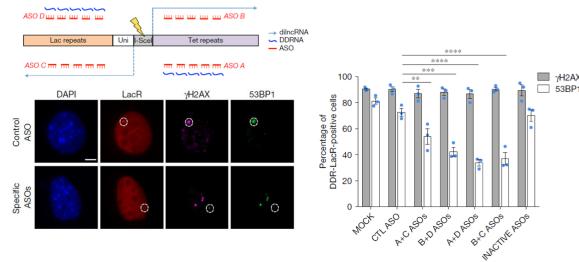


Figure 6.6: Michelini *et al.*, *Nature Cell Biology* 2017

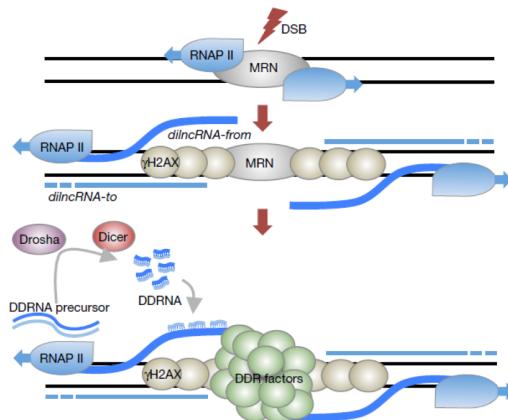


Figure 6.7: Michelini *et al.*, *Nature Cell Biology* 2017

6.3.1 DDRNAs and dlnrcRNAs regulate DDR and DNA repair

53BP1 interacts with DDRNAs and dlnrcRNAs. DDRNAs and dlnrcRNAs might act as scaffolds and stabilizers of the complex in DDR and DNA repair.

Antisense oligonucleotide gene knock-down: no phosphodiesteric bond, phosphotionate bond. This chemistry allows a stabler interaction and avoids endogenous enzyme degradation. Cannot be used with mutated ?.

ASOs prevents dlnrcRNA-DDRNa interaction, which affects 53BP1 focus formation.

The extremities of DSBs act as transcriptional promoters regardless of the genomic location, through MRN-PolII interaction. Damage-induced transcription represents one of the earliest events following DSBs generation, together with γHSAX. DSBs-induced transcription nucleates DDR foci formation at least in part due to DDRNAs-53BP1 interaction and is important to DSBs repair.

6.4 RNA Pol II recruitment at DSBs

ChIP shows that 2,000 bp or more downstream we do not have a signal from relevant TFs in RNAPolII recruitment.

6.4. RNA POL II RECRUITMENT AT DSBS

Stochastic optical reconstruction microscopy (STORM) reveals colocalization between transcription effectors and γ H2AX in NCS treated cells. NCS is a drug promoting DNA breaks randomly.

Enrichment of the indicated transcription effectors and MRN at a site-specific DSB in the indicated HeLa Kd cells. Transcription repression by multiple means reduces DDR signaling.

Main findings of the study

- The MRN complex recognizes the extremities of DSBs and recruits the PIC, Mediator and CDK9 to promote RNA polII transcription
- Inactivation of PIC and inhibition of transcription impair DDR signaling and repair
- DSBs act as sites of sequence- and promoter- independent recruitment of the transcriptional machinery
- The function of transcription in the DDR signaling is dependent on dilncRNA synthesis
- γ H2AX acts as docking site for the recruitment of DDR factors at DSBs, dilncRNAs and DDRRNAs can act in promoting and stabilizing these complexes at DSBs.

Chapter 7

Telomeres and telomere transcription in cancer

7.1 What are telomeres?

A DNA double strand break (DSB) generates new chromosome ends, which are structurally similar to a telomere. But what is so special about the extremities of our chromosomes?

Telomeric DNA is made of *telomeric repeats*, 2-20kb in length and forming GC rich sequences (seen when talking about R-loops). They depend on the current cell cycle phase. A few nucleotides before telomeres we find *subtelomeric repeats* ("patchwork" sequences duplicated during evolution, devoid of genes and less conserved).

Study from Barbara McClintock: she treated chromosomes with X-Rays and she observed large chromosomal aberrations, since breaks can be repaired by the fusion among different chromosomes. However, she noticed that such fusions never occurred between extremities. Somehow the physiologic ends of chromosomes were protected from these fusions. Even if chromosomes end with the exact same sequence, there are no identical extremities, as the t-loop can invade in different positions.

7.2 Telomeres mask chromosome ends from being recognized as double-strand breaks

Telomeres form t-loops: mammalian telomeres end in a large loop. A T-loop does not form by itself, but some proteins are involved. In particular, TRF2 dimer facilitates the strand invasion - in fact telomeres t-loops are TRF2 dependent.

Telomeric DNA is coated with telomere binding proteins → *shelterin complex*: It contains: TRF1 and TRF2 (dsBP), POT1 (ssBP), RAP1, TIN2 and TPP1 (bridging). The complex allows dsbp and ssbp to interact. We know that there are also subcomplexes on telomeres.

