



HAL
open science

Roles of RNA Helicases and ncRNAs in genome spatial and functional organisation

Olivier Fosseprez

► **To cite this version:**

Olivier Fosseprez. Roles of RNA Helicases and ncRNAs in genome spatial and functional organisation. Cellular Biology. Université Paul Sabatier - Toulouse III, 2023. English. NNT : 2023TOU30299 . tel-04575726

HAL Id: tel-04575726

<https://theses.hal.science/tel-04575726>

Submitted on 15 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE
Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par
Olivier FOSSEPREZ

Le 4 décembre 2023

**Rôles des Hélicases d'ARN et des ARN non-codants dans
l'organisation spatiale et fonctionnelle du génome**

Ecole doctorale : **BSB - Biologie, Santé, Biotechnologies**

Spécialité : **BIOLOGIE CELLULAIRE**

Unité de recherche :

MCD - Molecular, Cellular and Developmental Biology Unit

Thèse dirigée par
Olivier CUVIER

Jury

Mme Reini LUCO, Rapporteur
M. Bernd SCHUTTENGRUBER, Rapporteur
M. Thierry FORNE, Rapporteur
M. Tom SEXTON, Examineur
M. Marcus BUSCHBECK, Examineur
M. Olivier CUVIER, Directeur de thèse
Mme Pascale BELENGUER, Présidente

Acknowledgements

My acknowledgment will be somewhat usual, as like any other PhD student, my thesis has been a collective work in many aspects. First of all, I need to thank the BSB doctoral school as well as the Fondation ARC for their grants, without which this work would simply not have existed.

First in the list of acknowledgment is my supervisor, Dr. Olivier Cuvier, whose constant trust, excessive optimism and transient scepticism have made a superb mentor in the arcane ways of Science. Thank you for accepting the sometimes-extravagant enthusiasm of the student I was, and for helping me shape this eagerness in the (mostly) fruitful experiments of a proper researcher. Thank you also for giving me the material means and freedom to perform said experiments.

Of course, I would like to thank also the whole Cuvier Lab for a great scientific atmosphere. A special thank goes to my collègues de bureau, Charlène Perrois for experimental help and teachings, and David Depierre for his bioinformatic pedagogy; and both of them for their friendship as well as their mental support throughout my thesis. I also thank all the bioinformaticians who participated in the analysis of my data: Nicolas Chanard, Refka Askri and especially Robel Tesfaye. Finally, I thank all of the internship students I had throughout my thesis for teaching me about myself just as much as I taught them about Biology.

I would also like to thank my follow-up committee, Séverine Chambeyron and Marcus Buschbeck, whose critics and advice were of great help.

I thank my PhD defense jury, Dr. Reini Luco, Dr. Thomas Sexton and Dr. Thierry Forné who kindly accepted to evaluate my thesis and review my manuscript, providing expert and skilful feedback.

I would like to thank Pascale Belenguer for tutoring me during my teaching to molecular biology students, as well as for accepting to preside my defense jury.

Many thanks to all friends and people at the CBI that I had the chance to meet for the countless discussions and breaks, for their gracious advices, support and lending of material. Similarly, I thank the administrative, IT and platform staff who give their best to help us work in a good environment, even though they must know by now that we desperately often fail our experiments.

Finally, I thank my family, who gave me continuous support, undeterred by their lack of understanding of my work. And last but not least, I thank Aurore Corset for her unrelenting support and love throughout the fluctuations of my mood in this arduous marathon. Love you 3000.

ABSTRACTS

Genes' transcription is an essential nuclear process, influencing cells and organisms at virtually all stages of life and development. It is therefore tightly controlled to ensure proper gene expression programs and avoid deleterious mis-regulations. In the last decades, the blooming fields of Chromatin and Epigenetic have become prevalent in the study of gene expression regulations. Truly, from nucleosomes and their modifications, to the large-scale 3D structures formed to mediate and constraint spatial interactions of genomic elements, all the components of chromatin have been demonstrated as influential in controlling transcription at all steps. Insulator elements, for example, were found to limit spatial contacts between neighbouring regions to avoid unwanted regulatory interactions, thus forming Topologically Associating Domains (TADs). More recently, transcription was shown to happen not only at defined genes, but more pervasively in nearly all contexts. Notably, Enhancers and Promoters, the two main genomic elements enabling gene transcription, are transcribed into relatively unstable non-coding RNAs, named eRNAs and PROMPTs, which are thought to participate in gene regulation.

The search for the various factors that mediate and influence genome organisation, both spatially and functionally, is still ongoing, with new roles and factors discovered regularly. Insulators in particular have been shown in the last two decades to be composed of around 15 proteins in *Drosophila*, and at least 4 in Human. Recently, RNAs, and in particular non-coding RNAs, have been proposed as regulators of insulator proteins, influencing their binding and oligomerization at specific loci. In this context, we found several RNA helicases in interactions with insulator complexes in *Drosophila melanogaster* S2 cells. Hence, we focused on three of these RNA helicases, namely Mtr4, Rm62 (DDX5 in Human) and Abstrack (DDX41 in Human), to decipher their potential roles as insulator co-factors. In this study, we use predominantly genome-wide sequencing and bioinformatic analyses to investigate coding and non-coding RNAs regulations (RNAseq), as well as the recruitment and binding to chromatin of several insulator proteins (ChIPseq), and the influences of these RNA helicases on 3D chromosome conformation (HiC). We complement these genomic approaches with microscopy experiments to examine the impact of RNA helicases and ncRNAs on the formation of transcriptional and insulator condensates. Finally, we also make a proof-of-concept experiment using CRISPR-dCas9 to target the accumulation of specific eRNAs and thus potentially enact changes in insulator recruitment and gene expression.

Despite our relative lack of understanding of the underlying mechanisms, we have identified that all three RNA helicases have contextualized effects that depend on the type of insulator complexes

that are considered. We propose that these helicases exert different effects and at different sites, possibly through separate mechanisms. Thus, we support that RNA helicases may be an important branch of insulator complexes regulators that account for some of the previously described loci-dependent behaviour of insulators. Along this course, the global accumulation of non-coding RNAs upon depletion of Mtr4 has also permitted us to study the effects of non-coding RNAs accumulation when uncoupled from their transcription, a current subject of intense debate in the field. This study brings new evidence of the diversity of mechanisms impacted by non-coding RNAs and their regulation, laying the ground for future investigations of local chromatin regulations by RNAs.

RÉSUMÉ

La transcription des gènes est un processus nucléaire majeur qui influence les organismes et les cellules à toutes les étapes de leur développement. La transcription est donc précisément contrôlée pour assurer le bon fonctionnement des processus transcriptionnels. Au cours des dernières décennies, les domaines de la chromatine et de l'épigénétique sont devenus prévalents dans l'étude de la régulation de l'expression des gènes. Des nucléosomes et leurs modifications jusqu'aux grandes structures 3D formées pour médier et contraindre les interactions des éléments génomiques, l'importance de la chromatine dans le contrôle de la transcription a été démontré à toutes ses étapes. Par exemple, les insulateurs limitent les contacts entre régions voisines pour empêcher certaines interactions régulatrices délétères, formant ainsi les Domaines Topologiquement Associés (TADs). Plus récemment, il a été découvert que la transcription n'a pas lieu seulement sur des gènes bien définis, mais de manière plus pervasive. Ainsi, les Promoteurs et les Enhancers, éléments principaux permettant la transcription des gènes, sont transcrits en ARN non-codants appelés PROMPTs et eRNAs. Ces derniers sont également proposés comme régulateurs de l'expression des gènes

La recherche des divers facteurs qui composent et influencent l'organisation du génome, à la fois spatialement et fonctionnellement, est encore en cours, tandis que de nouveaux facteurs et rôles sont régulièrement découverts. Parmi ceux-ci, les insulateurs sont composés d'environ 15 protéines chez la Drosophile, et au moins 4 chez l'Humain. Récemment, les ARNs, et en particulier les ARNs non-codants, ont été proposés comme régulateurs des protéines insulatrices, influençant leur liaison et leur oligomérisation à certains *loci*. Dans ce contexte, nous avons découvert plusieurs hélicases d'ARNs en interaction avec les complexes insulateurs dans des cellules S2 de *Drosophila melanogaster*. Nous nous sommes alors concentrés sur 3 de ces hélicases d'ARNs, à savoir Mtr4, Rm62 (DDX5 chez l'Humain) et Abstrakt (DDX41 chez l'Humain), pour caractériser leur potentiel rôle de co-facteurs d'insulateurs. Dans cette étude, nous utilisons principalement des méthodes de séquençage du génome entier analysés en bio-informatique, pour mesurer la régulation des ARNs codants et non codants (RNAseq), le recrutement à la chromatine de plusieurs facteurs insulateurs (ChIPseq) ainsi que l'influence de ces hélicases d'ARN sur la conformation des chromosomes (HiC). Nous complétons ces approches génomiques par des analyses en microscopie portant sur l'impact de ces hélicases d'ARNs et des ARN non-codants sur la formation des condensats insulateurs et des condensats de transcription. Enfin, nous avons produit une preuve de concept pour cibler l'accumulation de eRNAs spécifiques par CRISPR-dCas9, et donc potentiellement modifier l'expression des gènes et le recrutement des insulateurs.

Malgré notre manque de compréhension des mécanismes sous-jacents, nous avons pu déterminer que les 3 hélicases d'ARN d'intérêt ont des effets contextuels dépendants du type de complexe insulateur considéré. Nous proposons que ces hélicases aient des rôles différents, à des sites distincts, potentiellement *via* plusieurs mécanismes. Nous émettons l'hypothèse que les hélicases d'ARN soient un groupe important de potentiels régulateurs des complexes insulateurs. Ils pourraient alors expliquer une partie des changements de comportement des insulateur en fonction des sites. Au cours de cette étude, l'accumulation globale d'ARNs non-codants nous a permis d'évaluer leur effet indépendamment de leur transcription, ce qui fait encore débat dans ce domaine. Cette étude apporte de nouveaux éléments quant à la diversité des mécanismes impactés par les ARN non-codants et leur régulation et pose les bases de futures investigations relatives aux régulation locales de la chromatine par les ARN.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	3
ABSTRACTS	5
INTRODUCTION	13
I. CHROMATIN ORGANIZES THE GENOME FUNCTIONALLY AND SPATIALLY	15
A. Chromatin as a mean to regulate transcription.....	15
<i>Heterochromatin & Euchromatin</i>	15
<i>Transcription and its regulation</i>	17
<i>DNA accessibility regulate transcription</i>	18
B. A spatially organized genome	19
<i>Roles and mechanisms behind chromatin loops and enhancer-promoter interactions</i>	20
<i>Insulators separate Topologically Associating Domains (TADs)</i>	21
<i>Compartments and Phase Separation</i>	24
II. NON-CODING RNAs, THEIR REGULATION AND THEIR MANY ROLES	27
A. Transcription & regulation of PROMPTs and eRNAs.....	27
<i>Transcription at regulatory regions and its diversity: PROMPTs and eRNAs</i>	27
<i>PROMPTs & eRNAs targeted degradation by the Exosome</i>	30
B. Roles for PROMPTs and eRNAs in functional mechanisms	31
<i>ncRNAs direct impact in Pol II transcription</i>	32
<i>ncRNAs influence on chromatin regulation</i>	36
<i>ncRNAs influence on 3D genome organisation</i>	38
<i>Deleterious effects of ncRNAs</i>	40
III. RNA HELICASES AND RNA DEGRADATION	42
A. DEAD-box & Ski2-like RNA helicases and their roles.....	42
B. RNA helicases found in insulator complexes	44
<i>Mtr4</i>	45
<i>Rm62 – DDX5</i>	46
<i>Abstrakt – DDX41</i>	48

RESULTS	51
CHAPTER 1 - MTR4-CONTROLLED NON-CODING RNAs MAY REGULATE GENE EXPRESSION THROUGH INSULATOR COMPLEXES IN DROSOPHILA.....	52
Mtr4 interacts with insulators and regulates ncRNAs and genes.....	52
Mtr4 depletion alters the binding of distinct insulator complexes to chromatin.....	54
Types of insulator complexes have distinct chromatin context, and associate with ncRNAs and gene deregulation upon Mtr4 KD	58
Mtr4 and ncRNAs accumulations influence the formation of insulator and transcriptional condensates.....	61
Mtr4 does not affect insulation by insulator complexes.....	63
Mtr4 influences long-range interactions.....	65
Locus-specific artificial accumulation ncRNAs by dCas9 is not generally sufficient to influence Beaf32 recruitment and gene expression	68
CHAPTER 2 – RM62 & ABS IMPACT ON GENE REGULATION, INSULATORS & GENOME SPATIAL ORGANISATION	70
Rm62 and Abs interact with insulators and regulate gene expression, but not ncRNAs.....	70
Rm62 & Abs depletion cause changes in distinct insulator complexes.....	72
Rm62 & Abs may participate in regulating 3D genome organisation.....	74
DISCUSSION.....	78
Limitations to the study of ncRNAs and their roles.....	78
The conundrum of Drosophila insulators.....	79
Multiple interplays between ncRNAs and genome organisation	81
Rm62 and Abs as potential insulator cofactors.....	83
METHODS.....	85
Wet-lab experiments.....	85
<i>Cell culture, dsRNA depletions & transfection.....</i>	<i>85</i>
<i>Cloning</i>	<i>85</i>
<i>FACS.....</i>	<i>86</i>

<i>Reverse Transcription & qPCR</i>	86
<i>RNA-seq</i>	86
<i>Western Blot & Co-Immunoprecipitation</i>	86
<i>Immunofluorescence & microscopy acquisition</i>	87
<i>Beaf32 & Pol II cluster detection</i>	87
<i>ChIP</i>	88
<i>HiC</i>	89
<i>iCLIP</i>	89
<i>CRISPR-dCas9</i>	89
<i>Antibodies</i>	90
Bioinformatic analysis	90
<i>RNAseq</i>	90
<i>ChIPseq</i>	91
<i>HiC</i>	92
APPENDIX	93
Supplementary figures - Chapter 1	93
Supplementary figures - Chapter 2	101
BIBLIOGRAPHY	106

INTRODUCTION

Evolution, although known since prehistoric times through the selection of plants and animals for better traits, has only begun to be understood in the 19th century, with the theories of Evolution and Genetic inheritance of Charles Darwin and Gregor Mendel. In the following century, amidst the accumulating discoveries on DNA (Dahm, 2005), its capacities to mutate and be transmitted (Morgan, 1910), the teachings of Genetics became an axiom. At the same period was developed the current paradigm of genes, transcribed into messenger RNA (mRNA)(Brenner, Jacob and Meselson, 1961; Gros *et al.*, 1961), to themselves be translated into proteins (Cobb, 2017). These breakthroughs explained many biological processes observed before, and were essential to the understanding of countless diseases. Thus, it became evident that genes, their mutations and their capacities to produce mRNA, were seminal to living organisms throughout development, both healthy and disordered. Yet, it was soon apparent that pinpoint mutations (Single Nucleotide Polymorphism, SNP) of the DNA of genes and the combination of alleles were not the only players in this game. First, because with the development of DNA sequencing in the late 1970s (Sanger, Nicklen and Coulson, 1977), it was quickly confirmed that most of our genome is not only composed of genes, but also of so called “junk DNA” (Ohno, 1972). It is this inadequately named DNA that challenged the perception of how mutations impact living organisms, as numerous deletions and SNPs of intergenic DNA were shown to be causative to specific phenotypic changes and diseases (Dunham *et al.*, 2012). Secondly, the genetic principles applied poorly to embryogenesis and the development of eukaryotes, in which a single cell multiplies to become various cell types, albeit with the exact same set of genetic material.

This question of how could non-genic material have an impact led to the combination of genetics, i.e. the study of genes, with previous studies pertaining to embryology and cell differentiation. In particular, the groundbreaking model of Conrad Waddington (*Figure 1*) which explained cell specification through an “epigenetic landscape”, a series of valleys representing various cell lineages, and mounts between them representing the increasing difficulty for a differentiating cell to re-differentiate into another cell type (Waddington, 1942, 1953). This model became an accurate description of how a totipotent cell can differentiate in any cell type, but as it goes down a valley, or steps of specification, from pluripotent to stem cell to progenitor and finally to fully specialised cell, its possibilities shrink, and its capacities transform to be adapted to its functions.

All of the processes of differentiation have a common basis: the regulation of gene expression. Especially of interest for this manuscript, gene expression is regulated through various processes that

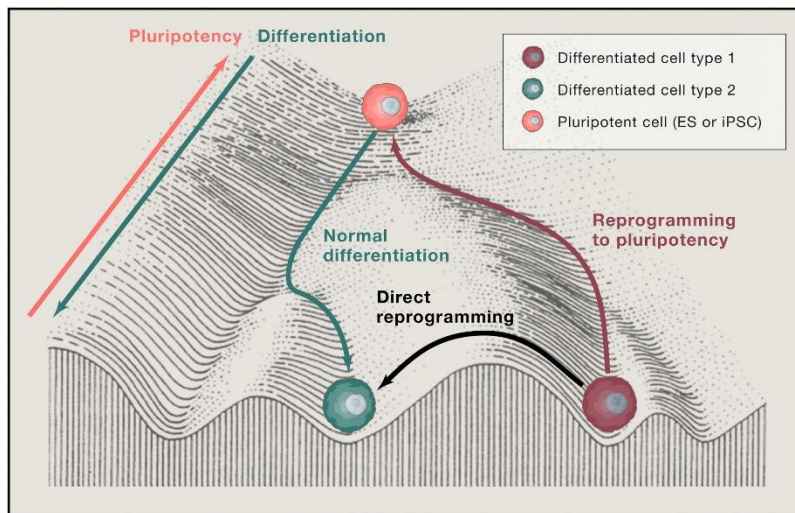


Figure 1: Schematic of the epigenetic landscape model by Conrad Waddington. Valleys represent cell differentiations, and the heights between them highlight the increasing difficulty for a cell to reprogram as it goes down the path of a particular lineage. Adapted from (Srivastava and DeWitt, 2016) for their added representation of reprogramming to the original schematic.

are encompassed in the generic term of Epigenetic. Although this term implies the transmission of information from one organism to its progeny without altering the genetic code itself, it has come to be used also for the maintenance of gene expression through cell division, thereby preserving cell type specificities. In particular, Epigenetics is often used to refer to chromatin and its regulation of three major nuclear processes: transcription, replication and DNA repair. Chromatin states, being more or less permissive to the transcription machinery, are indeed central to the regulation of gene expression. It was first demonstrated in the 1930's by the phenotypic observation of the *white eye* gene on/off expression switch in *Drosophila*, after X-rays induced random translocation (Muller, 1930) and translocation of that same gene to an "inert" region of a chromosome (Schultz, 1936). Interestingly, this effect had the additional phenotypic trait of forming clusters of cells with the same colour in a mottled pattern, demonstrating the conservation of the gene's expression for a group of cells, inherited through several mitosis events without changes to the DNA sequence. Since then, many other studies have enriched our knowledge of the numerous mechanisms through which gene expression is regulated by chromatin and its organisation at all levels. Notably, the acclaimed study by (Takahashi and Yamanaka, 2006) demonstrated the power of gene expression regulation when achieving de-differentiation of fibroblasts using only 4 defined transcription factors that remodelled gene expression programs. In the last two decades, this field has been propelled to new heights by the rise of genomics, i.e. the use of New Generation Sequencing (NGS) to study gene expression, protein binding and histone modifications distribution, as well as chromosome conformation, genome-wide. This manuscript will attempt to add a new stone to the wall of knowledge regarding epigenetics, focusing on the regulation of transcription by spatial organisation of the genome, while including the role of insulators and non-coding RNAs in these processes.