What is the function of this complex? Expression of a dominant negative form of TRF2 (TRF2 B M), which is unable to bind telomeres, destabilizes the shelterin complex. By removing the basic

7.3. TELOMERE DYSFUNCTION IN A P53 AND RB MUTANT SETTING DOES NOT INDUCE CELL CYCLE BLOCK

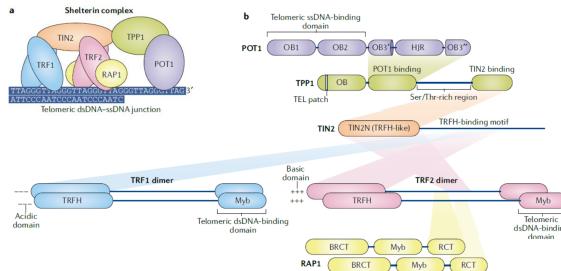


Figure 7.1: Lim and Cech, Nature Reviews Mol Cell Biol, 2021

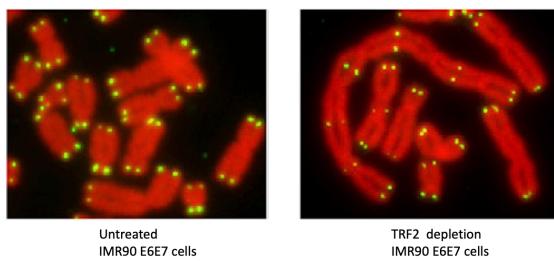


Figure 7.2: Cesare et al., Mol Cell 2013

and main domain of the stabilizing complex, the complex is disassembled. Telomere dysfunction results in activation of DNA damage response at telomeres. Cells stop dividing with this treatment, if the DDR pathway works correctly. The impaired telomere structure induces DDR at chromosome ends leading to dysfunctional telomeres.

TRF2 inhibits the recruitment of ATM through the formation of the t-loop and then there's POT1, which represses the ATR pathway. POT1 can displace RPA, which is not recruited and cannot interact with ssDNA.

7.3 Telomere dysfunction in a p53 and Rb mutant setting does not induce cell cycle block

Dysfunction of telomeres induces cell apoptosis through the usual P53- and ATM Dependent pathways. Problem: in cancer the apoptosis pathways are often impaired.

IMR90 (Human Lung fibroblast), cell line HT1080 (sarcoma cell line) → repress Cdk6, activate Cdk1. P53/Rb mutant cells continue dividing despite the presence of dysfunctional telomeres! (γ H2X)

Telomeres detected by DNA FISH on chromosome spread nicely co-localize at chromosome ends in untreated IMR90 E6E7 cells. If we apply TRF2 depletion, we observe something different.

While the DNA damage response is not active, DNA repair pathways are still active: cells continue proliferating and activating repair mechanisms (in particular NHEJ) for what is perceived as a DSB. This results in the fusion of chromosome ends → source of massive genome instability.

7.4. END OF REPLICATION PROBLEM

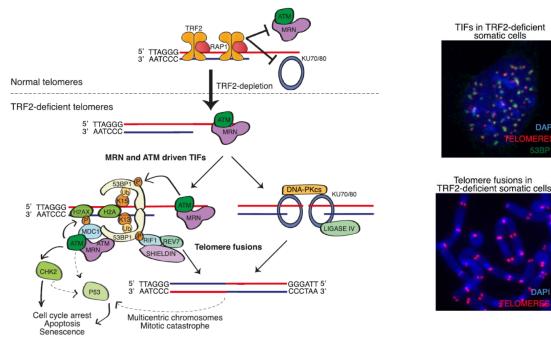


Figure 7.3: Screen Shot 2022-12-15 at 17-49-51.png

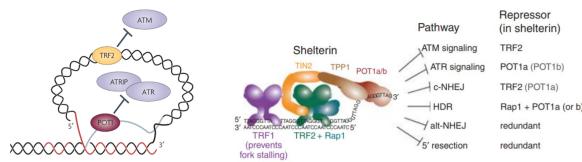


Figure 7.4: Doksan & de Lange. *Cold Spring Harbor Press*. 2014 Nature Reviews Cancer, 2008

Dysfunctional telomeres in a p53/Rb mutant background associate with genome instability.

The shelterin proteins inhibit DNA damage response and DNA repair mechanisms at chromosome ends. A key function of telomeres is to inhibit DDR and DNA repair pathways at chromosome ends. By blocking shelterin proteins, we reach *senescence* (does not allow cell division, incompatible with survival) or apoptosis, or chromosome fusions.

A few points to remember:

1. Chromosome ends are capped by structures called telomeres
2. Telomeres are nucleoprotein complexes protecting the extremities of chromosomes from being recognized as DSBs (**The end protection problem**)
3. Telomere function depends on the proper telomere structure and telomere binding proteins
4. Altered telomere structure leads to dysfunctional telomeres which are recognized as DSBs and trigger a DDR leading to senescence or apoptosis
5. If DDR pathway is impaired now genome instability arises

7.4 End of replication problem

“The end replication problem” (eukaryotes have linear chromosomes) : chromosome ends are not fully copied during DNA replication . Also, the 5' end is shorter than the lagging strand (because the template is short). In addition, in the lagging strand, the RNA primers do not anneal exactly at the last nucleosome, leading to uncertainty.

Chromosome ends shorten at every cell division: the sooner it replicates, the shorter the chromosome end will become, due to an intrinsic limitation of the DNA replication machinery.