I. Chromatin organizes the genome functionally and spatially

A. Chromatin as a mean to regulate transcription

The generic term of chromatin refers to the association double strand DNA with proteins. Primarily, DNA wraps itself around histone octamers to form nucleosomes, in stretches of roughly 200 base pairs (bp), separated by linker DNA (*Figure 3A*). This foremost level of organisation of the DNA enables several things: first, to protect DNA against double strand breaks; second to compact DNA to facilitate its packaging in the nucleus; third, to enable the localised maintenance of information through the histone code of Post-Translational Modifications (PTM); and fourth, to restrict the accessibility of the DNA to other factors. Beside this prevailing organisation, hundreds of diverse proteins and RNA are also incorporated into chromatin, such as transcription factors (TF), DNA replication and repair machineries, histone PTM writers, readers and erasers. All of these factors impact transcription at each step, i.e. formation of the Pre-Initiation Complex (PIC), Initiation, Pause ~50bp after the Transcription Start Site (TSS), productive Elongation and Termination at the Transcription Termination Site (TTS) (*Figure 2A*) (Cramer, 2019). Transcription was seen for a long time as a relatively homogeneous process in time, yet in the past 15 years, it has been accepted as discontinuous in the form of transcriptional “Bursts” separated by periods of inactivity, complexifying the vision of how gene expression may be regulated (Tunnacliffe and Chubb, 2020).

Heterochromatin & Euchromatin

One of the earliest observations made about chromatin was its heterogeneity, as described by Heitz in 1928. Already in these pioneer works, euchromatin, less condensed and transcriptionally active is opposed to heterochromatin that is compacted and transcriptionally inert. This concept has since been refined, in particular with the development of NGS-coupled techniques such as genome-wide Chromatin Immunoprecipitation (ChIP-seq), as ample evidence was accumulated to demonstrate that several sub-classifications are needed. For instance, heterochromatin can be divided into constitutive and facultative. Constitutive heterochromatin, found at peri-centromeric regions, telomers, repeats and Lamin-Associated Domains, is conserved in all cell types and throughout development, and is marked by factors such as Heterochromatin Protein 1 (HP1) and the histone PTMs H3K9me3 and H3K9me2 (Saksouk, Simboeck and Déjardin, 2015). Contrarily, facultative heterochromatin is established during development amid cell differentiation to silence specific groups

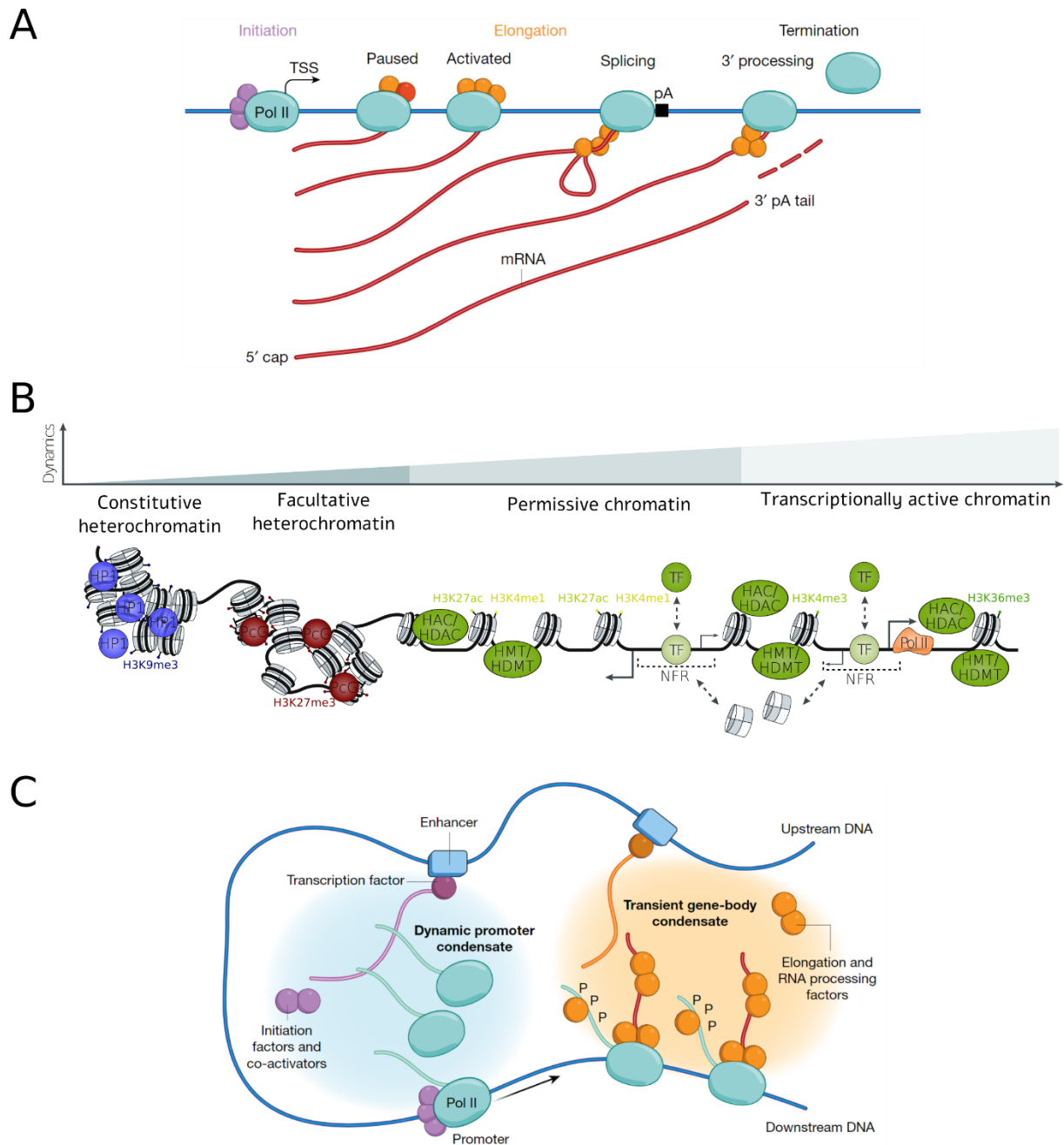


Figure 2: Schematics Transcription models & chromatin 1D organisation **A.** Schematic of the conventional steps of transcription, from initiation to termination **B.** Schematic of chromatin states and domains. In each state is represented the histones PTMs and proteins associated with it **C.** Schematic of the transcriptional model where two separate microphases or condensates are formed for initiation and elongation. **A & C** are from Cramer 2019. **B** is adapted from Klemm, Shipony & Greenleaf, 2019.

of genes through the deposition of H3K27me3 and H2K119ub by the Polycomb complexes (PcG), thus maintaining a transcriptional program specific to each cell lineage (Kim and Kingston, 2022). Euchromatin on the other hand has long been seen as relatively homogenous in its permissiveness to transcription. Yet in 2010, a key study by Filion et al. showed using the distribution patterns of 57 chromatin factors and histones PTMs that two distinct classes of euchromatin can be sorted out. Both

produce substantial amount of RNAs, and share factors such as histone acetylases and deacetylases as well as H3K4 and H3K79 methylations, yet these classes are dissimilar notably in their replication timing and the type of genes they harbour. One, enriched in housekeeping genes, is decorated with H3K36me3 and its reader MRG15, while the other is composed of genes expressed only in certain tissues and cell types, regulated by various DNA binding factors. As a development of this axis, prediction & annotation tools such as ChromHMM have demonstrated great efficiency at classifying genomic elements using the combination of chromatin markers present (Ernst and Kellis, 2017). When active, these elements are found in euchromatin domains, where they enable an extensive range of gene regulation, dependent on their precise dosage of DNA accessibility and transcription factors binding (*Figure 2B*).

Transcription and its regulation

Enhancers and promoters are DNA regions that enable regulation of gene expression by facilitating the formation of the Pre-Initiation Complex (PIC), and hence loading the transcription machinery, i.e. the RNA Polymerase 2 (Pol II), its General Transcription Factors (GTF), and accessory factors. Promoters, localised directly upstream from the TSS, are nucleosome-depleted regions when genes are active, and contain motifs, such as the TATA element, to enable binding of Pol II and its GTF. Other specialised TF bind to promoters to help guide the polymerase to specific target genes, thus exerting another layer of gene regulation. Enhancers have similar properties, albeit localised further away from genes, up to hundreds of kilo base pairs (Kb). They also bind a large variety of TF, and may regulate several genes simultaneously (Fukaya, Lim and Levine, 2016; Cramer, 2019). The exact mechanism by which enhancer enact their effect from such large distances has remained elusive, even though their role was discovered more than 40 years ago (Banerji, Rusconi and Schaffner, 1981). The discovery of the 3-dimensional organization of the genome, discussed in section I.B., brought a first answer by confirming the previously-proposed enhancer-promoter interactions, that can occur between elements separated by up to mega bases pairs (Mb). However, a general and comprehensive mechanism, explaining how enhancer select which gene they regulate amongst the overwhelming number of possibilities, is still lacking. A recent branch of studies brings evidence that eukaryote transcription factors may exert their role through Intrinsically Disordered Regions (IDR). Indeed, about 80% eukaryotic transcription factors possess such unfolded domains (compared to ~5% in prokaryotes) along with their DNA-binding domain (Ferrie *et al.*, 2022). In this model, IDRs facilitate the clustering, or microphase-separation of transcription factors at promoters and enhancers in the form transcriptional condensates, thus causing transcriptional burst induction (Boija *et al.*, 2018; Cho *et al.*,

2018; Sabari *et al.*, 2018; Cramer, 2019; Shrinivas *et al.*, 2019; Henninger *et al.*, 2021; Kawasaki and Fukaya, 2023). Importantly, a model of transcription is currently emerging in which phase separation enables the formation of separate condensates dedicated to initiation on the one hand, and to elongation, RNA processing and splicing on the other hand. The transition between these two phases would then be ensured by phosphorylation on the Pol II CTD (Cramer, 2019; Guo *et al.*, 2019)(Figure 2C).

Of note, both enhancers and promoters are bi-directionally transcribed into enhancers RNAs (eRNAs) and PROMoteur uPstream Transcripts (PROMPTs) (Seila *et al.*, 2008; Core *et al.*, 2008; de Santa *et al.*, 2010; Kim *et al.*, 2010; Preker *et al.*, 2008; Rennie *et al.*, 2018), as discussed further in part II.A. For enhancers, this discovery has unveiled the question of whether their transcription was a mere by-product of gene enhancement, or if it played a part in gene regulation. Even though these non-coding RNAs (ncRNA) are mostly short lived, several studies have demonstrated that some eRNAs participate in gene regulation (Ørom *et al.*, 2010; Li *et al.*, 2013; Mousavi *et al.*, 2013; Hsieh *et al.*, 2014; Lai *et al.*, 2015; Sigova *et al.*, 2015; Weintraub *et al.*, 2017; Tsai *et al.*, 2018; Abdalla *et al.*, 2019; Lee *et al.*, 2021; Chen *et al.*, 2023). However, the extent to which this can be interpreted as a global rule remains to be determined, as surveyed in section II.B. of this manuscript.

DNA accessibility regulate transcription

DNA accessibility to the transcription machinery is pivotal to the initiation of transcription (Cramer, 2019). Transcription factors, by binding to precise motifs at promoters and enhancers, improve the efficiency of Pol II loading and initiation of transcription. To that end, they may also recruit or activate an array of protein cofactors, such as chromatin remodelers, histone PTM readers and modifiers, transcription coactivators, or directly components of the general transcription initiation complex (Cramer, 2019; Ferrie *et al.*, 2022). The dynamic positioning of the nucleosomes themselves, their PTMs and composition in specific histone variants, govern the accessibility of DNA motifs to transcription factors and the Pol II machinery (Venkatesh and Workman, 2015). Hence the major role of chromatin remodelers such as ISWI and SWI/SNF in transcription regulation (Barisic *et al.*, 2019). Histone variants, such as H2A.Z (Weber, Ramachandran and Henikoff, 2014) and H3.3 (Armache *et al.*, 2020) have also been shown to control transcription initiation, in the case of H2A.Z by destabilizing nucleosomes. Inversely, studies showed that the histone PTM H3K9me3 directly compacts chromatin via the binding of HP1 (Hiragami-Hamada *et al.*, 2016; Machida *et al.*, 2018). In the same way, H3K36me3, though associated with processive elongation, inhibits transcription initiation through the

recruitment of the histone deacetylase (HDAC) Rpd3, supposedly to avoid random cryptic Pol II loading in otherwise permissive regions (Carrozza *et al.*, 2005)(Figure 2B).

B. A spatially organized genome

Eukaryotes genomes have the particularity to be organized not only in one dimension (1D), but also spatially (3D). It was first demonstrated by microscopy and electron microscopy experiments that interphasic chromosomes are spatially organized (Stack, Brown and Deweyf, 1977). With the evolution of microscopy techniques, the link between 3D genome organization and gene regulation was made, as the global position of a gene in a nucleus was correlated with its transcriptional activity (Kurz *et al.*, 1996; Dietzel *et al.*, 1999; Cremer and Cremer, 2001). However, the causative effect of genome spatial organization on gene expression became clearer only with the development of Chromosome Conformation Capture (3C) technologies, starting with Dekker *et al.* in 2002. This method truly reached its potential when coupled with NGS technologies in the form of HiC, allowing for a genome-wide visualisation of spatial organization at various scales (Lieberman-Aiden *et al.*, 2009; Rao *et al.*, 2014). More recently, other ligation-based techniques such as Micro-C (Hsieh *et al.*, 2015) and HiChIP (Mumbach *et al.*, 2016) have been devised to overcome specific limitations of the HiC original protocol, namely the intrinsically limited resolution, and the lack of information regarding the mediators of the observed contacts, respectively. SPRITE, a ligation-free technique based on split-pool manipulations and tag elongation to identify multi-fragment interactions, has since overcome a major limitation of ligation-based techniques, which is their inability to confirm the simultaneous interaction of more than two DNA fragments (Quinodoz *et al.*, 2018, 2021). To this day, other limitations remain, e.g. the incomplete single-cell resolution that confine conclusions to either averages over populations of cells, or very low resolution in each cell, as well as non-exhaustive possibilities due again to the inability of ligation-based techniques to observe multi-fragment interactions (Nagano *et al.*, 2013, 2017). For this question, microscopy in the form of highly multiplexed hybridization of fluorescent probes is currently spear-edging the study of spatial organization, yet with difficulties to attain the genome-wide resolution of the above sequencing-based methods (Cardozo Gizzi *et al.*, 2019; Cattoni *et al.*, 2017; Jerković & Cavalli, 2021; Nguyen *et al.*, 2020). In the following sections, the currently accepted model of a multi-layered, nested genome organisation will be described. This model depicts enhancer-promoter interactions and other chromatin loops being mostly restricted by Topologically Associating Domains (TADs), while TADs come together to form compartments of similar transcriptional status. Likewise, all of the above interactions are mostly restrained inside of separate chromosomes, forming so called Chromosome Territories, with remarkably little interactions between different

chromosomes. Finally, the more recent addition of phase separation models to explain the behaviour of genome organization will be discussed (*Figure 3*)

Roles and mechanisms behind chromatin loops and enhancer-promoter interactions

Arguably the main process by which spatial organisation regulates gene expression is the regulation of enhancer-promoter (E-P) interactions (*Figure 3F*). The current mechanistic view to explain the formation of these interactions is the loop-extrusion model (Davidson and Peters, 2021) (*Figure 3E*). In this model, chromatin loops are formed by the loading of the Cohesin ring complex onto the DNA, which then dynamically extrudes the DNA strand (Rao *et al.*, 2017; Davidson *et al.*, 2019). This extruding loop may then be slowed, blocked, or Cohesin may be unloaded from DNA, by different factors. The main human architectural protein, CCCTC-binding factor (CTCF), has been thoroughly demonstrated as an anchor to Cohesin, obstructing loop extrusion progress (Davidson *et al.*, 2023), while the transcribing Pol II has been advanced as slowing and/or maintaining extruding loops (Banigan *et al.*, 2023; Zhang *et al.*, 2023). Mediator, the transcriptional coactivator complex that transmit enhancer-bound TFs information, also interacts with Cohesin to regulate gene expression (Kagey *et al.*, 2010). However, whether Mediator is really necessary to architecturally bridge enhancer and promoters is still debated (El Khattabi *et al.*, 2019; Ramasamy *et al.*, 2023). In essence, the loop extrusion mechanism makes all genomic interactions possible, by bringing any two regions into close proximity at some point. The observed increased, or more correctly the more frequent contacts, are thus due to the slowing or the anchoring of Cohesin, which forms a transient Long-Range Interaction (LRI) (*Figure 3D*). Importantly, the loop extrusion model is not as well-demonstrated in other non-mammal organisms, such as *Drosophila*, where direct evidence lacks, notably to demonstrate loop-extrusion by Cohesin. It is however assumed to be conserved to some extent, as it is between Yeast and Human, and because different types of LRIs are also observed (Sexton *et al.*, 2012; Liang *et al.*, 2014; Espinola *et al.*, 2021). However, more factors than those observed in mammals are involved, as discussed in the following section on TADs (Vogelmann *et al.*, 2014).

Regarding gene regulation by chromatin loops, although some mechanisms are still being investigated, it is already quite clear that LRIs are at the basis of E-P interactions and their specificity inside TADs, thus becoming major regulators of gene expression. Depletion of Cohesin for instance certainly affects transcription at all steps (Reviewed by Dorsett & Merkenschlager, 2013), regulating genes both in *Drosophila* (Rollins, Morcillo and Dorsett, 1999; Schaaf *et al.*, 2013), and in Mammals (Kagey *et al.*, 2010; Schmidt *et al.*, 2010). Interestingly, two studies observed in *Drosophila* that DNA

loops pre-empt gene-activation during development, and may not be directly causative of gene expression levels, suggesting that formation of E-P contacts is in fact a framework to potentiate and facilitate gene expression regulation (Espinola *et al.*, 2021; Ing-Simmons *et al.*, 2021). It is through this lens of development that the importance of LRIs formations is emphasized, particularly to time and localise the expression of specific genes (Batut *et al.*, 2022; Levo *et al.*, 2022). In Human, aside of the prevalent CTCF, Yin Yang 1 (YY1) has been associated with the formation of E-P loops (Weintraub *et al.*, 2017), demonstrating that the mediation and regulation of specific E-P interactions involves others factors than those directly involved in the structuration of the genome through chromatin loops and TADs. This specificity enables the precise control of gene expression for cell identity (Schoenfelder and Fraser, 2019; Batut *et al.*, 2022). Importantly, long-range interactions may also help silence genes, as exemplified by Polycomb deposition and maintenance of H3K27me3 islets (Heurteau *et al.*, 2020; Cai *et al.*, 2021; Kraft *et al.*, 2022) and domains. In Drosophila, the Polycomb loops are the only apparent and observable loops, as the TAD corner-loops are not present (Sexton *et al.*, 2012). However, other methods such as 4C-seq, aggregation of HiC with deeper sequencing, and microscopy have shown that enhancers and promoters for example do preferentially interact together, even though such interactions are not readily seen on HiC maps at single loci (Ghavi-Helm *et al.*, 2014; Espinola *et al.*, 2021; Batut *et al.*, 2022).

Insulators separate Topologically Associating Domains (TADs)

Topologically Associating Domains, or TADs, are continuous genomic regions that predominantly interact with themselves, rather than with the neighbouring TADs (*Figure 3C & G*). Depending on the insulating frontier strength and the density of contacts inside each TAD, the contacts found between two contiguous TADs may be rare, or simply less probable than those inside each TAD. The field generally agrees that TADs are formed through the process of loop extrusion, with prevailing anchors forming the frontiers, called insulators (Cai & Levine, 1995; Kellum & Schedl, 1991). The Insulator Binding Protein (IBP) CTCF in Mammals would thus, when its motifs are in converging orientations, block Cohesin extrusion. This would favour all contacts in-between the Cohesin loading sites and the CTCF boundary, while preventing loop extrusion to bring into contact two sites separated by a frontier made of CTCF (Mach *et al.*, 2022). This results in HiC, and other 3D-probing technologies, in large triangles, from tens of kilobases to few megabases, of more frequent interactions, separated by less frequent interactions (Szabo, Bantignies and Cavalli, 2019)(*Figure 3G*). Of note, different model organisms and the increase in sequencing depth and resolution demonstrated that various TADs and sub-TADs particularities exist, and a wide length distribution can be observed (Beagan and Phillips-

Cremins, 2020). Commonly however, active TADs tend to be smaller and more gene-rich, while inactive TADs are large and harbour fewer genes.