7.4. END OF REPLICATION PROBLEM

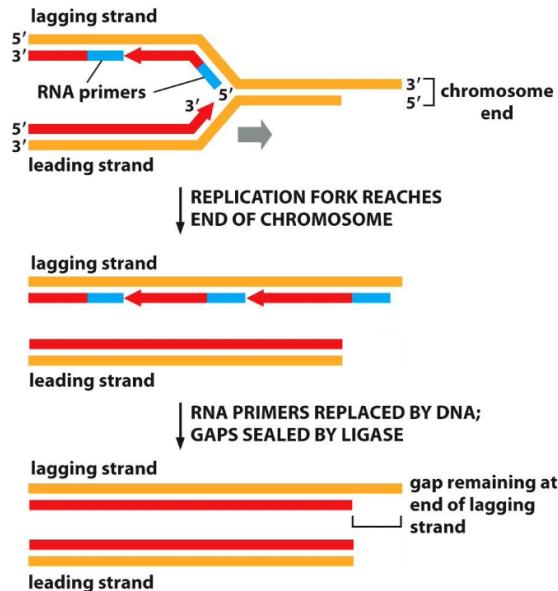


Figure 7.5: Screen Shot 2022-12-15 at 18-00-30.png

Experimentally, we can check telomere length with:

- **Southern blot** analyses: genomic DNA is digested with capping enzymes, which do not cut inside telomeric repeats → full length telomeric sequences. Due to the fact that they are repetitive and replication is not complete, the first thing that we can observe is that telomere lengths are heterogenous, every telomere has a different length (stochastic).
- **DNA-FISH**: usage of fluorescent probes binding is a sequence specific manner, the intensity of the signal is proportional on the quantity. The technique is usually applied during metaphase. If we were to do this on a population of fibroblasts, we could appreciate also the fact that different cells have different telomeres lengths.

Telomere repeats are lost with age in human leukocytes: we are born with a pool of telomeres of a specific size, which will shorten over time. But the shortening is not linear: 0-1 shortening of 400-500 hundred of bases per year and slows down. The shortening rates are not even the same for every cells! There's also a sex bias: females have on average longer telomeres both at birth and throughout life.

Telomeres shorten at each replication cycle up until a threshold, when DNA breaks are introduced. In human stem cells, this results in a low self-renewal capacity.

7.4.1 Hayflick limit

Discovery by Hayflick in 1961 (Hayflick limit): after a logarithmic growth, cells reach a plateau. There exists a limited population number that human cells can go through. The experiment involved

7.4. END OF REPLICATION PROBLEM

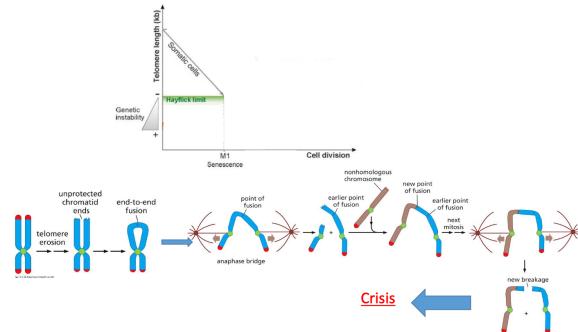


Figure 7.6: Stewart & Weinberg. Ann Rev Cell Dev Biol, 2006

fetal stem cells, with lung cells being the best ones at diving. Inverse correlation of age of the donor with max population doubling was observed. **Telomere shortening associates with decreased proliferative capacity of cells.**

- Telomere length correlates with the proliferative capacity of cells = our biological clock.
- Cells with short telomeres enter senescence earlier than cells with longer telomeres.

Subsequent ChIP-on-chip experiments confirmed the presence of DNA damage at the ends of chromosomes in senescent cells. Also, by removing repairing pathways, it was found that cell cycle arrest in senescent cells depends on DNA damage response pathways.

7.4.2 Triggering senescence

It's not actually the length of the telomeres that induces *senescence*, but their *structure*. If we find a way to promote t-loop formation, we can postpone senescence.

Study: TRF2 increased expression on telomeres promotes telomeres shortening with a higher shortening rate - probably because of a more stable t-loop, which impairs replication. Since telomeres are shorter, we expect that they enter early senescence, but this is not observed → senescence entrance does not depend on telomere length. It is the altered structure triggering the activation.

Somatic cells will normally grow and reach the Hayflick limit, triggering senescence. However, if we remove p53/Rb, we reach a M2 crisis, with increased genetic instability. The cells cannot sustain so much genome instability, so they die because of the crisis. Replicative senescence and crisis are two proliferative barriers controlled by telomere deprotection.

It is thought that there are intermediate states of telomere length, long enough to activate DNA damage response (sensed by cells) but not to fusion (not short enough to cause disruptive events, uncapped). Intermediate states are still kind of able to recruit the sheltering complex.

Replacitive senescence and crisis are independent cellular programs induced by distinct telomere states and associated with distinct biological outcomes.

7.5 Cell crisis

It is not yet clear how a cell crisis occurs (no p53 dependent death). When cells approach crisis, by bypassing senescence, the time of mitosis is extremely elongated: *mitotic arrest* (from 30mins to hours).

During the **mitotic arrest**, we are at the spindle assembly checkpoint, chromosomes are not properly attached by the two opposite spindles. Due to the genomic instability provoked by short telomeres, mitotic arrest amplifies telomeric deprotection, fostering the crisis.

The way cells die is only now being discovered: during the prolonged mitotic arrest exposed DNA activates DNA sensing mechanisms, triggering a particular kind of **autophagic cell death**. Telomeric DNA damage generates cytosolic DNA species with a fragile nuclear envelopes that undergo spontaneous disruption. The cytosolic chromatin fragments activate the cGAS-STING pathway and engage the autophagy machinery, which is an integral component of the tumor suppressive crisis mechanism.