At this level can be observed an extensive overlap between chromatin transcriptional states and their epigenetic marks, highlighting that TADs are really the building blocks of genome structuration. Recently, TADs were observed as separate entities in microscopy, confirming their

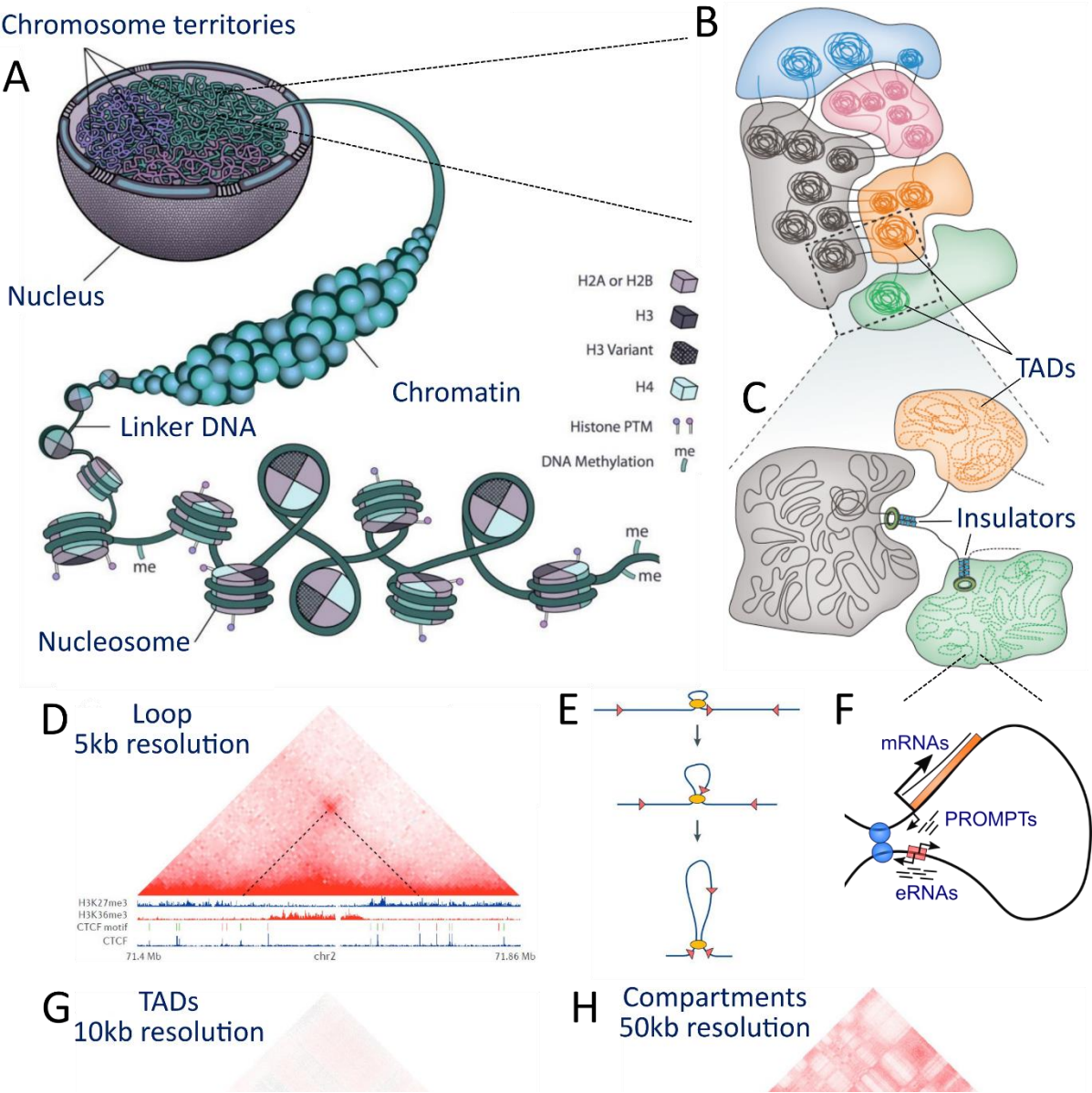


Figure 3: Spatial organisation of the genome: from nucleosomes to compartments **A.** Schematic of nucleosomes forming chromatin, to be packaged into the nucleus into chromosomes territories **B.** & **C.** chromosome territories are composed of compartments, each clustering TADs of similar transcriptional types (in different colours) **D.** A loop as observed in HiC, 5kb bins **E.** Model of loop extrusion, with Cohesin (yellow) extruding, and CTCF (orange triangle) anchoring the loop **F.** Representation of an E-P interaction, with production of mRNAs, eRNAs & PROMPTs **G.** Map HiC at 10kb with clear TADs observed **H.** Map HiC at 50kb resolution, at which resolution the compartments are evident. **B, C, D, G & H** are adapted from Bonev & Cavalli 2019. **E** is adapted from Davidson & Peters 2021.

existence in single-cells while highlighting their inconsistency between cells, in partial contrast with the view that they may simply be the result of averages over large population of cells (Szabo *et al.*, 2020). In fact, insulators do not only delineate epigenetic marks deposition and spreading (Gaszner and Felsenfeld, 2006; Emberly *et al.*, 2008; Lhoumaud *et al.*, 2014), but also help to favour and constrain specific E-P interactions inside TADs, while prohibiting contacts between genes and enhancers separated by insulators (Zuin *et al.*, 2022). In this sense, TADs and their boundaries are also important regulators of gene expression. Surprisingly however, the depletion of CTCF in mammals or that of other insulator proteins in *Drosophila* leads to remarkably little gene-deregulation (Hsieh *et al.*, 2022), even though CTCF is essential from the first steps of development to enable correct cell-specification (Chen *et al.*, 2019; Franke *et al.*, 2021). This was explained in the way that insulators are particularly important to regulate cell-to-cell variation in gene expression at both TADs and E-P interactions levels, rather than for specific up or down-regulations of genes (Ren *et al.*, 2017). Nonetheless, it remains unclear and debated how insulating frontiers can have both important roles in gene regulation in some loci, and yet be almost negligible in other cases.

To date, the model system of *Drosophila melanogaster*, although much used in 3D genomics, has revealed somewhat more complex than Human or Mouse to study TADs and chromatin loops. The numerous factors that influence genome organization, contrarily to the almost omnipotent CTCF and Cohesin in mammals, complexify experiments. Through the study of insulators, various IBPs were shown to form chromatin loops and have enhancer-barrier properties, notably Boundary element associated factor of 32kD (Beaf32)(Zhao, Hart and Laemmli, 1995), *Drosophila* CTCF homolog (dCTCF)(Moon *et al.*, 2005), Suppressor of Hairy Wings (Su(Hw))(Roseman, Pirrotta and Geyer, 1993), GAGA binding factor (GAF)(Ohtsuki and Levine, 1998; Li *et al.*, 2023), and M1BP (Bag *et al.*, 2021), among others (Cuartero *et al.*, 2014; Maksimenko *et al.*, 2015). In this peculiar model organism, ample evidence also shows that, beside Cohesin, several other insulator cofactors - i.e. non-DNA-binding proteins that are necessary for insulator functions – exist, such as Centrosomal Protein 190kD (CP190)(Liang *et al.*, 2014; Vogelmann *et al.*, 2014; Kaushal *et al.*, 2022; Cavalheiro *et al.*, 2023; Kahn *et al.*, 2023), Chromator (Vogelmann *et al.*, 2014) or modifier of *mdg4* (*mod(Mdg4)*).

In fact, the current view in the field regarding TAD formation, chromatin loops and E-P interactions in *Drosophila* is at best ambiguous. That is because DNA loops, although present in *Drosophila* are not as visible as those observed in vertebrates using HiC (Sexton *et al.*, 2012), and because the removal of most IBPs and cofactors on their own has limited impacts on the general 3D organisation of the genome (Cavalheiro *et al.*, 2023). That is also true of Cohesin, which does not seem to play a key role in *Drosophila* loops and TAD formation. Cohesin genomic repartition is also not as clear as in Mammals, where its peaks are found mostly at CTCF binding sites and Pol II enriched regions.

In *Drosophila*, although it can be found co-localized with insulators, it is not enriched as in Human at most IBPs binding sites, and is found rather at genes and enhancers, although non-exhaustively (Dorsett and Merckenschlager, 2013; Dorsett, 2019; Pherson *et al.*, 2019). It has been proposed that Cohesin and Condensin II compete to organize loops and global 3D genome organisation in *Drosophila*, yet with little molecular insight (Rowley *et al.*, 2019). Few tries were made at comprehensive studies of the different insulator complexes found in *Drosophila*, with the notable exceptions of Nègre *et al.*, 2010 and Ramírez *et al.*, 2018. The latter in particular showed that *Drosophila* boundaries could be subdivided into 8 clusters defined by motifs for different IBPs. Furthermore, a separation between promoter-associated boundaries (ZIPIC, Beaf32, M1BP and two other unidentified motifs) and non-promoter boundaries (Ibf1, dCTCF & Su(Hw)) was made. However, the lack of several IBPs and cofactors in these studies, the mostly bioinformatic analysis, as well as the aging methods used in Nègre *et al.*, 2010, call for updates of the conclusions made. Also, these studies focused on boundary functions of insulator complexes, leaving open the questions on loop-mediation by these IBPs.

Regarding this aspect, recent studies from the Levine lab and the Gambetta lab provided unprecedented demonstrations in single-cells that gene-regulatory 3D contacts are indeed mediated in *Drosophila* at various ranges, arguing for a dual role of IBPs: formation of regulatory LRI on the one hand, and enhancer blocking by insulating boundaries on the other hand (Batut *et al.*, 2022; Levo *et al.*, 2022; Li *et al.*, 2023; Mohana *et al.*, 2023). These HiC and microscopy experiments complement previous studies which observed, over cell populations, that specific contacts are established by insulator associating proteins, on top of their proper insulating function (Liang *et al.*, 2014; Vogelmann *et al.*, 2014). To summarize, the currently prevailing view of *Drosophila* TADs is that redundant and/or specialized IBPs act as loop anchors, while their cofactors act as the ligand between them, with more or less specificity, to form LRIs and TADs. Yet, the interactions are also restricted by insulator frontiers to avoid spurious gene regulation. The underlying mechanisms behind these observations are poorly understood, and the exact molecular role of Cohesin compared to its Mammal and Yeast counterpart remains to be elucidated.

Compartments and Phase Separation

Compartments are structures observable in microscopy as large globules, partitioning the genome in active and inactive chromatin territories (Bickmore and Van Steensel, 2013). They were observed in HiC maps in the form of a checkerboard pattern of interacting TADs. In fact, although TADs do not interact with their contiguous neighbour, they do interact with other TADs farther away that have similar transcriptional status, thus forming compartments of euchromatin (or A compartment) and heterochromatin (B compartment) (Erez Lieberman-Aiden *et al.*, 2009) (*Figure 3B*). Although

compartments and TAD frontiers overlap well, the mechanism behind compartment formation seems to be different from those involved in the formation of other 3D structures (Hildebrand and Dekker, 2020). As a matter of fact, the exact drivers of compartmentalisation are still the subject of intense research, as a comprehensive understanding of the mechanisms involved has not yet been achieved. The prevailing view tends towards a major role for phase separation in the segregation of compartments as biomolecular condensates that aggregate affine factors together to separate different types of chromatin states. Thus, interactions between chromatin sharing the same state would simply be physically easier to achieve, A-A and B-B interactions hence becoming more probable than A-B interactions. The drivers of these interactions are for example HP1 in constitutive heterochromatin (Strom *et al.*, 2017; Zenk *et al.*, 2021); or in the case of euchromatin, Brd2 has been shown to facilitate compartmentalisation through its binding to acetylated histones, while being antagonized by Cohesin and Brd4 (Xie *et al.*, 2022). It is not clear yet whether the processes involved in transcription microphases are the same as those driving large-scale phase separation of euchromatin (Boija *et al.*, 2018; Henninger *et al.*, 2021; Xie *et al.*, 2022). Likewise, depletion or accumulation of loop-extruding Cohesin also shows an antagonism between compartment strength and Cohesin's formation of TADs, that would hence be two distinct mechanisms in interplay with each other (Haarhuis *et al.*, 2017; Rao *et al.*, 2017; Schwarzer *et al.*, 2017; Wutz *et al.*, 2017; Vian *et al.*, 2018). Overall, compartments and TADs are formed on the same chromatin polymer yet through different mechanisms, and their balance may participate in the correct regulation of chromatin functions. Of importance is the dissimilarities between the mechanisms observed for compartmentalisation in mammals compared to those observed in other model organisms. In the case of *Drosophila*, the full overlap between TAD borders and compartment positions was petitioned as evidence that compartments (named in those articles "compartmental domains") formation in this organism may drive TAD formation (Rowley *et al.*, 2017, 2019). This conclusion may highlight the previously discussed differences in Cohesin roles in different organisms, however caution must be taken as it may be due also to simple differences in resolution obtained in HiC. Indeed, the *Drosophila* genome being ~20 times smaller than the Human genome, much higher sequencing depth can be achieved, hence rendering comparisons somewhat vague and difficult.

Surprisingly, a perfected study of compartments using ultra-deep sequencing was published recently (Harris *et al.*, 2023), with the conclusion that Human compartment organisation is actually finer than previously thought. In this study, they point out that single active enhancers for instance can localise in the active compartment, even though their surrounding is in an inactive compartment. In the same vein, they make the unprecedented observation that genes may have their TSS and TTS in separate compartments. These observations dispute the previous conclusion about compartment

formation, where they would be mere clusters of TADs with similar transcriptional status, and will likely bring to further studies regarding the mechanisms behind this phenomenon. Similarly, it was recently proposed that a dedicated compartment appears to cluster DNA repair foci (Arnould et al., 2023). This in turn highlights the long-standing question of the role of compartmentalisation. In fact, is the clustering of chromatin into compartments important for its functions? Does it impact gene expression, DNA replication, and if so why and how? Those questions are hard to tackle, as it is intricate to uncouple compartmentalisation from other 3D genome structuration mechanisms, and because many factors predicted to participate in compartmentalisation processes are also involved directly in gene expression. Hopefully, the coming years will bring new insight regarding the drivers of compartments formation, and the reasons for their existence.

II. Non-coding RNAs, their regulation and their many roles

Aside of messenger RNAs transcribed by the RNA polymerase II, a multitude of RNAs are not translated into proteins, hence named non-coding RNAs (ncRNAs). RNA polymerase I and III transcribe all of the ribosomal RNA (rRNA), transfer RNA (tRNA) and other small nuclear RNAs (snRNAs), composing ~70-90% of all RNAs in cells. Other numerous ncRNAs are transcribed by Pol II, or derive from Pol II transcripts, having a myriad of roles, notably in RNA processing, maturation and regulation (St.Laurent, Wahlestedt and Kapranov, 2015; Jarroux, Morillon and Pinskaya, 2017). The sheer amount of different RNA classes and the diversity of their roles make an encyclopaedic review irrelevant to the work developed here. Hence this manuscript will focus on subclasses of Pol II-transcribed ncRNAs, and in particular enhancer RNAs (eRNAs) and PROMoteur uPstream Transcripts (PROMPTs). Even though upon discovery these two RNA species were seen as mere by-products of transcription due to their relatively short half-life, of the order of minutes (Schwalb *et al.*, 2016), they have since been the subject of much attention. Several eRNAs have notably been associated with cancers and diseases (Wang *et al.*, 2022), as well as with the regulation of gene expression and 3D genome organisation (Chen *et al.*, 2023; Syed & Hon, 2021). Most intergenic RNAs, sometimes referred to as long non-coding RNAs (lncRNA, for lengths above 200bp) are in fact eRNAs, while PROMPTs and “conventional lncRNAs” are comparatively a minority (Hon *et al.*, 2017). In this manuscript, the term lncRNA will not be used, as the distinction between previously described lncRNA compared to eRNAs is now thoroughly disrupted by their shared properties and roles. These canonical RNA classes are now seen as the two extremes of a continuous spectrum (Syed and Hon, 2021). This chapter aims at defining these ncRNAs, their production, and discussing the roles that they have already been attributed.

A. Transcription & regulation of PROMPTs and eRNAs

Transcription at regulatory regions and its diversity: PROMPTs and eRNAs

eRNAs were first described by de Santa *et al.* and Kim *et al.* in 2010 as a widespread bi-directional transcription at active enhancers. Since, these observations of mostly short, rarely spliced, 5'capped but non-polyadenylated, low-abundance eRNAs have been generalised to all active enhancers in all cell types, both in vertebrates and invertebrates (*Figure 4A*) (Djebali *et al.*, 2012; Andersson *et al.*, 2014; Rennie *et al.*, 2018). Moreover, a global correlation was demonstrated between the level of eRNA expression and the level of expression the genes they regulate (Andersson *et al.*, 2014). Interestingly, a non-negligible fraction of eRNAs tend to be longer, poly-adenylated and more stable, resulting in a bias towards one strand that produces more steady-state transcripts compared

to the other strand (*Figure 4B*) (Syed and Hon, 2021). This bias may reveal that, although eRNAs do not appear to have a conserved sequence, enhancer TSSs may be subject to evolutionary selection to favour accumulation of a specific eRNA (Jin *et al.*, 2017; Syed and Hon, 2021). A study of eRNAs role in myogenesis even observed a population of “bi-stable” eRNAs, where both strand of the enhancer produced more stable eRNAs (Zhao *et al.*, 2019). This heterogeneity in eRNAs may also highlight different properties and functions for different subsets. For example, enhancers producing longer and more stable eRNAs tend to have higher H3K4me3 deposition, as well as more CTCF, TF and Pol II binding, perhaps underlining different roles for these enhancers (Gil and Ulitsky, 2018; Syed and Hon, 2021). In turn, the effects of these various kinds of eRNAs may be different, as will be discussed further in section B of this chapter.

PROMPTs, also referred to as Upstream Antisens transcripts (uaRNAs) or Promoter Antisens transcripts (PAS RNAs), were first described as a widely distributed divergent transcription at promoters (Seila *et al.*, 2008; Core *et al.*, 2008; Preker *et al.*, 2008), similarly to eRNAs. They are transcribed on the promoters of genes, on the opposite strand and direction from the mRNA (*Figure 4C*). The fundamental difference between eRNAs and PROMPTs is still blurry however, as they share the same properties of early termination, low stability and splicing, as well as relatively few polyadenylations (Almada *et al.*, 2013; Core *et al.*, 2014; Chen *et al.*, 2016; Rennie *et al.*, 2018). In fact, this comes down to the still debated difference between promoters and enhancers, as both can initiate transcription, and promoter can participate in other genes’ regulation, in the same manner as enhancers (Engreitz *et al.*, 2016; Fukaya, Lim and Levine, 2016; Mikhaylichenko *et al.*, 2018). This advocates for a common property of transcription initiating regions to be intrinsically bi-directional, and to then be selected to favour the transcription, stability and maturation of one transcript over the other (Andersson *et al.*, 2015; Chen *et al.*, 2016; Jin *et al.*, 2017). Although the stability and length of PROMPTs have been associated with the proximity of other promoters (Chen *et al.*, 2016), there does not seem to be a particular population to be separated from the rest as it can be done for eRNAs. Potential effects of PROMPTs transcription would thus be mediated solely by the transcription of short-lived RNAs, and not by a sub-population of higher steady-state PROMPTs (Syed and Hon, 2021).

Of note, eRNAs and PROMPTs matrices appear to be depleted in U1 motif, while enriched in Poly(A) Sites motifs compared to mRNAs, explaining their low splicing and higher rate of early transcriptional termination, as well as their higher Exosome degradation (*Figure 4*) (Almada *et al.*, 2013; Ntini *et al.*, 2013; Syed and Hon, 2021). The global epigenetic marking of eRNAs and PROMPTs is also similar, with deposition of H3K4me1 and H3K4me3 next to the Nucleosome Depleted Regions (NDR). However, the ratio between those marks may be different, as PROMPTs templates exhibit low H3K4me1/H3K4me3 ratios, similar to genes; whereas that of eRNAs display higher

H3K4me1/H3K4me3 ratios. In the case of PROMPTs, it was demonstrated by Churchman & Weissman in 2011 in Yeast that the deacetylase Rpd3S helps prevent their transcription. The integrator complex on the other hand has been shown to be central to eRNAs biogenesis by enabling their termination and release from Pol II (Lai *et al.*, 2015). In the same vein, the ubiquitous Pol II-associated factor SPT6 was shown in Human cells to regulate ncRNA transcription, including both PROMPTs and eRNAs, by ensuring correct H3K36me3 deposition only on coding genes, as well as enabling the recruitment of the Integrator complex to chromatin to favour early termination of ncRNAs (Nojima *et al.*, 2018).

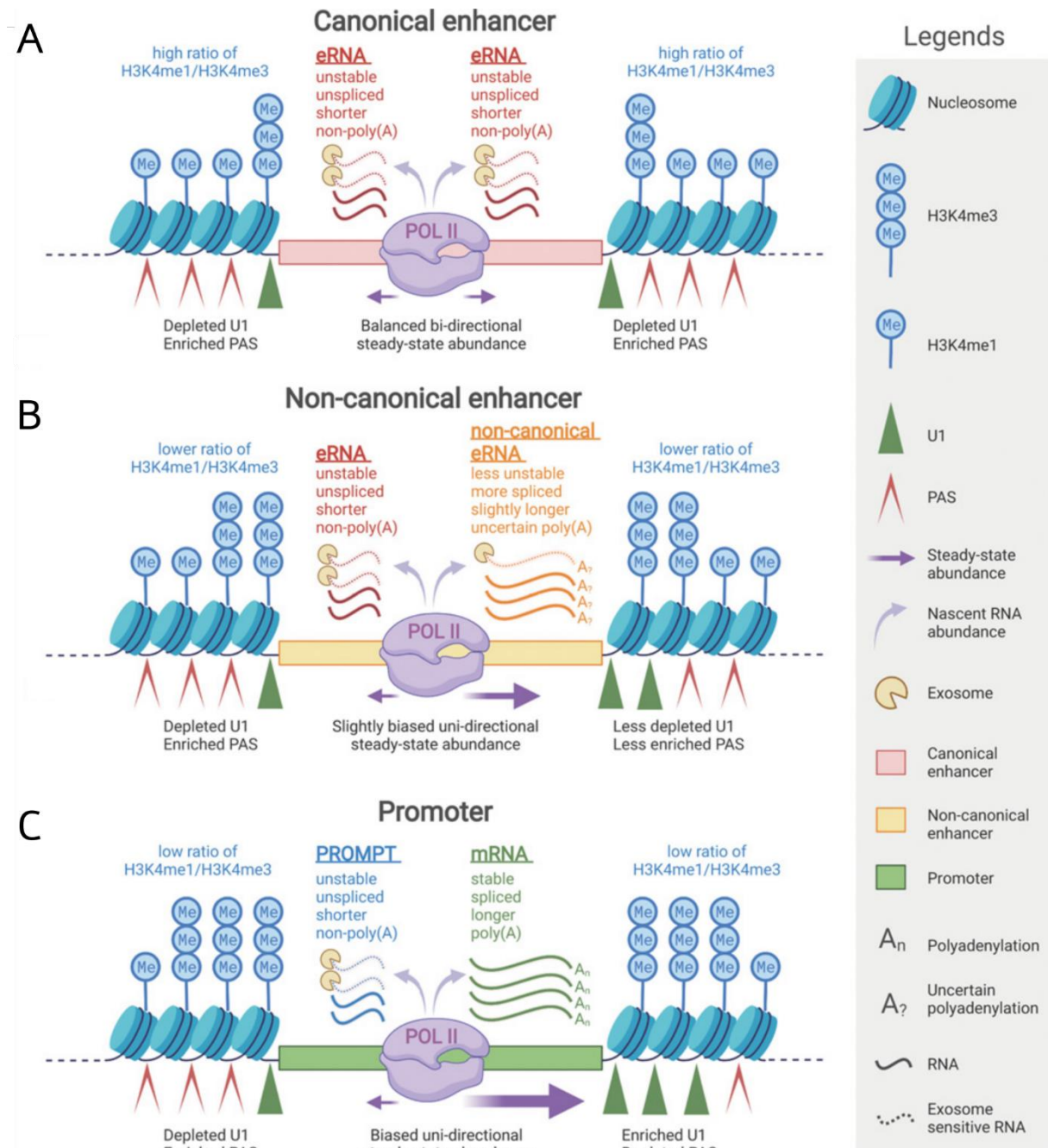


Figure 4: Transcriptional & epigenetic contexts of eRNAs and PROMPTs Schematics representing eRNAs & PROMPTs, their properties and transcription contexts. Canonical eRNAs, similarly to PROMPTs, have low steady states due to their fast degradation via the exosome, while non-canonical eRNAs can have longer half-lives and accumulate. Figure adapted from Syed & Hon 2021.

Conversely, DNA-RNA hybrids (R-loops) were shown to promote the transcription of ncRNAs, both at promoters and enhancers (Pefanis *et al.*, 2015; Nojima *et al.*, 2018; Tan-Wong, Dhir and Proudfoot, 2019). The production of eRNAs and PROMPTs may in turn favour the formation of R-loops, as eRNAs and PROMPTs are able to form such structures (Chen *et al.*, 2023). Still, most of the possible impacts and functions of these differences in histone PTMs and ncRNAs regulation are yet to be explored (Syed and Hon, 2021; Yang, 2022).

PROMPTs & eRNAs targeted degradation by the Exosome

When first described, both eRNAs and PROMPTs were suggested to be simple by-product of transcription due to their high turnover and low steady-state level. Both are degraded by the nuclear Exosome machinery (*Figure 5A*), usually swiftly after transcription, which explains their usually very low accumulation levels in conventional RNAseq data. However, the Exosome itself has poor specificity and cannot process RNA secondary structures (Kilchert, Wittmann and Vasiljeva, 2016). To bring specificity, the Nuclear EXosome Targeting complex (NEXT) (*Figure 5B & C*) and the Poly(A) eXosome Targeting complex (PAXT) are the bearers of binding and helicase properties (Lubas *et al.*, 2011; Meola *et al.*, 2016; Gerlach *et al.*, 2022), even though the exact way by which they target specific RNAs is yet to be determined. The NEXT complex has been demonstrated as being necessary to enable the proper degradation of most PROMPTs and eRNAs, while the PAXT complex tends to target longer, processed and poly-adenylated targets (Lubas *et al.*, 2011; Meola *et al.*, 2016). Of note, studies on the secondary structures of these ncRNAs are scarce, thus the requirement for the NEXT complex to enable their degradation may be due to the formation of secondary structures, such as hairpins. The RNA helicase Mtr4 in particular, present in both PAXT and NEXT, is necessary for the degradation of both PROMPTs and eRNAs, but also of some introns, which may again highlight the formation of secondary structures to be unfolded by this RNA helicase (Lubas *et al.*, 2011; Meola *et al.*, 2016; Gerlach *et al.*, 2022).

Of interest for this manuscript, it is yet to be understood whether this degradation of PROMPTs and eRNAs can be regulated, or if it is a simple “all you can degrade” mission for the exosome and its targeting complexes. In fact, the studies performed to date were mostly limited to the observation of accumulations of ncRNAs upon degradation of either an Exosome component or a targeting complex factor (Lubas *et al.*, 2011; Kilchert, Wittmann and Vasiljeva, 2016; Meola *et al.*, 2016; Gerlach *et al.*, 2022). In the hypothesis of a function for eRNAs and PROMPTs, it would make sense indeed for the degradation of these ncRNAs to be dynamically regulated or even locally prevented, so as to enable their accumulation and enhance their effect in precise contexts. Interestingly, two contradicting studies recently reported that m⁶A methylation of eRNAs and PROMPTs appear to either protect them against early termination and degradation (Xu *et al.*, 2022), or facilitate their degradation via YTHDC1

binding (Liu *et al.*, 2020), revealing an unexpected layer of regulation that may be context-dependent. Supporting this view, the recently published structure of the NEXT complex has revealed regulatory configurations preventing processing of the bound RNA, especially via the ZCCHC8 subunit (Figure 5B & C) (Gerlach *et al.*, 2022). Furthermore, the NEXT complex interacts with various factors outside of the Exosome, notably via ZC3H18 (Winczura *et al.*, 2018), which leaves ample space to elucidate the molecular mechanisms by which different classes of ncRNAs are recognized and targeted for degradation by the Exosome.

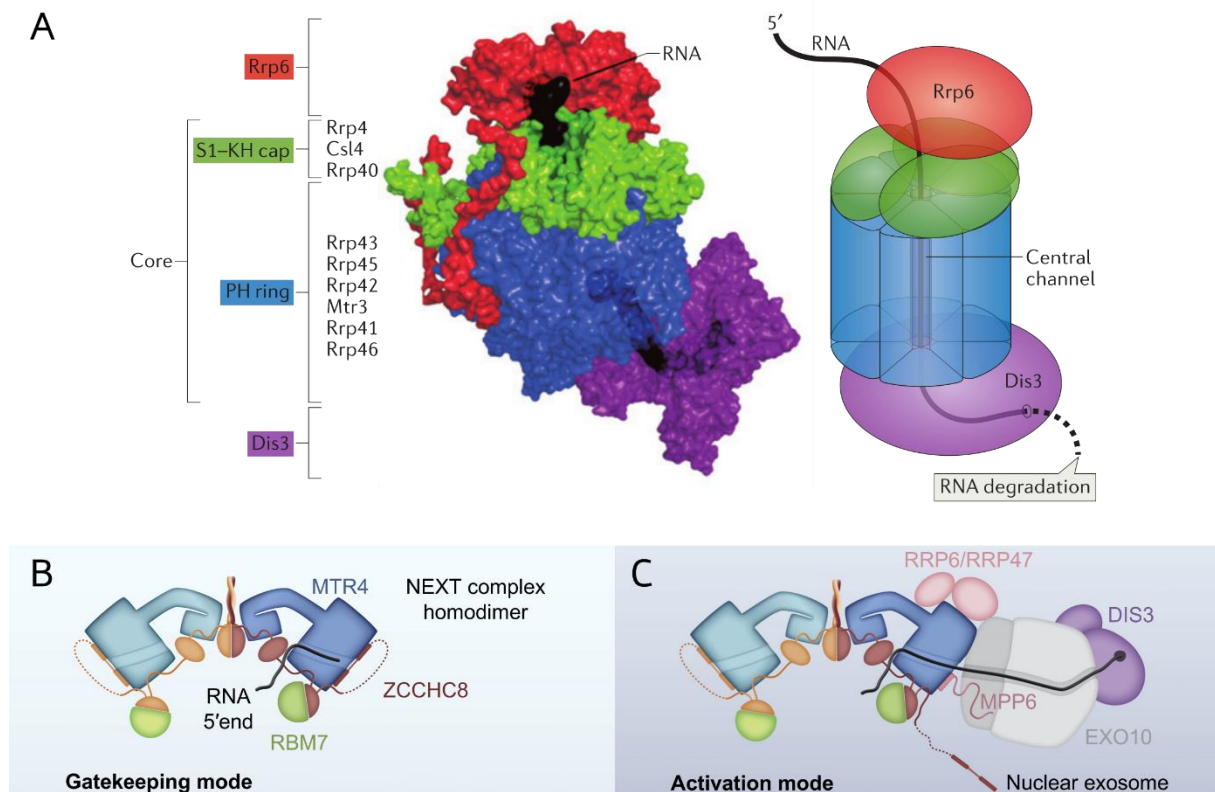


Figure 5: Structure & schematics of the exosome and its nuclear targeting complex NEXT **A.** Structure of the nuclear exosome complex, with a simplified view of its Dis3-catalytic conformation, adapted from Kilchert 2016. **B. & C.** Schematics of the NEXT complex in its gatekeeping mode and processive mode, adapted from Gerlach 2022. For EM structure of the NEXT complex with and without exosome, see Gerlach *et al.* 2022. No proper structure of the PAXT complex is available to date.

B. Roles for PROMPTs and eRNAs in functional mechanisms

The aforementioned assumption that eRNAs and PROMPTs are useless transcriptional by-products, based on their rapid degradation, has been overturned by the discovery of regulatory roles, particularly for specific eRNAs (Figure 6A). However, the question of the role of the actual RNAs compared to that of their transcription itself was raised. In fact, if the function of enhancer transcription is to keep them transcriptionally active and their chromatin open, then perhaps only their

transcription is really necessary, and the eRNAs produced are not – hence their degradation. Yet, other purposes for eRNAs, as well as for PROMPTs, that necessitate their binding to other factors have been described, as discussed below. An overcast shadow nevertheless is whether these functions are common to all eRNAs – and thus possibly common to PROMPTs – or if specific subgroups sustain the observed effects. It may be expected for example that less stable ncRNAs act locally, as their fast degradation would not allow otherwise; while more stable ncRNAs could possibly act in *trans*, i.e. further away from their loci of transcription. As many ncRNAs are predicted to be 3-dimensionally structured, it is also possible that for some their structuration, rather than their sequences, is necessary to mediate their roles (Schwalb *et al.*, 2016; Harrison and Bose, 2022).

ncRNAs direct impact in Pol II transcription

Various studies have demonstrated that ncRNAs can mediate gene expression regulation. Specific eRNAs in particular were shown to participate to the regulation of their enhancers' target genes, even though the mechanisms by which they do so remain obscure. Most of these studies used a depletion, or more sporadically an accumulation, of the target eRNA to demonstrate a deregulation of either their gene of interest (i.e. not necessarily local) or the genes neighbouring the transcribed enhancer (Lai *et al.*, 2013; Lam *et al.*, 2013; Li *et al.*, 2013; Mousavi *et al.*, 2013; Hsieh *et al.*, 2014; Maruyama, Mimura and Itoh, 2014; Alvarez-Dominguez *et al.*, 2017; Kim *et al.*, 2018; Abdalla *et al.*, 2019). In 2013, another study by Melo *et al.* performed a Gal4-MS2 targeting of specific exogenously transcribed eRNAs to a promoter to demonstrate their capacity to enhance gene expression, with still little understanding of the mechanism by which these eRNAs could impact gene regulation. Since, a method with the same goal was developed using CRISPR-dCas9 to target ncRNAs (Shechner *et al.*, 2015; Sigova *et al.*, 2015; Yang *et al.*, 2021; Hou and Kraus, 2022). Another study also observed that RNA-exosome-regulated ncRNAs, i.e. mostly PROMPTs and eRNAs, were influential in the regulation of super-enhancers activity (Pefanis *et al.*, 2015). Finally, hnRNPL, a protein involved in RNA alternative splicing and transcriptional regulation has been shown by several reports to bind eRNAs to regulate gene expression levels. It may do so in different ways and at several levels of the transcription process, because one of these reports observed repressing effect of the ncRNA-hnRNPL interaction, while the two others found up-regulating effects (Li *et al.*, 2014; Atianand *et al.*, 2016; Zhao *et al.*, 2019).

Interestingly, the structures of ncRNAs may participate in their regulatory functions, particularly to provide specificity to proteins binding. This hypothesis, although poorly investigated, is supported by the global lack of sequence conservation among most ncRNAs, and especially for the less stable PROMPTs and eRNAs. Regrettably, only few eRNAs/PROMPTs structures have been

experimentally characterised (Yang *et al.*, 2021; Gorbovytska *et al.*, 2022; Hou and Kraus, 2022; Przanowska *et al.*, 2022), even though many have been predicted to be spatially folded (Schwalb *et al.*, 2016). In a few cases however, the structures of eRNAs and PROMPTs have been directly associated with their role (Yang *et al.*, 2021; Przanowska *et al.*, 2022). In the case of ^{DDR}eRNA in myogenesis for example, it is striking to see different domains, and their respective unpredicted but experimentally-confirmed structures, regulate separate sets of genes, through different mechanisms (Przanowska *et al.*, 2022).

Aside of these rather general association of gene regulation with modifications of ncRNA levels, certain eRNAs were observed to directly impact Pol II pausing (*Figure 6A5*). They do so by acting as decoys and/or destabiliser for the Negative ELongation Factor (NELF), hence facilitating pause release into productive elongation (Schaukowitch *et al.*, 2014; Shii *et al.*, 2017; Gorbovytska *et al.*, 2022). This action was demonstrated *in vitro* to be mediated through direct interactions between eRNAs and NELF-A/E subunits, yet with no effects of structures nor sequences, needing only a certain length of RNA (>200 bases) and unpaired guanosines to be effective. This argues for a role of the RNA themselves, regardless of motif recognition (Gorbovytska *et al.*, 2022), in tune with the previous demonstration that NELF-E binds a large range of RNA with no sequence or structure preference (Yamaguchi *et al.*, 2002). In the same vein, the pause-release kinase p-TEFb was shown to be bound and activated by an eRNA, thus promoting Pol II phosphorylation and productive elongation. In this case, the structure-motif of the eRNA was shown to be influential, as it substitutes the well characterised p-TEFb inhibitor RNA 7SK (Zhao *et al.*, 2016). It remains to be determined whether these fascinating mechanisms can be generalised to all eRNAs, as suggested for NELF (Gorbovytska *et al.*, 2022). Nonetheless, it is clear that at least eRNAs and maybe other ncRNAs have regulatory roles on Pol II pause release and thus gene expression, both dependent and independent on their structures-motif.

The accumulation of ncRNAs have also been reported to participate in the formation of transcriptional hubs, to favour and regulate transcription at all steps. Chiefly, transcription initiation has been proposed to be substantially impacted by the concentration of ncRNAs in the transcriptional condensates. In effect, transcriptional condensates are formed mainly via the Intrinsically Disordered Domains of TFs (Boija *et al.*, 2018; Cho *et al.*, 2018; Sabari *et al.*, 2018; Shrinivas *et al.*, 2019). Interestingly, these IDRs are canonical RNA-binding domains in many proteins involved in transcriptional condensate formation (Roden and Gladfelter, 2021). In 2021, Henninger *et al.* proposed a model, where the synthesis of relatively small and quickly degraded eRNAs and PROMPTs would, at low concentrations, facilitate the formation and efficiency of the initial transcriptional condensates. Upon induction of the transcriptional burst, the massive accumulation of mRNAs would

then concentrate too many negative charges, making the phase crumble on itself, and thus terminating the phase of hyper-efficient transcription (*Figure 6C*). Although this model, supported by other studies (Maharana *et al.*, 2018; Boeynaems *et al.*, 2019; Garcia-Jove Navarro *et al.*, 2019; Sharp *et al.*, 2022), does not unite well with the previously mentioned transcriptional model of separate phases for initiation and productive elongation, it provides a brilliantly simple mechanistic framework to explain transcriptional bursts, a still obscure mechanism due to its dynamic aspect and the sheer number of factors involved (Tunnacliffe and Chubb, 2020). This simplistic view however has been recently challenged by quantitative live-imaging and super-resolution that argue for enhancer transcription to antagonize molecular crowding of transcription factors, suppressing transcriptional bursting of the linked genes, while certain developmental enhancers may be structurally optimized to support both coding and non-coding transcriptional activation (Hamamoto *et al.*, 2023). They argue that enhancer function would be tuneable through non-coding transcription and hub formation to perform gene regulation. Importantly, this study uses synthetic constructions where the effect of non-coding transcription is to displace TFs bound at the enhancer, perhaps explaining the contradicting effect observed, as they point out that only few natural – and mostly developmental – enhancers have such configurations. Likewise, this study lacks a proper distinction of the effects of non-coding transcription versus the accumulation ncRNAs themselves on which the other studies cited above were focused, possibly confusing the conclusions attained.