7.6 Telomerase enzyme

Stem cells need to undergo hundreds of cell divisions without stressing out. How can they do it?

In 1995, Greider and Blackburn managed to identify a specific telomere terminal transferase activity in *Teatrahymena* (thermophile). In thermophiles, there are a lot of very little chromosomes, with many telomeres. When they incubated oligonucleotides with the sequence of the telomeres, they observed elongated. An enzyme goes and duplicates it also adding six more nucleotides. This could not be done with a regular RNA polymerase, there is no template: the enzyme is perceiving and elongating the sequence.

They found out that there exists an enzyme which also provides for the template.

In addition to the shelterin complex, also telomerase enzyme is recruited on telomeric DNA. The telomerase RNA (TR) acts as a scaffold module for telomerase assembly. The template is 11 bases long. The RNA feature the presence of H-box domain, aiding in the recruitment of TCAB1.

Just after birth, telomerase activity drops. Some telomerase activity can only be found in germ cell production (testicles). Telomerase elongates telomeres in a stepwise manner, processive enzyme. Once it is recruited, it is able to add 6-mers (not only 6 nts).

We also need to generate the second strand of telomeric DNA, since the sheltering complex is recruited on dbDNA. Almost all eukaryotes use telomerase to elongate telomeres.

7.6.1 Usage of telomerase

“Creation of human tumor cells with defined genetic elements”: overexpressed hTERT to make up for telomere shortening, immortalization (inhibition of p53 and Rb).

Mutations in TERT promoter represent the most frequent noncoding mutations in cancer.

A DNA double strand break (DSB) generates new chromosome ends structurally similar to a telomere.

7.6. TELOMERASE ENZYME

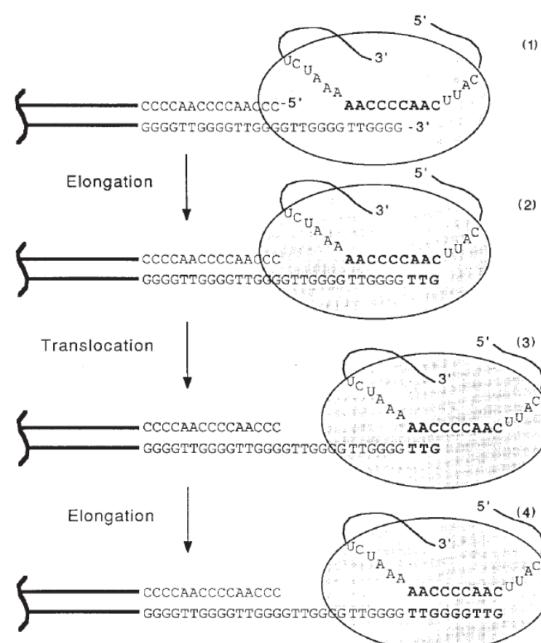


Figure 7.7: Screen Shot 2022-12-15 at 22-58-03.png

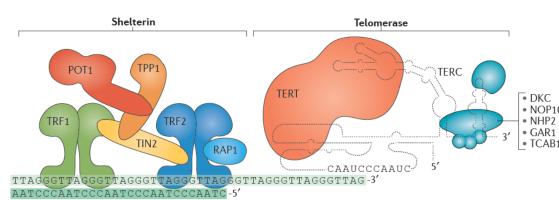


Figure 7.8: Screen Shot 2022-12-15 at 22-58-28.png

Chapter 8

Journal Clubs

8.1 Enhancers are activated by p300/CBP activity-dependent PIC assembly, RNAPII recruitment, and pause release

E1A binding protein p300/CREB-binding protein is a *transcriptional co-activator* and *histone acetyltransferase*. p300/CBP activity has been directly implicated in enhancer activation through acetylation (H3K27ac hallmark). p300/CBP-catalyzed acetylation promotes the recruitment of BRD4, and therefore the release of promoter-proximal-paused RNAPII. p300/CBP can regulate pro-proliferative pathways.

The aim of this article is to investigate the function and mechanisms of p300/CBP-catalyzed acetylation in dynamic enhancer activation.

To investigate the effects of p300/CBP inhibition on enhancers, SILAC (Stable isotope labeling by amino acids in cell culture) can be performed to obtain a relative quantification of A-485-induced acetylome and proteome changes. A-485 is a highly selective inhibitor of p300/CBP catalytic activity.

Time-resolved nascent transcription analyses in A-485-treated ESCs [5-ethynyluridine sequencing (EU-Seq)] revealed that p300/CBP inhibition causes rapid enhancer deactivation. In addition, p300/CBP inhibition selectively downregulates cell-type-specific (CTS) genes.