Interestingly, the m⁶A modification of eRNAs has been advanced as potentiating the formation of transcriptional condensates through its phase-separating reader YTHDC1, in turn activating gene expression (Lee *et al.*, 2021). It remains to be investigated whether this effect is really directly exerted by the m⁶A modification, for which the eRNAs would be mere carriers, or if it is really the eRNAs that are effective, in which case the m⁶A modification may play its part in regulating the stability of the eRNAs, as previously reported (Liu *et al.*, 2020; Xu *et al.*, 2022). To conclude, a brilliant study by Quinodoz *et al.* in 2021 proposed a general model based on the new technique they developed named RD-SPRITE, which enables the simultaneous observation of RNA and DNA spatial repartition genome-wide. This model proposes that ncRNAs globally act as seeds and recruiters of various protein factors to form specific and localised genome 3D structures, so as to regulate and shape gene expression, DNA contacts, but also heterochromatin assembly (*Figure 6D*). The great asset of this model, as well as that of the rather complementary transcriptional condensate model exposed above, is that they finally provide a general rule and framework to understand how ncRNAs can influence gene transcription.

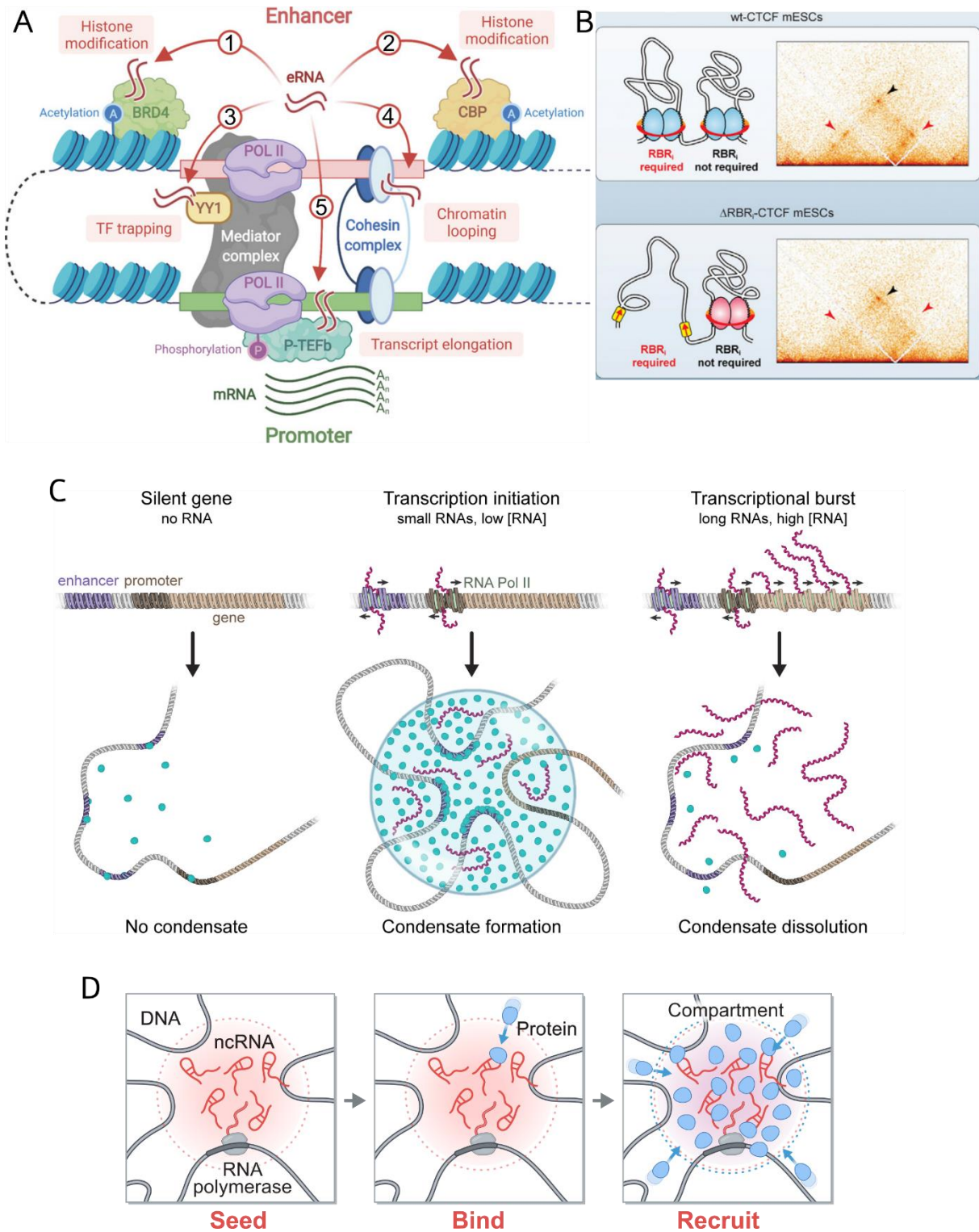


Figure 6: Non-coding RNAs roles in transcription, chromatin & genome spatial organisation **A.** Schematic summarizing of some of the roles identified for eRNAs to date, adapted from Syed & Hon 2021 **B.** Model and example proposed by Hansen et al. 2019 regarding effects of the RNA-binding domains of CTCF **C.** Model of transcriptional condensate formation to initiate transcriptional burst, proposed by Henninger et al. 2021 **D.** ncRNAs seeding model proposed by Quinodoz et al. 2021, which enables recruitment of various proteins to form transcriptional compartments.

ncRNAs influence on chromatin regulation

Indirectly involved in gene regulation, various chromatin changes mediated by ncRNAs have also been examined, either at the level of histone occupancy and histone PTMs, or at the level of transcription factors recruitment. First of all, one cannot avoid mentioning the extensively studied Xist ncRNA, which mediates X-chromosome Inactivation in Mammals by triggering gene silencing and structural reorganisation through the recruitment of a vast array of chromatin modifiers (Loda and Heard, 2019). However, this mechanism being very peculiar and restricted to the X chromosome, we will rather interest ourselves in the less-studied roles of the multitude of other ncRNAs. A 2012 study by Schubert et al. for example observed that the Densation factor 31 (Df31) binds various ncRNAs in *Drosophila* to establish and maintain open chromatin domains. Along similar lines, enhancer transcription, independently of the eRNAs, was demonstrated to precede and facilitate H3K4 methylation at de novo enhancers (Kaikkonen *et al.*, 2013). Interestingly, this argues that eRNA transcription is not a simple consequence of enhancer function, but rather is necessary to enable the activity of certain enhancers. This was later supported by a case study demonstration that the transcription of a distal enhancer, and particularly one of the eRNAs produced, influenced histone marks deposition at both enhancers and promoters (Pnueli *et al.*, 2015). In this instance, simple knockdown of only one strand of the bi-directional eRNA is sufficient to cause a loss of H3K27ac at the enhancer and replacement by H3K27me3, coinciding with H3K4me3 loss at the promoter. This result argues rather for a direct role of the transcripts themselves, rather than their transcription. In complete support, H3K27me3 deposition by PRC2 was later shown to be antagonized by RNAs, even though PRC2-RNA interactions are necessary for proper targeting of the silencing complex to chromatin. The current model explaining this paradox is that JARID2, a specific subunit of PRC2, timely inhibits RNA-binding when necessary, in order to enable PRC2 catalytic activity (Cifuentes-Rojas *et al.*, 2014; Beltran *et al.*, 2016; Long *et al.*, 2020; Mangiavacchi, Morelli and Orlando, 2023). Although most of these studies do not distinguish different RNA species, at least three studies demonstrate that specific ncRNAs, localised away from the gene they regulate, also mediate their regulatory role through the PRC2-RNA interaction mechanism (Rinn *et al.*, 2007; Grote *et al.*, 2013; Teo *et al.*, 2022). This would allow for a model in which many, if not all ncRNAs would participate in regulating H3K27me3 deposition to setup and maintain transcriptional programs. In the way of histone PTMs being regulated by ncRNAs, PROMPTs have also recently been shown to regulate the transcription of numerous genes, in part through the stabilization at promoters of KDM4B and KDM4C, the demethylases of H3K9me3 (Yang *et al.*, 2021). Similar to histone methylation, acetylation has also been shown to be regulated by eRNAs in particular. The Histone Acetyltransferase (HAT) complex CBP/p300 binds locally-transcribed eRNAs directly, targeting and stimulating histone acetylation at enhancer, and hence transcription of

target genes (*Figure 6A2*) (Bose *et al.*, 2017; Carullo *et al.*, 2020; Hou and Kraus, 2022). This mechanism is further supported by the eRNA-binding property of BRD4, a Bromodomain-containing protein that binds to histone acetylation to enhance transcription. This direct eRNA binding again increases BRD4 binding to acetylated histones and thus its recruitment to enhancers to function as a transcriptional cofactor (*Figure 6A1*) (Rahnamoun *et al.*, 2018). Altogether, it is clear that eRNAs and their transcription empower transcriptional regulation by snowballing the deposition of several histone PTMs favourable to transcription while prohibiting deposition of repressive ones, thus making chromatin more permissible.

ncRNAs may also control transcription through the recruitment of various transcription factors and cofactors. Several Bromodomains-containing factors, analogous to BRD4 as cited above, have been shown to bind eRNAs, even though a complete description of the genes they may regulate through ncRNA binding is still lacking (Rahnamoun *et al.*, 2018). The ncRNA ARIEL has also been shown to enhance the activation of an oncogenic transcriptional program by a specific TF named TAL1, through the recruitment of the Mediator complex (Hao Tan *et al.*, 2019). Similarly, other oncogenic TF, such as c-JUN (Shii *et al.*, 2017) and NF- κ B (Spurlock *et al.*, 2017; Huang *et al.*, 2018), were documented to depend on eRNAs for their recruitment to regulate gene expression. Finally, Sigova *et al.* demonstrated in 2015 that the ubiquitously-expressed YY1 is also affected by eRNAs, as they facilitate its recruitment and retention on chromatin at enhancers (*Figure 6A3*). Surprisingly however, general accumulation of eRNAs and PROMPTs through exosome depletion caused global decrease of YY1 at enhancers. The authors interpret this result as a titration of YY1 by the excessive accumulation of eRNAs, keeping too many YY1 molecules away from chromatin and with no possibility to bind anything but the eRNAs.

Complementary to the previously described mechanisms, the R-loops formed by eRNAs and PROMPTs can enhance transcription by serving as promoters through the opening of the DNA helix (Tan-Wong, Dhir and Proudfoot, 2019). Of note, the exact mechanisms afoot to regulate ncRNA-mediated formation of R-loops are still under investigation, as will be developed in chapter III of this manuscript. Aside from this mechanism, the ncRNA named KHPS1 may control, via the formation of R-loops, various chromatin regulators such as CTCF, p300 and an eRNA, to target the regulation of a gene, SPHK1 (Blank-Giwajna, Postepska-Igielska and Grummt, 2019). This rather complex mechanism highlights the possibility for ncRNAs to form DNA-RNA hybrids as a mean of targeting to specific regions, as will be discussed in depth in the next section regarding spatial genome regulation by ncRNAs. Cooperatively, it was suggested that the still controversial m⁶A modifications of a subset of eRNAs and PROMPTs forming R-loop may participate in their targeting to degradation via YTHDC1 (Liu *et al.*, 2020). This modification would in turn inhibit TFs recruitment and histone PTM deposition, allowing for the abrogation of the otherwise self-sustaining mechanism of transcription activation.

To conclude, substantial evidence demonstrates that ncRNAs enable the setup and enforcement of transcriptional programs by regulating chromatin. In particular, the deposition of histone PTMs and the recruitment of TFs facilitated by eRNAs is influential to regulate various genes and processes. However, it is still unclear in most cases whether these observations can be generalised, or if ncRNAs act in a context specific manner to regulate limited numbers of genes. In this sense, it will be of great interest to develop an exhaustive view of the transcription factors and histone modifiers that necessitate or are regulated by ncRNAs.

ncRNAs influence on 3D genome organisation

As exposed before, the 3D structuration of the genome also has substantial impacts on various cellular processes, and in particular gene expression. In this sense, ncRNAs have also been involved in regulating this spatial architecture by binding and recruiting various architectural factors (*Figure 6A4*). Cohesin was first shown to bind several eRNAs in 2013 by Li *et al.*, with implications in gene regulation by promoting Cohesin binding at enhancers. This finding was later supported by studies on the previously mentioned ^{DDR}eRNA, which interacts with the Cohesin complex and is required for its proper loading and activity to regulate the *Myogenin* gene (Tsai *et al.*, 2018; Przanowska *et al.*, 2022). Another biochemical investigation also demonstrated that the SA1 and SA2 subunits of Cohesin directly bind not only single stranded RNA (ssRNA), but also double stranded RNA (dsRNA) and R-loops, opening the door to many potential effects of ncRNA transcripts on Cohesin recruitment and its mediation of 3D interactions (Pan *et al.*, 2021). A more indirect study of YY1 demonstrated that this architectural TF enables the formation of certain E-P interactions by anchoring Cohesin in an eRNA-binding dependent manner, because YY1 is regulated by eRNA interactions, as detailed in the previous section (Sigova *et al.*, 2015; Weintraub *et al.*, 2017). Similarly, Mediator binding to ncRNAs was demonstrated to promote chromatin looping (Lai *et al.*, 2013). Given the previously cited ongoing debate regarding Mediator role in chromatin looping, we may speculate that the RNA-Mediator interaction effect on chromatin looping also involves Cohesin. In the same vein, the observed modification in chromatin looping stabilized by eRNAs in the regulation of *Nanog* and *Dppa3* by super-enhancers is supposed to involve Cohesin, although it was not properly addressed in this case.

The other major regulator of chromatin 3D organisation in vertebrates, CTCF, has also been shown in Human and Mouse to bind RNAs. First reported in a case study with the SRA ncRNA, the RNA-binding property of CTCF was then proposed as necessary for its insulating activity at few loci (Yao *et al.*, 2010). The RNA-binding regions of CTCF were later identified, and demonstrated to bind a large variety of transcripts, including ncRNAs, with surprisingly no particular sequence or motif preference (Saldaña-Meyer *et al.*, 2014, 2019; Hansen *et al.*, 2019). Even more intriguing, although CTCF is well

known to be required for virtually all TAD borders and loops in vertebrates, the deletion of its RNA binding regions (RBR) impacted only about half of all loops and TADs mediated by CTCF (*Figure 6B*) (Hansen *et al.*, 2019). This result is mirrored by the specific disruption of CTCF binding to DNA only in certain loci that depend on the deleted RBR (Hansen *et al.*, 2019; Saldaña-Meyer *et al.*, 2019). Two classes of loops could even be distinguished, between RBR dependent loops that are abolished upon RBR deletion, and RBR independent loops that are not affected. A new report from Harris *et al.* in 2023 has shown using ultra-deep sequencing that the RBR of CTCF may actually distinguish different types of loops. Hence, the more diffusive loops would depend on CTCF's RBR, while the punctuate loops remain unaffected, perhaps due to different mechanisms behind their formation. The model proposed for the influence of RNA on CTCF is that RNAs would facilitate the oligomerization and clustering of CTCF, hence maintaining loops, perhaps somewhat independently from Cohesin (Hansen *et al.*, 2019; Saldaña-Meyer *et al.*, 2019). Yet a proper demonstration of this speculative model remains to be generated, as it is still obscure how exactly RNA binding by CTCF would cause or favour loop formation in some cases and not others. The lack of specific sequence motif is just as puzzling, as the current understanding would mean that any RNA can be bound by CTCF, making it conceptually hard to appreciate context specificities. The potential of RNA structures mentioned above may bring a first layer of specificity, but remains to be demonstrated as impacting RNA-CTCF interactions. Yet another layer of comprehension may be derived from two studies of HOTTIP, a ncRNA involved in the setup of the leukemic transcription profile (Luo *et al.*, 2019, 2022), as well as two studies of Jpx, a ncRNA involved in Xist expression for X-chromosome inactivation (Sun *et al.*, 2013; Oh *et al.*, 2021). In these studies, they observed that these two ncRNAs are targeted widely in the genome, via the formation of sequence specific R-loops in the case of HOTTIP, to regulate a large transcriptional program. To mediate this regulation, HOTTIP seemingly helps recruits CTCF and Cohesin to form TAD boundaries where it forms R-loops, in a sort of targeted enhancement CTCF boundary formation (Luo *et al.*, 2022). These findings, if generalised to other ncRNAs, could mean that CTCF does not need to bind only free RNAs, but rather chromatin-tethered RNAs that would retain or empower CTCF at certain loci, hence the specificity. Oppositely, Jpx seem to act as a competitor for CTCF binding, displacing it from low-affinity sites, thus favouring the formation of other loops and possibly TAD frontiers, although a complete description of the effects of Jpx on TADs and Cohesin recruitment remains to be produced. Regarding ncRNAs, their formation of R-loops and their binding to chromatin proteins may thus be shown influential to regulate CTCF architectural purposes in precise contexts.

In conclusion, although comprehensive models are still lacking to explain how architectural proteins are impacted by ncRNAs, it is clear that such a layer of regulation does exist in Mammals. It is however surprising that all of the studies to date are limited to Human and Mouse models. In fact,

there is yet no studies on the effects of ncRNAs on the various architectural proteins present in *Drosophila*, aside from a study by Lei and Corces in 2006 that argue for an RNA-dependent interaction between CP190 and an RNA helicase, Rm62. However, this study is severely over-reaching in its speculative mechanistic model as to the potential roles of RNAs. In fact, their data is restricted to genetic experiments that do not present evidence for RNA-binding-mediated effects, and biochemical experiments that only observe CP190-Rm62 interactions in co-immunoprecipitation, that are diminished upon RNase A treatment. Another study from the same author also tried to demonstrate that mRNAs may participate in the formation of insulator complexes, but with yet again little evidence to support the claim further than the interaction of CP190 and Su(Hw) with RNAs (Matzat, Dale and Lei, 2013). To summarize, the potential effects of ncRNAs on 3D genome organisation in *Drosophila* are still vastly unexplored, and left to be speculated from observations made in Mammals, which is an error-prone task in regards of the poor homology of architectural factors mentioned previously.

Deleterious effects of ncRNAs

Even though the evidence and models presented in this manuscript point mostly toward physiological roles for ncRNAs, it is important to note that their deregulation may also have or develop deleterious effects. eRNAs, for example, have been associated with multiple human diseases, ranging from cancers to neurodegenerative disorders and metabolic diseases (Wang *et al.*, 2022). Evidently, the dysregulation of all of the priorly mentioned mechanisms in which ncRNAs regulate gene expression and chromatin is likely to be consequential in diseases development. In this sense, the numerous SNPs found in regulatory sequences outside of coding regions that are associated with various diseases could be in part due to such dysregulations, on top of the first-thought-of TFs motif mutation (Harrison and Bose, 2022). Indeed, as at least several ncRNAs mediate their effects via their structure, they could be easily tempered by pinpoint mutation overturning their essential structures. In turn, such mutations in promoters and enhancers and the regulatory motifs present in their transcribed matrices have been suggested as an evolutionary pool to shape new transcripts, coding or not (Chen *et al.*, 2016; Espinosa, 2017). There is no doubt however that the transcription of previously inexistent transcripts has the potential for adverse effects, particularly if it escapes degradation. Likewise, the accumulation of transcripts escaping degradation will probably cause the accumulation or R-loops, which have shown times and again to be a source of Double Strand Breaks and mutations, on top of gene dysregulations (Skourti-Stathaki and Proudfoot, 2014; Allison and Wang, 2019; García-Muse and Aguilera, 2019).

Altogether, the various genome-wide and loci-specific data produced to date argue for both promiscuous, local effect of ncRNAs and their transcription, as well as more long-range effects of specific and more stable ncRNAs. Even so, most studies do not examine the stability of the ncRNA they study, complexifying interpretations as to whether they exert their effect locally or distantly. Importantly, the myriad of mechanisms through which ncRNAs impact chromatin, 3D genome organisation and transcription support a general and prevalent role for these sometimes-rare ncRNAs, albeit not in all contexts nor by all ncRNAs. It may be expected that the diversity of mechanisms in which eRNAs act, for example, is comparable to the variety in enhancers and transcription factors bound to them. Clearly, a comprehensive model of how ncRNAs impact genome regulation is yet to be achieved, even though still-debated preliminary models have been proposed. The development of new methods to study ncRNAs in their native contexts while decoupling their effects from that of their transcription will be a major challenge and prospect to understand the multitude of ncRNAs double-edged-sword functions.

III. RNA helicases and RNA degradation

RNA helicases are a large group of enzymes with a common capacity to process and unwind RNA double-strand structures using ATP hydrolysis as a motor. The RNA helicases activities are essential throughout all RNA metabolism processes, from transcription, splicing and translation, to shaping RNA structures. Thus, RNA helicases are primordial to RNA degradation, as many RNA processing factors can only cleave single strand RNA, and are blocked or slowed by folded RNA. Likewise, although some proteins can directly cleave the RNA in RNA:DNA hybrids, these structures usually need to first be unwound by RNA helicases, so that the RNA can be degraded. Nowadays, the term “RNA helicase” for this family of protein is seen as restrictive, as the vast diversity of RNA helicase functions makes for unwinding-incapable RNA helicases. These usually act as sensors or binders of RNAs, such as the eIF4A-III DEAD-box helicase, that is part of the exon-junction complex and serves as an assembly platform for other factors.

A. DEAD-box & Ski2-like RNA helicases and their roles

RNA helicases are divided into super-families (SF) depending on the conserved features and sequence motifs they have. I will focus on the SF2, which comprises the DEAD box family and the Ski2-like helicases that are of interest for this manuscript. All RNA helicases of the SF2 share a common core motif, comprising two RecA-like domains (*Figure 7A*), that possess the ATP-catalytic site and unwinding activity, although with low basal efficiency and no sequence preference on its own (Donsbach and Klostermeier, 2021). The addition of subfamily-specific domains then usually stimulates the unwinding activity, by modulating the RNA helicase core conformation, or its interaction with RNA. Similarly, substrate specificity is often controlled via N- or C-terminal domains, or through cofactors (Donsbach and Klostermeier, 2021). Interestingly, many RNA helicases have been shown to participate in phase-separation, either from having IDRs themselves, or through interactions with cofactors that do. They are thought to be global regulators of these condensates, as their conformational changes can influence the formation of separated phases (Hondele *et al.*, 2019). Additionally, their regulation of RNA localisation and/or stability may participate in this establishment of transient condensates, as RNA is an important component of phase separating condensates, as discussed before (Hondele *et al.*, 2019). The study of RNA helicases roles can be difficult, as many of them possess several roles in either separate or intricate processes. In the same vein, some of them

are redundant, in part or all of their functions, and simple knockdown of one RNA helicase is not always sufficient to observe all of its effects. Thus, to decipher a given RNA helicase effect on gene expression for example can be challenging, because they may participate at several steps of gene expression, from transcription regulation to mRNA splicing and maturation, as well as mRNA transport, translation and decay.

The DEAD box family (DDX) is the largest RNA helicase family, with a broad variety of functions among its 40 members. Generally, DDX helicases are non-processive, meaning that they only unwind local secondary structure, and do not translocate along the RNA to continue unwinding. They are characterised by a conserved Asp-Glu-Ala-Asp (DEAD) amino-acid motif. Upon binding of both ATP and RNA by either Rec-like domain, the DDX helicases undergo a conformational switch, from an open state to a close one, in which both Rec-like domains interact with each other and with ATP and RNA (*Figure 7B*) (Sengoku *et al.*, 2006). This conformational switch in itself is essential to catalyse RNA duplex separation, as it disrupts a few base pairs and destabilizes the RNA duplex, subsequently facilitating the release of the first strand of RNA (Sengoku *et al.*, 2006). ATP hydrolysis then enables the separation of the Rec-like domains to release the second strand of RNA and recycle the helicase for another round (*Figure 7B*) (Linder and Jankowsky, 2011; Donsbach and Klostermeier, 2021). The sequential release of both strands, in contrast to a simultaneous release, is seen as preventing re-annealing to avoid spurious cycles (Rudolph and Klostermeier, 2015). The numerous and diverse roles played by DDX helicases in RNA metabolism – and their multitude of physiological consequences – exclude a relevant exhaustive review of their functions here. Yet the reader may be interested in Linder and Jankowsky, 2011 or Donsbach and Klostermeier, 2021 for reviews of DDX helicases biochemical properties and functions, and in Cargill, Venkataraman and Lee, 2021 for their impact in genome stability. Also, the meeting report by Andrisani *et al.*, 2022 give a synthetic summary of the implications of a few selected DDX helicases in health and disease. The two members of the DDX family that are of interest for this manuscript will be described further in the next section.

The Ski2-like family on the other hand is particularly involved in RNA processing, via its splicing, maturation or degradation. Two notorious members of this family, Mtr4 and Ski2, are paralogs involved in RNA degradation via the exosome, with the first being restricted to the nucleus, and the second acting mostly in the cytoplasm. Also, Brr2 and Slh1, respectively involved in pre-mRNA splicing and mRNA transcription and translation, have been extensively studied (Johnson and Jackson, 2013). Their structure share similarities with the DEAH box helicases, their 2 conserved Rec-like domains forming a ring-like structure with the C-terminal ratchet-helix and WH-domains (*Figure 7A & C*) (Johnson and Jackson, 2013). These domains are supplemented by an arch domain and a KOW domain in the case of Mtr4 (*Figure 7C*). Unlike DDX helicases, the Ski2-like helicases unwind RNA by

translocation along one strand of the RNA duplex in a 3'->5' orientation (*Figure 5B & C*) (Patrick *et al.*, 2017; Donsbach and Klostermeier, 2021). These helicases translocate in single-nucleotides steps by nucleotide-induced movement of the Rec-like domains, which causes in turn a shift in the ratchet helix and a destabilisation of the interaction with ssRNA (*Figure 7C*) (Büttner, Nehring and Hopfner, 2007). This manuscript will further review Mtr4 in the following segment, focusing particularly on the functions of this proteins, and its impacts in genomic processes.

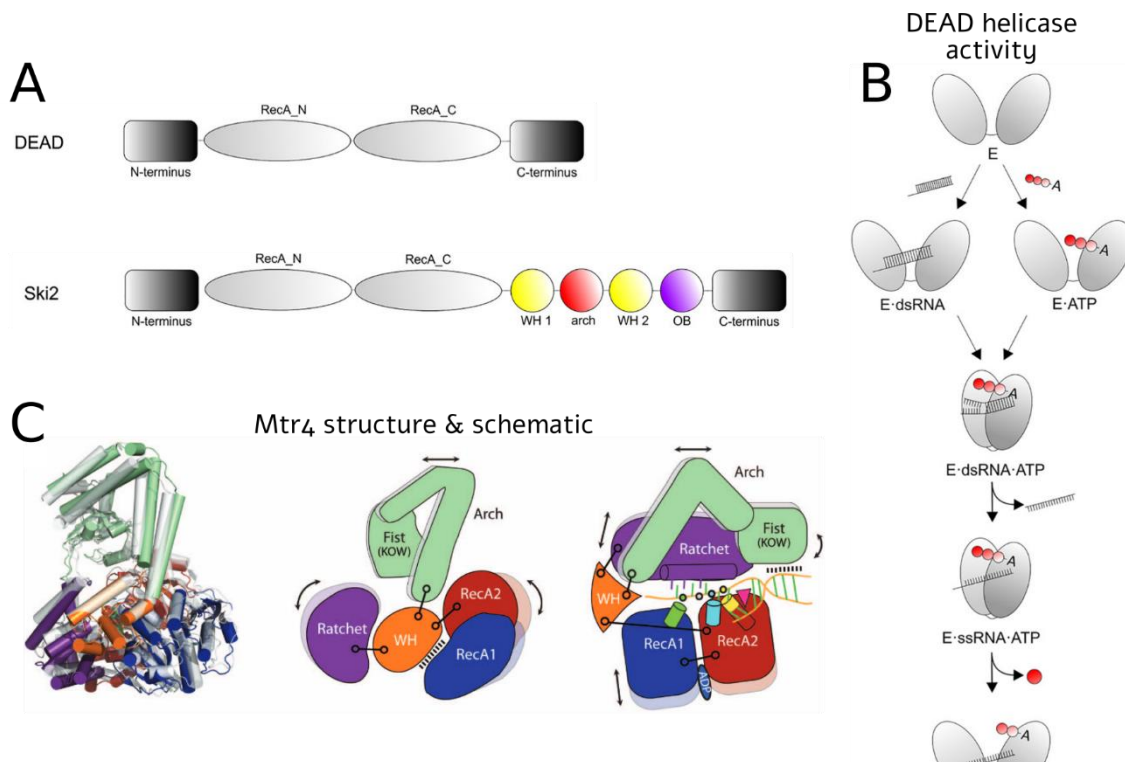


Figure 7: SF2 RNA helicases structures **A.** Schematic domains of DEAD-box & Ski2-like RNA helicases, adapted from Donsbach and Klostermeier, 2021 **B.** Schematic of the conformational changes of DEAD-box RNA helicases during their catalytic activity, adapted from Donsbach and Klostermeier, 2021 **C.** Electron Microscopy & schematic of the changes of Mtr4 structure upon binding of RNA and ATP/ADP. Covalent linkages between each domain are represented as black lines, adapted from Johnson and Jackson, 2013.

B. RNA helicases found in insulator complexes

This study was initiated by the discovery of 8 RNA helicases in insulator complexes using mass spectrometry. Given the sparse literature available for most of these helicases, and for feasibility concerns, the experiments were focused on 3 of these RNA helicases: Mtr4, Rm62 (homolog of DDX5, also named p68 in Human), and Abstrackt (homolog of DDX41 in Human). The literature available to date on these 3 RNA helicases will be developed in following paragraphs, focusing particularly on their

function that could pertain to chromatin, transcription and genome organisation; i.e. the functions of insulators.

Mtr4

Mtr4 is a strongly conserved enzyme of the Ski2-like family, and has been demonstrated to have major biological roles, notably in Ribosomal RNA (rRNA) biogenesis, and in RNA degradation. Although active alone *in vitro* (Olsen and Johnson, 2021), Mtr4 associates with less-conserved mutually-exclusive cofactors to form different complexes, that target specific RNA species, in both the nucleolus and the nucleus (Thoms *et al.*, 2015; Falk *et al.*, 2017; Lingaraju *et al.*, 2019; Schmid and Jensen, 2019). Its first description as an RNA helicase came from the study of rRNA biogenesis in Yeast, where it was demonstrated to be needed for proper formation of the 5.8S rRNA 3' end. Subsequently, Mtr4 was demonstrated to form nucleolus specific complexes involved in rRNA metabolism in both Human and Yeast (Schmid and Jensen, 2019).

The most studied role of Mtr4 is its targeting of RNAs for degradation. Via the NEXT and PAXT complexes notably, Mtr4 facilitates exosome-mediated degradation of ncRNAs (*Figure 5*), as well as the degradation of various coding transcripts. Several studies have clearly shown in Human that NEXT and PAXT regulate the degradation of both PROMPTs and eRNAs (Lubas *et al.*, 2011, 2015; Meola *et al.*, 2016; Gerlach *et al.*, 2022). Similarly, NEXT is necessary for the 3'-end trimming of snRNAs, snoRNAs and histone encoding mRNAs (Lubas *et al.*, 2011, 2015). The current view of mRNA degradation via the exosome targeting complex is a competitive mechanism between RNA decay and RNA export. In this model, the nuclear mRNA 5' Cap-Binding Complex form mutually-exclusive interactions with either proteins of the nuclear RNA export pathway, or proteins that act as bridges with NEXT or PAXT, thus controlling the fate of mRNAs to avoid potential translation of unwanted or defective mRNAs (Schmid and Jensen, 2019). Its roles in different contexts and complexes have made difficult the study of Mtr4 structures, as reviewed in details by Olsen and Johnson in 2021, yet the complete structure of the exosome-adaptor NEXT complex has recently been reported (*Figure 5B & C*)(Gerlach *et al.*, 2022). As discussed previously, this recent study of Mtr4 structure in the NEXT complex reveals that the whole Mtr4-ZCCHC8-RBM7 complex acts as a dimer, and has two possible conformations: 1) gatekeeping (*Figure 5B*), where ZCCHC8 blocks RNA processing by Mtr4; and 2) activation (*Figure 5C*), where the RNA is actively processed and extruded from Mtr4 helicase core. This argues for a regulation of RNA degradation by Mtr4 complexes. Supporting this view, NRDE2 is also involved in restricting Mtr4 binding to specific transcripts, supposedly to protect mRNA against degradation (Wang *et al.*, 2019). Incidentally, the NEXT complex may also have a minor participation in the degradation of some spliced-out introns, as RBM7 interacts with splicing factors (Lubas *et al.*, 2015; Falk *et al.*, 2016).

Surprisingly little is known in *Drosophila* about Mtr4, the exosome and its targeting complexes. In fact, most studies on this subject have been performed in Yeast and Human, and the conservation was thus inferred for *Drosophila*. This gap in knowledge is considerable, as although the exosome itself is well conserved and can be expected to function similarly, the targeting complexes are poorly conserved outside of the multitasking Mtr4 (Schmid and Jensen, 2019). The degradation of PROMPTs and eRNAs by the exosome has been observed (Rennie *et al.*, 2018), but the functionality of exosome-targeting complexes remains to be demonstrated. In fact, a single study produced in *Drosophila* has involved Mtr4 itself in the degradation of small RNAs in a case study of the Integrator-complex-mediated cleavage of a particular gene (Tatomer *et al.*, 2019). Hence, it is still obscure how non-coding RNAs such as PROMPTs and eRNAs are regulated in *Drosophila*, even though it is inferred from other species. Of great interest for this manuscript however, the exosome itself has been suggested to be recruited to insulator complexes in *Drosophila*, hypothetically to regulate RNAs involved in these genome regulatory complexes (Lim *et al.*, 2013). Yet, as discussed in section II. B., such an influence of RNAs on *Drosophila* insulators remains to be properly demonstrated.

To date, Mtr4 has never been associated with roles in the regulation of transcription itself, nor in regulation of chromatin and genome spatial organisation. That may be surprising, as ncRNAs and their stability have been associated with various aspects of these processes, as discussed in chapter II. It can be hypothesized for example that defaults in ncRNAs degradation, either general or specific, may influence their capacity to regulate chromatin processes. Such an observation was made by Sigova *et al.* in 2015 in Human by total exosome depletion that modulated YY1 binding through eRNAs, yet the impact of targeted regulation of PROMPTs and eRNAs by Mtr4 has not been examined.

Rm62 – DDX5

Rm62 is a *Drosophila* DDX helicase, conserved from Yeast under the name of Dbp2, to Human in the form of DDX5 (previously named p68, and Dmp68 or Lip in *Drosophila*). In both Human and *Drosophila*, it has a close paralog named DDX17 in Human (CG10077 in *Drosophila*, predicted but unstudied) that have both redundant functions and paralog-specific functions (Xing, Ma and Tran, 2019). DDX5's RNA helicase activity was demonstrated early (Hirling *et al.*, 1989), yet the manner in which it unwinds RNA duplexes was subject to some debates. Some argued that the length it can unwind *in vitro* (162bp, considerably higher than the 12-20bp unwound by other DDX helicases), as well as its need for 3' overhangs and directionality meant it would be processive (Rössler, Straka and Stahl, 2001); a feat unprecedented in the DDX family. Yet further work in various models contravene this conclusion and directionality of unwinding, although DDX5 is indeed capable of unwinding

unusually long duplexes (Xing, Ma and Tran, 2019). As for many RNA helicases, DDX5 has been the subject of numerous investigations in various physiological contexts, showing pleiotropic effects, of which the exact causes are hard to confirm. Hence it has been attributed roles in a wide range of cancers and diseases, with sometimes little understanding of the underlying mechanisms (Fuller-Pace and Moore, 2011; Cheng *et al.*, 2018; Xing, Ma and Tran, 2019).

This RNA helicase has been well studied in its participation to mRNA processing at various steps. First, in both Human and *Drosophila*, it has been associated with splicing, as it was identified as a component of the spliceosome in RNAi screens and proteomic analysis (Xing, Ma and Tran, 2019). In particular, DDX5 is involved in alternative splicing, as it controls the inclusion of numerous exons by tuning the folding of specific transcripts and the recruitment of splicing regulators (Camats *et al.*, 2008; Dardenne *et al.*, 2012, 2014; Lee, Wang and Rio, 2018; Terrone *et al.*, 2022). DDX5 has also been shown to participate in miRNA maturation (Suzuki *et al.*, 2009; Dardenne *et al.*, 2014). In *Drosophila*, Rm62 was also suggested to participate in mRNA release from chromatin after transcription, thus impacting mRNA export as well as gene de-activation after heat shock or Ecdysone induction (Buszczak and Spradling, 2006).

Rm62, and particularly its Human homolog DDX5, have been described as coregulators of transcription factors by associating with chromatin modifiers and remodelers in a variety of manners (Giraud, Terrone and Bourgeois, 2018). DDX5 direct interaction with several TFs and signalling pathways is a clear and relatively well-explored process through which it participates in gene regulation (reviewed by Fuller-Pace and Moore, 2011; Giraud, Terrone and Bourgeois, 2018). More interesting to this study however is the reported interaction of DDX5 with chromatin modifiers such as CBP-p300 (Rossow and Janknecht, 2003; Warner *et al.*, 2004), Su(Var)3-9 (Boeke *et al.*, 2011) HDAC1 (Wilson *et al.*, 2004) and PRC2 (Zhang *et al.*, 2016). Along the same lines, DDX5 and Rm62 have both been shown to be associated with insulator factors, thus possibly participating in genome spatial organisation (Lei and Corces, 2006; Yao *et al.*, 2010), but also pre-mRNA processing (Terrone *et al.*, 2022). Considering the previously described influences of ncRNAs in the regulation of chromatin processes, transcription and 3D genome organisation, a striking link can be made, wherein DDX5 aforementioned roles in these processes may be in part due to its regulation of ncRNAs and their structures. Such a link is supported by several studies that have implicated DDX5-ncRNA interactions in various mechanisms. The most studied of these interactions is with SRA (Steroid nuclear Receptor Activator), an alternatively spliced ncRNA which generates several isoforms, some of which even code for the translation of a protein (Giraud, Terrone and Bourgeois, 2018). SRA and its interaction with DDX5 have been reported to regulate transcription in various physiological and pathological processes (Giraud, Terrone and Bourgeois, 2018). Importantly, it remains mostly unclear in these studies whether

it is one, the other, or the combinations of both the ncRNA and the RNA helicase that mediate the effects. The current understanding of these mechanisms involving SRA, as it is with other such ncRNAs that interact with DDX5, is that the ncRNA may act as a multimodal platform for the recruitment and assembly of various factors and complexes to regulate gene expression and chromatin. Thus, the main role of DDX5 would be that of a chaperone, shaping the ncRNAs structures in order to facilitate the formation of the correct RNA-proteins interactions depending on the cellular context (Giraud, Terrone and Bourgeois, 2018).

Possibly related to these mechanisms is the recently reported capacity of DDX5 to resolve R-loops (Mersaoui *et al.*, 2019; Yu *et al.*, 2020) and G-quadruplexes (Wu *et al.*, 2019). Indeed, as discussed in Chapter II., R-loop in particular have been involved in multiple cases in chromatin and gene expression regulation, as well as genome organisation and stability. DDX5 regulation of R-loops may thus be yet another mechanistic layer through which it may achieve various effects.

As the reader may find evident at this point, the study of Rm62/DDX5 roles is difficult in its intricacy. These various roles are apparently impossible to uncouple, as they all rely on the unwinding capacity of the RNA helicase. It remains to be investigated whether certain domains or isoforms of the helicase, or even cofactors may enable specificity for various roles. However, it may be interesting to point out that precisely, this may all be a common story that has been, until now, observed from different points of view. This DDX helicase would thus be a sort of master-regulator of coding and non-coding RNA through their shape and binding to various factors. This could explain many of the roles in gene expression, alternative splicing and chromatin regulation with the common basis of RNA structures, yet leave ample gaps of knowledge to be filled by the identifications of the exact RNA motifs and structures that are recognized and regulated by DDX5 in all the different contexts.