TSA, a KDAC inhibitor, is able to fully recover acetylation levels after inhibition. TSA-induced transcription changes are not affected by A-485 co-treatment.

p300/CBP inhibition does not impair TF and p300 binding:

- increased p300 binding at TSS inversely correlates with reduced expression of downstream genes
- p300/CBP activity appears to contribute to its dynamic dissociation from chromatin
- chromatin binding of TFs and p300/CBP per se is not sufficient to activate enhancers

To investigate whether A-485-induced transcription inhibition results from a change in chromatin accessibility, ATAC-Seq can be applied to detect chromatin accessibility levels. It was observed that

8.2. SPT5 STABILIZATION OF PROMOTER - PROXIMAL RNA POLYMERASE II

p300/CBP inhibition reduces chromatin accessibility, hence p300/CBP-catalyzed acetylation and/or ongoing transcription likely contributes to chromatin accessibility and rapid transcription inhibition by A-485 is unlikely due to a direct consequence of altered chromatin accessibility.

p300/CBP functions through BET-BRD-dependent and -independent mechanisms, BET inhibitors e.g. JQ1 slightly downregulate A-485 genes, but not as much as A-485. Indeed, p300/CBP activity promotes RNAPII recruitment independently of promoting pause release independently from BRD4.

p300/CBP-catalyzed acetylation promotes TFIID binding and PIC assembly:

- genes showing a slow TBP and RNAPII dissociation kinetics, such as Scl2a3 and Tfcp2l1, contain TATA-box
- genes displaying rapid TBP and RNAPII dissociation such as Nanog, Zfp42 and Esrrb, lack canonical TATA-box
- p300/CBP activity regulates the PIC assembly of enhancer-regulated, but not housekeeping genes.
- reduction in TAF1, TBP and RNAPII binding is positively correlated

p300/CBP-catalyzed acetylation is crucial for TFIID anchoring, and, consequently, for PIC assembly and transcription initiation at enhancers and enhancer-regulated genes.

Conclusions

- p300/CBP and deacetylase activities regulate dynamic (de) activation of enhancers
- p300/CBP-catalyzed acetylation promotes PIC assembly and RNAPII recruitment
- BRD4 acts as a p300/CBP downstream effector to promote RNAPII pause release
- Coupling of RNAPII recruitment and pause release enables rapid enhancer activation

8.2 SPT5 stabilization of promoter - proximal RNA polymerase II

Aoi et al. find that the conserved transcription elongation factor SPT5 stabilizes RNA polymerase II (RNA Pol II) in yeast and human cells. SPT5 loss triggers degradation of a RNA Pol II subunit through CUL3, VCP, and CDK9. Stabilizing RNA Pol II may be a SPT5's primary function to safeguard accurate gene expression.

Acute depletion studies: RNAPolII undergoes a 2-step pausing at NELF and at the +1 nucleosome: a depletion of NELF causes the RNAPolII to still be paused at the +1 nucleosome. Additional regulatory mechanism for the release of RNAPolII into the gene body.

SPT5 is an evolutionarily conserved and essential elongation factor. It forms the DSIF complex with SPT4. It interacts with RNAPolII throughout the elongation phase, probably active as a positive elongation factor, but impossible to knock-out.