Abstrakt – DDX41

Abstrakt (Abs) is a still putative RNA helicase from the DDX family, whose enzymatic helicase capacity has not been demonstrated to date, while its homolog in Human is named DDX41. Generally, the cellular and molecular functions of this enzyme are poorly understood, even though it is involved in innate immunity (Jiang *et al.*, 2017), and considered as a tumour suppressor and as necessary for viability and cell growth in plants, invertebrates and vertebrates (Jiang *et al.*, 2017; Liu and Imai, 2018; Tsukamoto *et al.*, 2020; Weinreb *et al.*, 2021). In *Drosophila*, Abs was reported almost 25 years ago as essential at all stages of development, its loss causing defects in various developmental processes, and notably RNA localisation (Irion and Leptin, 1999; Schmucker *et al.*, 1999). A few years later, it was

shown to control asymmetric divisions through the regulation of the protein Insc, with a notable direct interaction with its mRNA, but with no understanding of the underlying molecular mechanism (Irion *et al.*, 2004). In Human on the other hand, DDX41 has been reported to associate with the spliceosome to ensure proper splicing of certain exons, thus exerting a tumour suppressor role in myeloid neoplasms (Polprasert *et al.*, 2015). This interaction with spliceosome components and effects in splicing were later recapitulated in other contexts (Ma *et al.*, 2022; Weinreb *et al.*, 2022); nevertheless, DDX41 exact function in the spliceosome remains unknown.

Of greater interest to this manuscript, two studies demonstrated recently that DDX41 is capable to unwind R-loops *in vitro*, and antagonizes their formation *in vivo* (Mosler *et al.*, 2021; Weinreb *et al.*, 2021). These studies finally provide biochemical evidence of DDX41 catalytic activity, and describe the effects it can have at genomic and cellular levels. In fact, this regulation of R-loops by DDX41 is suggested to be particularly important to avoid Double Strand Break caused by hybrids at numerous promoters (Mosler *et al.*, 2021). This could be related to the previously mentioned roles of R-loops at promoters and enhancers, implicating DDX41 in gene regulation. It is important to note however that, as observed for most DDX helicases, the roles of DDX41 are probably not limited to a single activity, and its interactions with RNAs remain to be investigated. To conclude, most of DDX41 properties and functions remain unexplored, even though its implication in numerous cancers and diseases alludes to potentially interesting roles in gene expression and splicing regulations.

To summarize, the 3 RNA helicases discussed here have been heterogeneously studied. Mtr4 has been relatively well described in its functions to regulate RNA degradation via the exosome, while Rm62/DDX5 seems to function through RNA folding, in both splicing and gene regulation. DDX41, although poorly studied, has promising pleiotropic effects, and can resolve R-loops, similarly to DDX5. Altogether, we can note a recurrent role, sometimes only hypothetical, for these factors in the indirect regulation of gene expression. Although the last decades have been centred on the effects of proteins to catalyse effects, these RNA helicases may rather be intermediaries to regulate, disable or enable the effects of non-coding RNAs – among others. Such effects are indeed being reported in various context, particularly to regulate genome structure, organisation and expression; yet most underlying mechanisms and factors involved in ncRNA regulation still remain unknown. These open questions regarding ncRNAs could open new avenues to complete and adjust the current models of genome and chromatin organisation in particular. Hence, the project developed in this manuscript aims at deciphering such roles of ncRNAs and their regulations by RNA helicases. Prompted by the discovery of the association of RNA helicases with insulator complexes in *Drosophila*, we investigate their effects

as potential insulator cofactors. Particularly, we study their influence in gene expression and ncRNAs regulation, focusing on their impacts in insulator complexes formation. We then try to decipher how the changes observed in gene expression and insulator proteins recruitment may translate into 3D genome organisation genome-wide. Finally, we suggest that ncRNAs may enact changes wrought by RNA helicase depletion, as supported by the literature exposed above. Altogether, our data demonstrate that eRNAs and PROMPTs levels, are tightly controlled and may regulate certain insulator complexes to preserve 3D genome organisation and gene expression programs. This study provides a first genome-wide examination of the regulation of ncRNAs in *Drosophila*, and studies their effects on transcription control and 3D genome folding.

RESULTS

When starting this project, we looked for new cofactors of insulator complexes. To this end, I made use of previously produced Mass Spectrometry data (Lhoumaud *et al.*, 2014; Liang *et al.*, 2014) of Beaf32 CoIP. In this dataset, we were surprised to see that at least 8 RNA helicases were detected. This unexpected discovery encouraged us to investigate whether RNA helicases could be cofactors of the insulator complexes. When researching the literature for these various helicases, we observed that some were very poorly described, while others had more defined roles, in RNA degradation and RNA folding for example. Along the same line, we found another RNA helicase that had been described before as a potential interactor of insulator complexes. Thus, because of feasibility concerns, we decided to focus on fewer of these RNA helicases to characterise, choosing the three for which we had a better understanding of the potential roles relating to genome regulation and insulator functions.

The results presented in this manuscript will be divided into two chapters. First, I will present the results and analysis obtained for Mtr4 regarding its control of non-coding RNAs degradation, its influence on gene regulation, and its impact on genome organisation through insulators. In this chapter, I will try to develop and complete our current model, in which the stability of eRNAs & PROMPTs in particular is regulated by Mtr4, and may in turn control distinct insulator complexes' formation. We notably observe that 3D contacts, transcriptional condensates and insulator condensates behaviours are influenced by Mtr4, but not insulation. To do so, we make use of RNA-seq, ChIP-seq and HiC, and we complement the analysis of these datasets with microscopy and CRISPR-dCas9 experiments.

In the second chapter, I report the experiments and analysis performed to study the two other RNA helicases studied here, namely Rm62 and Abstrakt. This chapter is more descriptive, as we currently do not have a proper model to explain the effects observed. In fact, we observe little to no effect on the regulation of ncRNAs by these two RNA helicases, rendering difficult the conception of an RNA-based model to explain the effects observed. Using the same methods as for Mtr4, we note effects of both RNA helicases on gene expression, insulator complexes formation, insulation frontiers and 3D genome interactions, yet surprisingly not in the same contexts. In tune with the exposed literature, we can hypothesize other means through which Rm62 and Abs could mediate their effects, notably R-loop resolution and RNA folding. Nevertheless, this study does not provide a proper examination of such underlying mechanisms.

Chapter 1 - Mtr4-controlled non-coding RNAs may regulate gene expression through insulator complexes in *Drosophila*

Mtr4 interacts with insulators and regulates ncRNAs and genes

When searching for potential new insulator cofactors using previously produced mass-spectrometry data from Beaf32 CoIP, we surprisingly found peptides for Mtr4, an RNA helicase. To confirm this interaction, we performed several CoIP experiments to reproduce this result. Due to lack of IP-qualified antibody commercially available for Mtr4, we produced a stable cell line expressing an exogenous HA-tagged version of Mtr4. We thus confirmed that Mtr4 and Beaf32 do interact, as a fraction of Mtr4 is reproducibly immunoprecipitated with Beaf32 (*Figure 1A*). We also found a lower interaction between CP190 and Mtr4 (*Figure 1B*), that may be accounted for by indirect precipitation through the previously demonstrated Beaf32-CP190 interaction. As Beaf32 is a DNA-binding factor, we questioned whether its binding to DNA was necessary to observe its interaction with Mtr4. Similarly, as Mtr4 is an RNA helicase that can be bound to RNA in a stalled conformation, we questioned whether Mtr4 interaction with Beaf32 was dependent on RNA. Performing co-immunoprecipitation in the presence of DNase I to degrade DNA, or MNase to degrade both DNA and RNA (single and double strand), we demonstrate that the interaction between Mtr4 and Beaf32 is independent of both DNA and RNA presence when using complete cell lysates (*Figure 1B*).

Hence, we began to study the potential effect of Mtr4 as a cofactor of insulators. We performed depletions of Mtr4, Beaf32, as well as dCTCF (*Figure S1 A & B*). To validate Mtr4 depletion which was not confirmed by RTqPCR, we used western blot to confirm complete protein depletion. Also, we further demonstrated functional depletion by measuring the accumulation of a previously described PROMPT, namely that of Cg25C (hereafter P-Cg25C) (Rennie *et al.*, 2018), whose degradation is enhanced by Mtr4 (*Figure S1A*). Upon Mtr4 degradation, we observed a fold change of 5-20 for P-Cg25C.

We then used strand-specific RNAseq to probe gene expression changes upon Mtr4, Beaf32 and dCTCF KD. Mtr4 depletion reproducibly upregulates 1461 and downregulates 474 genes when using as threshold an absolute Log2FC of 0,7 (~1,62 FC) (*Figure 1C*). Beaf32 and dCTCF KD however only cause few deregulations, as expected from single depletion of insulator factors from previously described results in the literature (Rowley *et al.*, 2017) (*Figure 1C*). As Mtr4 is known to participate in ribosome RNA maturation, we validated that our depletion did not modify cell-cycle dynamics using FACS, despite a slightly lower number of cells in mitosis (*Figure S1C*). Clustering all genes deregulated

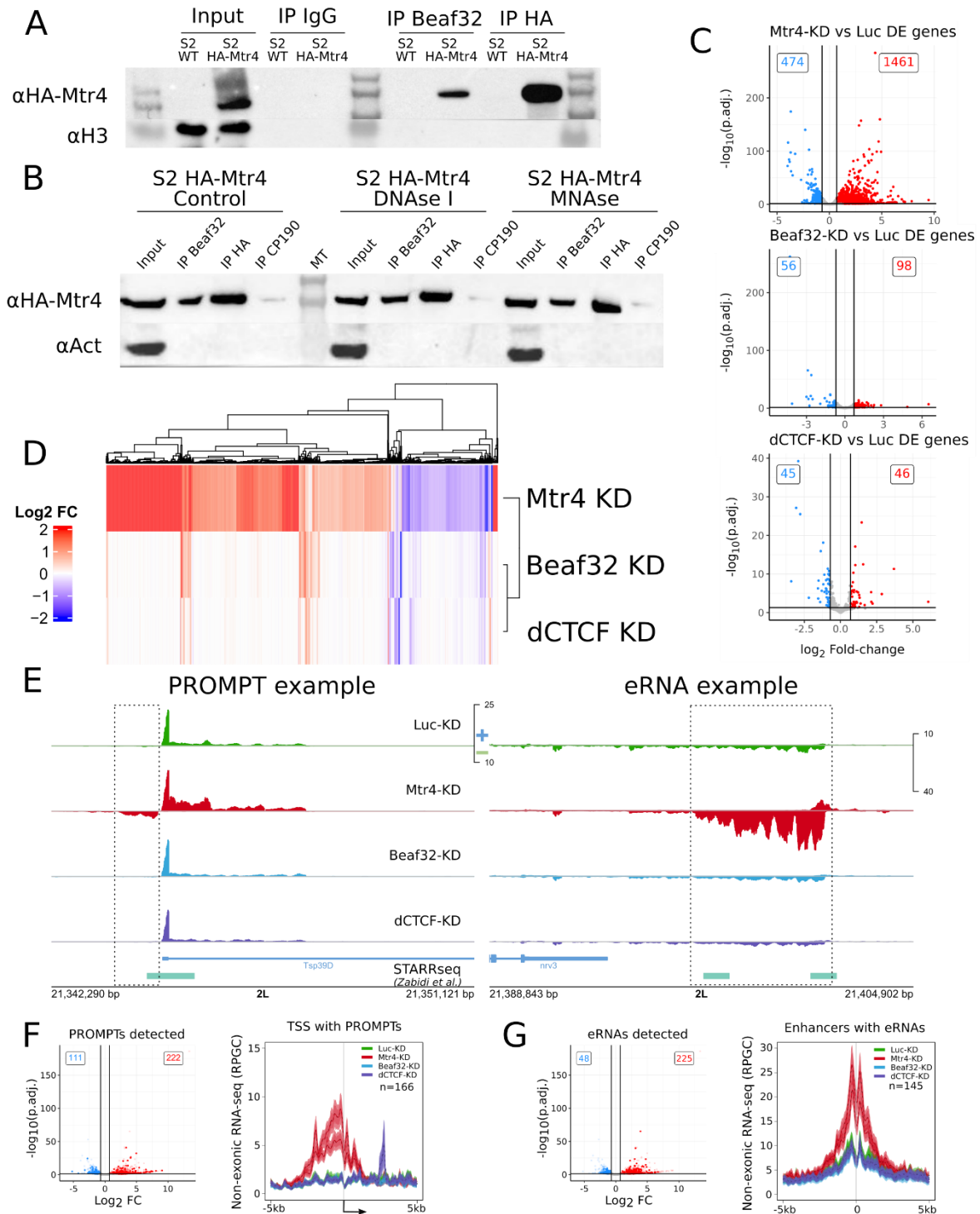


Figure 1: Mtr4 interacts with insulators and regulates ncRNAs and genes. **A.** Western Blot of CoIP showing interaction between HA-Mtr4 and Beaf32 in WT S2 cells and a stable S2 cell line expressing exogenous N-terminal HA-tagged Mtr4. Input material is lysate before IP, 2/3 the amount is deposited compared to IP samples. Irrelevant Rabbit IgG serves as control. **B.** Western Blot of CoIP showing levels of interactions between HA-Mtr4 and Beaf32 and CP190 in a stable cell line expressing exogenous HA-tagged Mtr4, under control (non-treated), DNase I treated lysate and MNase treated lysate. Input material is lysate before IP, 2/3 the amount is deposited compared to IP samples. **C.** Volcano plots of the de-regulated genes detected by strand-specific RNAseq in Mtr4 KD, Beaf32 KD and dCTCF KD. Coloured are the genes with an absolute log₂ fold change (FC) threshold of 0.7, corresponding to a FC of 1.62, and an adjusted p-value of 0.05.

(End of Figure 1 Legends) **D.** Clustered heatmaps of the FC of deregulated genes detected in all 3 conditions, coloured by the intensity of their FC. **E.** Example of a PROMPT and an eRNA in strand specific RNAseq, forward strand on top, reverse strand on bottom. Box highlights the accumulated ncRNAs. STARR-seq plotted is from Zabidi et al., 2015. **F.** and **G.** show volcano plots of the detected eRNAs and PROMPTs, which are mainly accumulated upon Mtr4 KD. Adjacent profiles show Promoters and enhancers accumulating ncRNAs in Mtr4 KD. All conditions are plotted with only non-exonic reads RPGC normalized.

in at least one of the conditions, it is clear that Mtr4 KD has much larger effects than either IBP depletions. Despite the small number of genes regulated by Beaf32 and dCTCF, we observed a small enrichment for genes regulated in the same way in all three conditions (Figure S1D). However, the small sample of genes deregulated in Beaf32 and dCTCF KD shades uncertainty over a conclusion regarding co-regulation by Mtr4 and the two IBPs. Analysing the Gene Ontology (GO) enrichment for the genes deregulated in either condition did not provide any GO relative to Chromosome Organisation, gene expression, or RNA metabolism (Figure S1E).

Mtr4 being known in Human and Yeast for its role in the degradation of ncRNAs, we used ribodepletion instead of poly(A) during library preparation. The combination of ribodepletion with strand-specific RNAseq enabled us to detect ncRNAs *de novo* using the Ictea R package (see Methods section), as exemplified in Fig1E. We identified with high confidence a total of 222 PROMPTs and 225 eRNAs, respectively on 166 different promoters and 145 different enhancers that are accumulated upon Mtr4 depletion (Figure 1F and G & S1G). This contradicts previous reports (Lim *et al.*, 2013) that eRNAs and PROMPTs might not be accumulated in Drosophila upon exosome depletion, confirming also their transcription as demonstrated by (Rennie *et al.*, 2018). Furthermore, we thus confirm that Mtr4's role observed in Human is indeed conserved in Drosophila. We also show that Beaf32 and dCTCF do not have any noticeable effect on eRNAs and PROMPTs accumulation (Figure 1F, G & S1F). Finally, we note that upon Mtr4 KD 225 annotated, and probably more stable ncRNAs are also accumulated (Figure S1H). They may correspond to selected and spliced ncRNAs, such as stable eRNAs and other regulatory ncRNAs, as it has been previously proposed in multiple cases, reviewed in Syed and Hon, 2021.

Mtr4 depletion alters the binding of distinct insulator complexes to chromatin

Having observed that Mtr4 interact with insulator complexes, we studied its impact in the binding and recruitment of insulating factors to chromatin. Due to a lack of commercially available antibodies for several insulator factors in Drosophila, and for feasibility concerns, we decided to focus our ChIP experiments on Beaf32, for which Mtr4 interaction has been observed, as well as CP190 and

Cohesin, which are currently seen in the field as summarizing most IBP binding sites between the two of them. Producing ChIPseq duplicates for each, we observed reproducible changes upon Mtr4 KD. Notably, Beaf32 and CP190 recruitment to chromatin is increased or unchanged, while Cohesin recruitment on the other hand is mostly reproducibly reduced or unchanged (*Figure S2A, B & C*).

Drosophila has been demonstrated to form distinguishable types of insulator complexes through combinations of its IBPs. However, an up-to-date and exhaustive classification of the different insulator complexes currently lacks. In order to characterise which types of insulator complexes were impacted by Mtr4, we performed a k-means clustering using first a single replicate of ChIP seq for each condition (α Beaf32, α CP190 and α Cohesin in Luc and Mtr4 KD and their differential coverages), as well as all other available ChIPseq for insulator proteins in S2 cells, namely dCTCF, M1BP, GAF, Pita, ZIPIC, Su(Hw), Ibf1&2, Mod(Mdg4), Chromator (Kc167 cells) and Putzig (Kc167 cells) (Ong *et al.*, 2013; Cuartero *et al.*, 2014; Li *et al.*, 2015; Maksimenko *et al.*, 2015; Zouaz *et al.*, 2017)(*Figure 2A*). We empirically obtained a best fit with a clustering of 8 k-centroid seeds, thus obtaining 8 types of insulator complexes that can be well-separated (named A to H), when taking all the peaks of Beaf32, Cohesin and CP190 combined. This result is in accord with the 8 clusters obtained by Ramirez *et al.*, 2018. However, our set of clusters do not overlap well with those obtained using only motifs found in HiC-detected insulating frontiers in Ramirez *et al.* 2018. This is probably due to the base of the approach, our study comprising many more loci that do not necessarily exhibit strong insulating capacities, and encompassing more insulating factors. However, we must note that clusters E and F, on top of being very large (respectively $n=4630$ and $n= 3253$) exhibit relatively lower binding by known *Drosophila* insulator factors (*Figure S3*), and have lower percentages of motif enrichment for IBPs compared to other clusters (Supplementary Table 1). This may underline that yet more insulating proteins are left to be described. The clusters obtained are of good confidence, as a separate clustering using all available ChIP replicates identify highly similar clusters, with only marginal changes observed in the number of peaks in each cluster (*Figure S2D*).