8.2. SPT5 STABILIZATION OF PROMOTER - PROXIMAL RNA POLYMERASE II

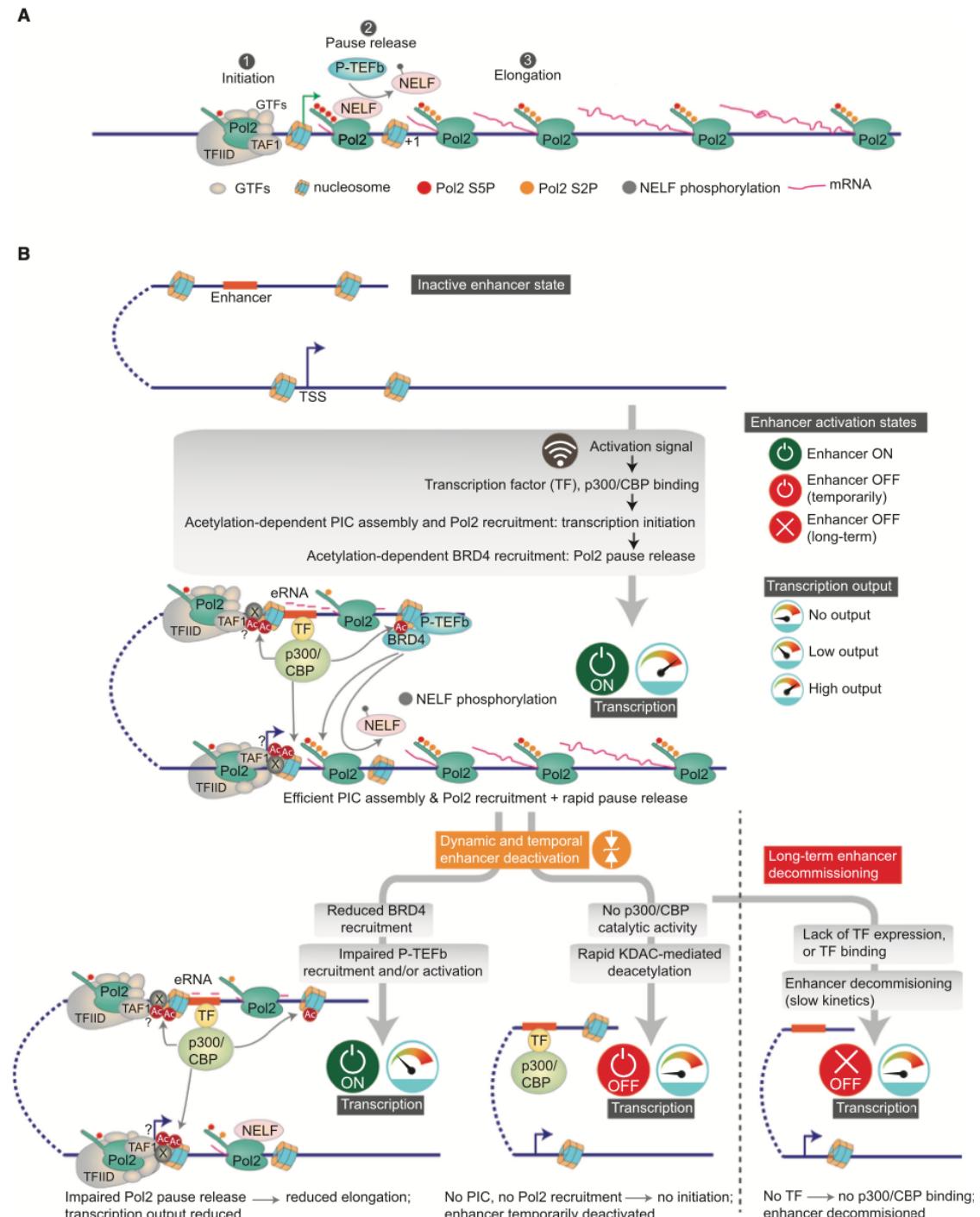


Figure 8.1: (A) Schematic depiction of the key steps in the transcription cycle. (B) Proposed hypothetical model of p300/CBP-activity-dependent dynamic enhancer activation. In response to activating signals, TFs recruit p300/CBP at enhancers where it acetylates chromatin-associated proteins. The acetylation of histones or non-histone proteins (X) promotes TFIID and RNAPII (Pol2) recruitment and transcription initiation. The question mark indicates the uncertainty of whether acetylation recruits TFIID directly or indirectly. Concurrently with promoting PIC assembly, p300/CBP-catalyzed acetylation promotes BRD4 binding, which facilitates Pol2 pause release by recruiting or activating P-TEFb. If BRD4 recruitment is impaired, then it leads to impaired Pol2 pause release and reduced transcription output. However, under this condition, p300/CBP continues to promote Pol2 recruitment, resulting in promoter proximal Pol2 pausing. If p300/CBP activity is inhibited, then PIC assembly, Pol2 recruitment, and pause release are impaired, leading to severe transcription inhibition. Under the conditions in which p300/CBP activity is impaired transiently, TFs and p300/CBP remain bound to chromatin and enhancers remain poised for rapid reactivation. However, if p300/CBP activity is inhibited for prolonged periods, then it will lead to reduced TF expression, which in turn will lead to impaired p300/CBP binding and long-term enhancer decommissioning.

8.2. SPT5 STABILIZATION OF PROMOTER - PROXIMAL RNA POLYMERASE II

Depletion of SPT5 through the generation of auxin-inducible degron (AID) avoids effects of long-term depletion and leads to an RBP1 degradation (the largest subunit of Pol II).

SPT5 physically localizes with RNAPolII at most transcriptionally active genes on chromatin. Therefore, SPT5 depletion may affect the protein-protein interaction properties/composition of RNAPolII transcription units. **CUL3** is one of the human Cullin family proteins that serve as scaffolds of Cullin-RING E3 ubiquitin ligase complexes. Cul3 knockdown partially stabilized RBP1 upon SPT5 depletion, but did not affect SPT5-AID degradation. Recent studies showed how DNA damage causes RBP1 degradation by NEDD4 or CUL5, but upon SPT5 degradation they did not show any stabilization ability. CUL3- RNAPolIII association on chromatin is triggered by SPT5-AID degradation.

VPC is known to mediate the unfolding of ubiquitylated proteins through its ATPase activity leading to efficient protein degradation. VCP inhibition did not stabilize RBP1 when inhibiting initiation by triptolide (TPL). There are (at least) two distinct pathways:

1. VCP-dependent degradation induced by SPT5 depletion
2. VCP-independent degradation induced by a TPL-stimulated initiation effect

RBP1 is degraded either during the early elongation stage at promoter-proximal region, or during the productive elongation stage, in absence of SPT5.

To determine how SPT5 depletion and VCP inhibition affect the steady state of transcription by RNA Pol II, SPT5-AID cells were treated with the VCP inhibitor, followed by auxin, and then ChIP-seq for RNA Pol II was performed. RPB1 degradation induced by SPT5 loss occurs during the early stages of elongation. RNAPolIII was able to travel substantial distances when SPT5 was depleted (although the distance was shorter than the control). RNAPolIII at the elongation stage is unlikely to be degraded in the absence of SPT5. Even when RNAPolII is stabilized at the promoter-proximal region, it still needs SPT5 for the elongation stage in the body of the gene.

NELF regulates the transition of pausing states. Upon NELF loss, RNAPolII travels from a first pause region to a second pause region corresponding to the +1 nucleosome dyad. The SPT5 depletion system, coupled with the VCP inhibition described earlier, may allow us to precisely investigate the SPT5 function in promoter-proximal pause/release. They performed precision run-on sequencing (PRO-seq) to map the RNA Pol II positions at base-pair resolution. RNA Pol II was paused at the second pause regions when depleting SPT5 and inhibiting VCP, which resembles NELF depletion.

Conclusions SPT5 regulates promoter-proximal pausing at the first pause region, likely through its interaction with NELF, and that the transition to the second pause region upon SPT5 depletion is through NELF dissociating from RNA Pol II. Two possible roles for SPT5:

1. SPT5 loss leads to unproductive elongation on nucleosomes, which may trigger RPB1 destabilization.
2. This raises another possibility: that SPT5 may regulate RPB1 stability and proper elongation (pausing and velocity) independently.

8.2. SPT5 STABILIZATION OF PROMOTER - PROXIMAL RNA POLYMERASE II

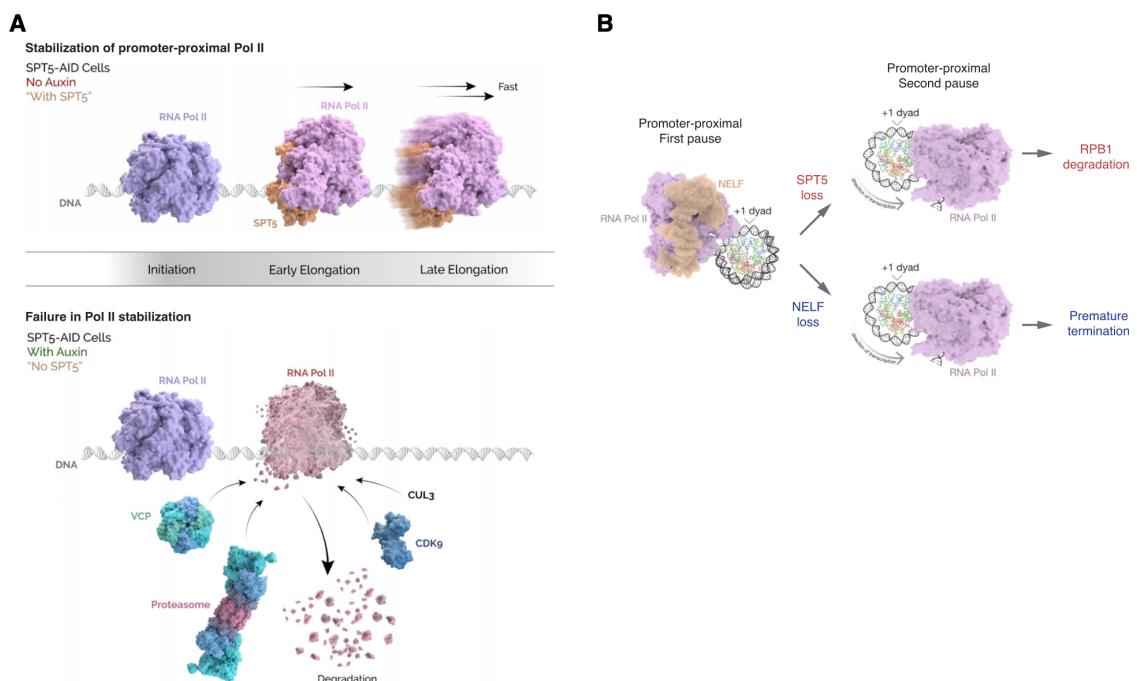


Figure 8.2: (A) SPT5 stabilizes RNA Pol II during the early elongation stage. With the SPT5 protein (upper panel), once RNA Pol II is initiated and escapes from the promoter, RNA Pol II undergoes productive elongation until termination. Without SPT5 (lower panel), early-elongating RNA Pol II undergoes RPB1 degradation that is mediated by the CUL3 ubiquitin ligase, the VCP unfoldase, and a novel form of the P-TEFb component CDK9. Late-elongating RNA Pol II without SPT5 can persist in transcription at a slower speed. (B) Failure in RNA Pol II passage on the +1 nucleosome leads to 2 distinct pathways to eliminate RNA Pol II from chromatin.

8.3. MYC RECRUITS SPT5 TO RNA POLYMERASE II TO PROMOTE PROCESSIVE TRANSCRIPTION ELONGATION

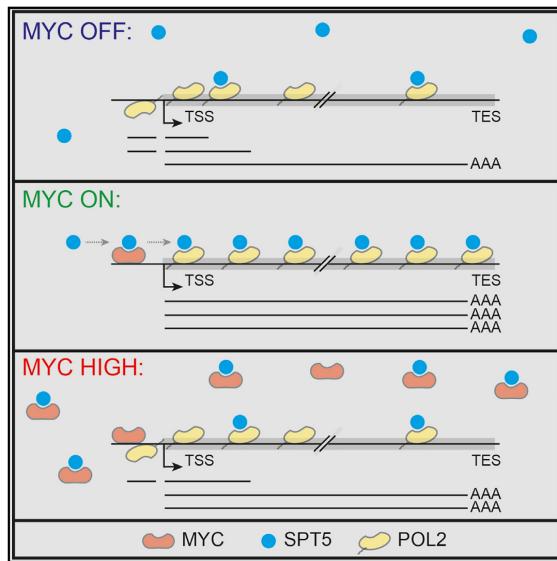


Figure 8.3

8.3 MYC Recruits SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation

MYC encodes for a nuclear phosphoprotein and is involved in cell cycle progression, apoptosis and cellular transformation. It forms heterodimers with MAX and binds to E-box containing DNA and regulates expression of non-coding transcripts and mRNA expression by RNA pol II.

ChiP-seq analysis revealed that MYC binds at the majority of open promoters. MYC is able to modulate the transcriptional cycle, and two mechanism are proposed:

- MYC can recruit CDK9 to mediate pause release
- MYC transfers Pol II associated factor PAF1 complex onto Pol II

Through the analysis of Tet-Off/On system, it was seen that MYC controls the assembly of transcriptionally engaged pol II complexes.

- 6 common proteins in pol II and MYC interactome, but only SPT5 and SPT6 changed their association with Pol II upon MYC depletion
- upon MYC depletion, read distribution at all expressed genes showed a global reduction of SPT5 binding → chromatin-associated Pol II binds to less SPT5 in the absence of MYC

By performing In situ Proximity Ligation Assay (PLA), we can achieve the detection of individual proteins or protein complexes using antibodies attached by DNA strands. PLA between SPT5 and pS2-Pol II shows clear reduction in nuclear proximity pair in absence of MYC in both U2OS and HMLE cells, meaning that upon MYC depletion we observe less interaction among SPT5 and Pol II.

8.4. BRD4-DIRECTED SUPER-ENHANCER ORGANIZATION OF TRANSCRIPTION REPRESSION PROGRAMS LINKS TO CHEMOTHERAPEUTIC EFFICACY IN BREAST CANCER

MYC recruits SPT5 by binding to its N-Terminal region.

CDK7 Activity is Required for Transfer of SPT5 from MYC to Pol II, as MYC facilitates the CDK7-dependent assembly of Pol II-SPT5 complexes. In addition, MYC Influences Processivity and Directionality of Pol II after pause release:

- Reduced processivity and directionality upon MYC depletion
- MYC Maintains High Pol II Elongation Rates
- Overexpression of MYC has an inhibitory effect on Pol II elongation rate. SPT4 is sequestered by MYC, and therefore its binding with Pol II is decreased. Certain genes lose processivity and directionality at oncogenic levels of MYC → MYC dependent sequestration of SPT5 might repress tumor-suppressive genes.

MYC is not a pioneer TF, it only binds to accessible binding sites. MYC is able to bind low-affinity binding sites in a scenario of MYC overexpression.

8.4 BRD4-directed super-enhancer organization of transcription repression programs links to chemotherapeutic efficacy in breast cancer

BET are epigenetic leaders, they can recognize marks through bromodomain. **BRD4** is a positive regulator of transcription. It is involved in some disorders, as inflammation, obesity and cancer, as it can enhance the transcription of oncogenes. JQ1 inhibits BET4 and is a promising antitumor drug, but a resistance can be developed.

Super-enhancers are clusters of enhancers occupied by master regulators e.g. BRD4 or Mediator, which control the expression of lineage-specific genes.

- NuRD complex: multi-subunit, repressor. Acts as chromatin remodeling complex and histone deacetylase.
- LSD1: histone demethylase
- Pellino (PEL1): E3 ubiquitin ligase

Thanks to MF we can identify interactors: all the components of LSD1/NuRD complex, confirmed by Western Blotting and coIP assays. BRD4 is directly interacting with LSD1 through the C-terminus.

Interaction with factors known to reside in super-enhancer regions e.g. MED-1 was consolidated with ChIP-seq, almost 23000 targets in common. The sequences bound are enriched in H3Kme1 and H3k27ac.

SE control genes in important signalling pathways and drug resistance e.g. PDPK1.

After 4 weeks cancer resistant cells emerge. BRD4 is quickly inhibited, while LSD1 or MTA3 in the early phase of treatment maintain chromatin contact, which is exactly lost at 4 weeks.

1. removal of LSD1
2. increase of target genes expression

8.5. HSF2 COOPERATES WITH HSF1 TO DRIVE A TRANSCRIPTIONAL PROGRAM CRITICAL FOR THE MALIGNANT STATE

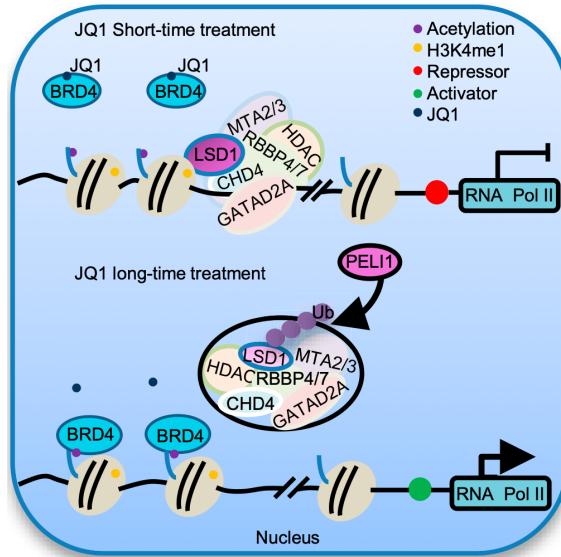


Figure 8.4

3. emergence of drug resistance

BRD4 is required for LSD1/NuRD recruitment at super-enhancers

Reduced occupancy could be due to the lower protein levels: LSD1 mRNA expression level is stable upon JQ1 inhibition, but protein levels are lower → proteasome action degrades the protein.

Elimination of PELI1 Improves the Therapeutic Efficacy of JQ1 in Breast Cancer Cells:

- Of 6 E3 ligases candidates, only PELI1 knockdown results in a strong increase of LSD1
- PELI1 inhibition improves JQ1 therapeutic efficacy
- the removal of BRD4/LSD1/NuRD complex from chromatin allows cancer cells to evade the selective pressure induced by JQ1 or other anti tumor drugs.

The Clinicopathological Significance of the PELI1-LSD1-BRD4/LSD1/ NuRD Axis in Breast Cancer:

- PELI1 mRNA levels are up-regulated when at least chemotherapy are applied
- high expression of PELI1 leads to a worse survival rate

Proposal: combined targeting of BRD4 and PELI1 as therapeutic for breast cancer

8.5 HSF2 cooperates with HSF1 to drive a transcriptional program critical for the malignant state

c