One of the clusters (E ($n=4630$)) gathers most of the reproducible Beaf32 increases (*Figure S2E*), despite showing relatively low Beaf32 binding, as for all other IBPs and cofactors. Only small and un-reproducible changes in CP190 and Cohesin are also observed in this cluster (*Figure S2F & G*). Two other clusters show reproducible reduction in Cohesin recruitment, without significant nor reproducible changes in CP190 and Beaf32 (*Figure S2E & F*). These changes are found at Cohesin strongest sites, in clusters A ($n= 510$) and B, ($n= 473$), marked mostly by strong recruitment of CP190, CTCF and Pita, medium M1BP, Su(Hw), Mod(mdg4), and low Beaf32, Chromator and Putzig (*Figure Sup3*). These two clusters are notably dissimilar in their Ibf1 & Ibf2 binding, with Cluster A marked by the strongest binding of these two factors, while cluster B shows low Ibf1/2 binding, as well as low GAF

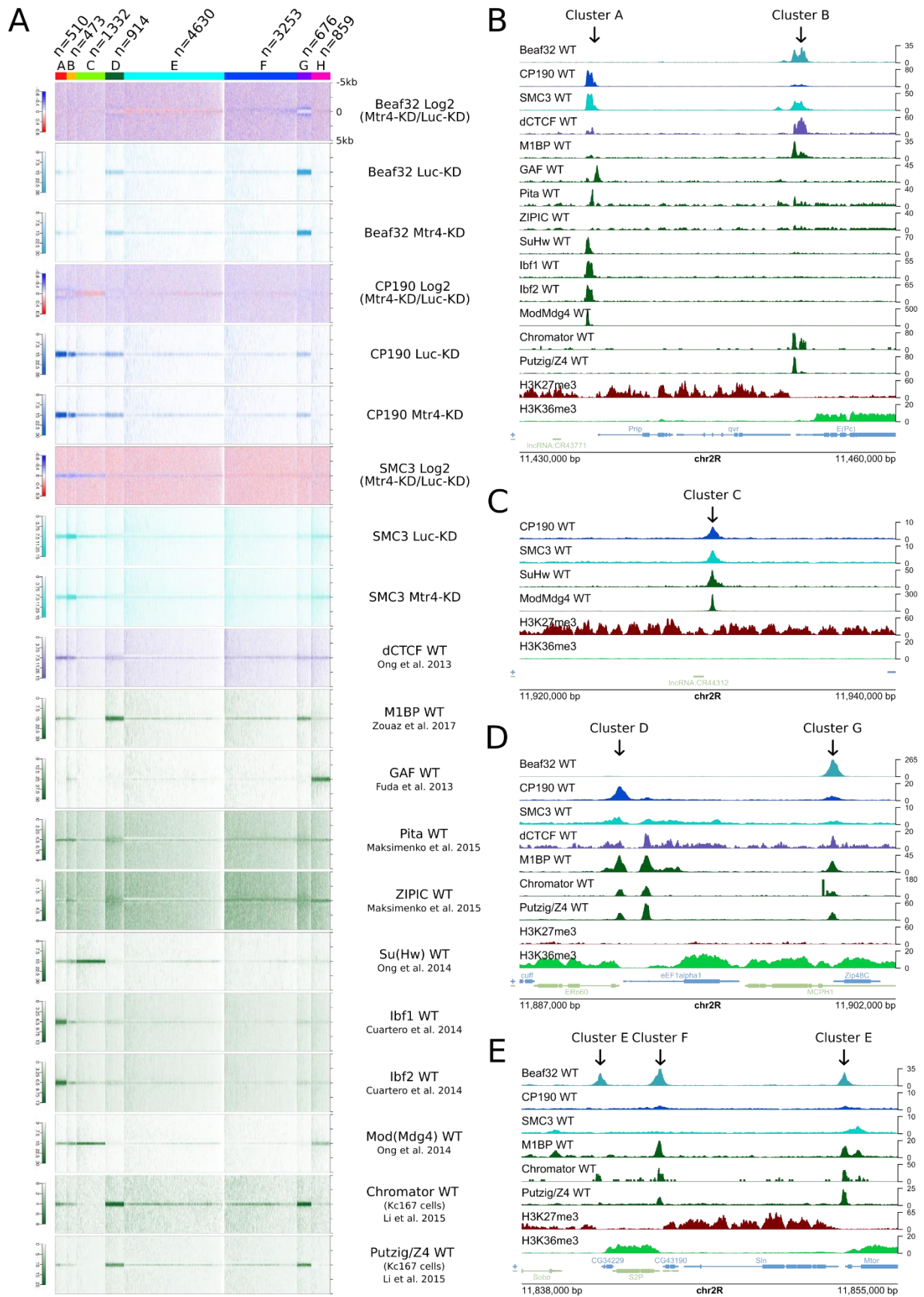


Figure 2: Drosophila insulators form distinct complexes, that are differentially influenced by Mtr4 KD. A. Heatmaps obtained by K-means with 8 k-centroid seeds, using all plotted ChIPseq as input, considering only

(End of Figure 2 Legends) the 500bp around the combined peaks detected for Beaf32, CP190 and SMC3. All tracks are normalised by RPGC. Beaf32, CP190 and SMC3 ChIPseq are scaled using factors computed by csaw from their peaks, for better inter-condition and inter-replicates comparisons. Number of peaks per cluster is indicated. B-E. Browser shots examples of peaks for each cluster, as indicated, showing only relevant tracks, as well as H3K27me3 and H3K36me3 ChIPseq from Depierre et al. 2023 as indicator of transcriptional status of the surrounding region.

and ZIPIC which are not found in cluster A. In the same vein, Cohesin and CP190 seem to be somewhat inversely proportional between those two clusters, with the highest CP190 in A, and highest Cohesin in B, while they are less recruited in the other cluster (*Figure S3*). Finally, Cluster C (n=1332) surprisingly exhibits mostly increased CP190 (*Figure S2F & H*), combined with lower Cohesin upon Mtr4 KD (*Figure S2G*). Aside of the medium recruitment of these two cofactors, this cluster is marked by high Su(Hw) and Mod(mdg4), and low dCTCF and M1BP, while other factors are mostly excluded. All 4 other clusters show no significant changes in Beaf32, CP190 or Cohesin recruitment, namely (1) cluster F (n=3253) marked by disparate and low binding of dCTCF, M1BP, ZIPIC, Chromator and Putzig; (2) Cluster D & G (respectively n= 914 and n=676) marked by high Beaf32, M1BP, Chromator and Putzig binding, and low recruitment of various other factors. Finally, cluster H forms a distinctive cluster marked mainly by high GAF binding, and medium SMC3, Mod(Mdg4), Pita, ZIPIC and dCTCF. While it is not impossible for these four clusters to be impacted by Mtr4 KD accumulation, it seems unlikely as they all are bound by either Beaf32, CP190, or Cohesin and no changes in their recruitment are observed. However, definitive evidence to elucidate whether these types of insulator complexes are unaffected by Mtr4 remains ChIPseq of every single IBP & cofactor.

Altogether, we can summarize that *Drosophila* insulating factors can form 8 types of complexes, which exhibit different changes upon Mtr4 KD. It is yet unclear from this clustering what makes an insulator complex sensible to Mtr4 KD. Excluding Cluster H, Mod(Mdg4) seem to be the only factor relatively enriched in the clusters showing effects from Mtr4 KD, and absent from other clusters. Beaf32 binding sites that are affected by Mtr4 KD have relatively low levels of binding, which are increased upon depletion, yet it is still unclear what makes these sites sensitive. The changes in CP190 and Cohesin recruitment are expected to be mainly indirect, and that it is primarily IBPs binding or their capacity to recruit cofactors that is changed. Yet with such a hypothesis, it is puzzling for cluster C to observe inversed effects between CP190 and Cohesin. Nonetheless, combinatorial effects may need to be considered, as we observe disparate and inconsistent changes between clusters. Substantially, further characterization of the genomic context and possible roles of each cluster is needed to achieve better understanding of the changes observed.

Types of insulator complexes have distinct chromatin context, and associate with ncRNAs and gene deregulation upon Mtr4 KD

Insulator complexes have been shown to operate in different context to regulate gene expression and 3D genome structure, acting both as constraints to Long-Range Interactions (LRIs), and as mediator of some LRIs. Also, they act as potent 1D regulators of histones-PTMs deposition regulators, and Beaf32 in particular has been shown to bind at a wide range of promoters (Jiang *et al.*, 2009; Liang *et al.*, 2014; Vogelmann *et al.*, 2014). To decipher the reasons and impacts of the changes in Beaf32, CP190 and Cohesin recruitment upon Mtr4 KD, we decided to characterise further the regions in which each type of insulator complex is found, keeping still to the 1D organisation of the genome.

First, we decided to probe whether certain clusters are particularly enriched in the two basic regulatory elements: promoters and enhancers. To that end, we made a simple overlap between each cluster's peaks and gene TSSs or enhancer, using an extension to 5kb around each peak, and comparing the proportions obtained with a group of random peaks of a size similar to that of the biggest cluster (E, 4630 peaks) (*Figure 3A*, left). With such parameters, we could readily observe that the peaks in clusters D and G overlap quasi-entirely with promoters, much higher than the ~40% expected from random. To facilitate and strengthen such comparisons, we performed a Fisher test to assess the relative enrichment of each element in each cluster depending on its size. For this analysis, we binned the entire genome in 5kb bins, attributing to each bin the presence of one feature (eg. Peak of cluster A), the other feature (eg. Enhancer from (Zabidi *et al.*, 2015), non-overlapping with a promoter), both or neither feature (*Figure 3A*, right). This statistical analysis clearly exposed that all clusters are enriched in promoters, although cluster D and G are clearly the most enriched. Also, cluster C, marked notably by Su(Hw), exhibits only a low enrichment in promoters (*Figure 3A*, right). As for enhancer, they seem to be enriched rather in Cluster H, marked by GAF, in particular, even though clusters B, E and F are also lowly enriched. On the other hand, clusters A, C, D and G are significantly depleted of enhancers (*Figure 3A*, right).

Having identified that the clusters we obtained are highly enriched in gene promoters, and sometimes enhancers, we wondered whether such an enrichment could also be observed when considering ncRNAs. Notably, we questioned whether the observed changes in CHIP of IBPs could be correlated with ncRNAs accumulation. Thus, we considered only promoters that accumulate PROMPTs, and enhancers that accumulate eRNAs, despite the low number of each feature (166 PROMPTs and 145 eRNAs accumulated) compared to the size of the clusters (500-4300 peaks each).

Following the same lines as Promoters in general, PROMPTs are also globally enriched in all clusters except cluster C (*Figure 3B*). Similarly, only the clusters enriched in enhancers are also enriched in eRNAs (i.e. clusters B, E, F & H), with the notable exception of cluster G which was depleted in enhancers, and yet is enriched in eRNAs. This apparent discrepancy between the clusters showing IBP changes from Mtr4 depletion (i.e. clusters A, B, C & E) and those accumulating PROMPTs and eRNAs argues against a local effect PROMPTs and eRNAs accumulation on the regulation of IBP binding or recruitment (*Figure 3B*).

Likewise, we questioned whether the local changes in IBP recruitment could be correlated with deregulation of genes in Mtr4 KD. In other words, we wondered if the TSS of genes deregulated upon Mtr4 KD tend to be found near the peaks of certain clusters. Thus, we separated up-regulated genes from down-regulated genes, and performed an overlap with the same parameters as for the previous overlap with all promoters. When plotting the fractions of de-regulated genes that were proximal to peaks of each cluster, we observed a clear correlation between cluster's size and the number of genes found overlapping (*Figure 3C, left*), thus we resorted again to relative enrichment using Odds Ratio from Fisher tests (*Figure 3C, right*). Surprisingly, we observed that up and down-regulated genes are similarly enriched in all clusters compared to random control sites, even though for cluster C only up-regulated genes show low yet significant enrichment. Cluster A and D however seem to be significantly more enriched in down-regulated genes rather than up-regulated genes. Altogether, these results tend to show that although insulator complexes are globally enriched near gene TSS and deregulated genes (*Figure 3A*), the deregulation of genes upon Mtr4 KD is not particularly correlated with the proximity to types of insulator complex that are affected in ChIPseq.

Finally, insulators complexes have been shown to form frontiers between domains of different transcriptional status (Depierre *et al.*, 2023). Thus, we studied the repartition of each cluster in the domains of heterochromatin (inactive) and euchromatin (active), and the frontiers between them. We used H3K27me3 domains from previously published domain detections (Depierre *et al.*, 2023) to call inactive domains, and considered their frontiers as the 5kb surrounding their edges. The remaining regions were attributed to active domains. In random regions, ~50% are expected to be in facultative heterochromatin domains, while ~30% are in euchromatin domains, and thus ~20% are found proximal to the frontiers detected in this manner (*Figure 3D*). For all clusters except cluster C, we can observe that larger fractions of peaks are found at frontiers (30-42%) (*Figure 3D*), in accordance with these clusters exerting roles of Histone PTMs-spreading barriers. On the other hand, clusters D, G and H, and to a lesser extent clusters E and F are found chiefly inside active domains (45-66%) (*Figure 3D*), concurring with their enrichment at promoters and enhancers, and their proximity to de-regulated genes (*Figure 3A&C*). Oppositely, cluster C is mostly present inside inactive domains (67%) (*Figure 3D*).

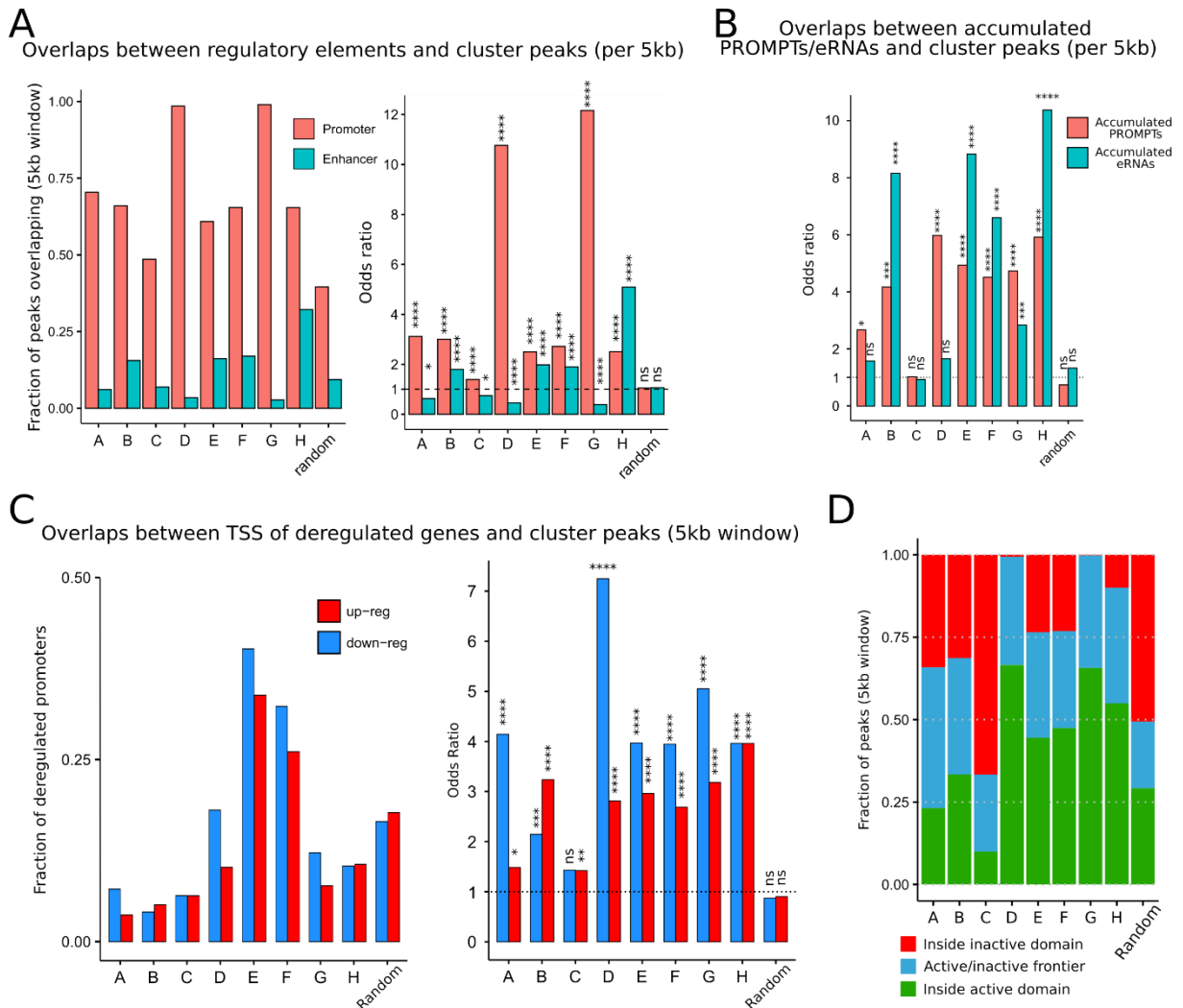


Figure 3: Insulator clusters have different genomic contexts. **A.** Left Barplot of the fractions of clusters' peak (+/-2.5kb around peak summit) that overlap with Promoter (+/-2.5kb around TSS) and Enhancers (STARRseq call (Zabidi et al. 2015) that do not overlap with promoters); Right Odds ratio from Fisher Exact tests, separately for each cluster and each features, computed from on the entire genome binned to 5kb bins, each bin is attributed with the presence or absence of corresponding feature or peak (i.e. bin with no peak nor promoter (-/-), bin with promoter and no peak (+/-), bin with both promoter and peak (+/+)). **B.** Same as A right panel, but promoters and enhancers are replaced by promoters with PROMPTs accumulated, and enhancers are replaced with enhancers with eRNAs accumulated. **C.** Left Barplot of the fractions of up- and down-regulated genes (+/-2.5kb around peak summit) that overlap with cluster peaks; Right Odds ratio from Fisher Exact tests, separately for each cluster and up- and down-regulated genes, computed from on the entire genome binned to 5kb bins, each bin is attributed with the presence or absence of corresponding gene or peak, as in A right panel. **D.** Cumulative barplots of the fractions of peaks from each cluster (+/-2.5kb around peak summit) in heterochromatin, euchromatin and at their frontier. Heterochromatin/inactive domains correspond to H3K27me3 domains from Depierre et al. 2023 and exclude non-mapable regions. Frontiers are called from the edge of H3K27me3 domains, considering them as the surrounding 5kb. The remaining fraction is considered as active domains.

This observation may explain its depletion for promoters, enhancers and their transcripts (Figure 3A&B), as well as the relative absence of enrichment for de-regulated genes for this cluster (Figure 3C).

Overall, all 8 clusters have distinct repartitions regarding regulatory elements, as well as heterochromatin, euchromatin and their frontiers. Shortly, Cluster A and B, seem to be found predominantly at frontiers, while Cluster C, specific to Su(Hw), is mainly present in heterochromatin. Conversely, Cluster D and G, marked by M1BP and Beaf32, are specific to active promoters, and Cluster H marks predominantly active enhancers bound by GAF. The two larger Clusters E and F are less distinct, although they do represent majorly active regions and frontiers, in continuity with their enrichment for promoters and enhancers, and their relatively low recruitment of IBPs and cofactors. Importantly, the clusters in which changes were observed for Cohesin, CP190 and Beaf32 recruitment upon Mtr4 KD (namely clusters A, B, C and E) do not show particular local enrichment for eRNAs and PROMPTs (*Figure 3C*), nor for deregulated genes (*Figure 3C*), when compared to other cluster where ChIP differences were not observed. This indicates that changes in insulator proteins recruitment is not correlated with local accumulation of ncRNAs, nor with local gene deregulation.

Mtr4 and ncRNAs accumulations influence the formation of insulator and transcriptional condensates

Recently, RNAs have been proposed as regulatory components of phase-separated condensates, such as transcriptional condensates. It has also been previously observed that insulators tend to form similar condensates (Amankwaa, Schoborg and Labrador, 2022). Namely, eRNAs have been proposed to be key elements in the formation of transcriptional condensates (Fukaya, Lim and Levine, 2016; Henninger *et al.*, 2021; Sharp *et al.*, 2022; Hamamoto *et al.*, 2023). However, most studies to date have observed the formations of such condensates in non-physiological *in vitro* conditions (Henninger *et al.*, 2021), or upon harsh *in vivo* treatments (complete arrest of transcription, complete RNA degradation (Henninger *et al.*, 2021)). Also, when more physiological conditions were achieved (Hamamoto *et al.*, 2023), the transcription of RNAs and their accumulation was not uncoupled, leaving ambiguous space to interpretations of the effects of RNAs in the formation of such condensates.

Presently, having observed the accumulation of numerous ncRNAs upon Mtr4 KD through their lack of degradation, we decided to challenge these conclusions. Precisely, we questioned whether the accumulation of those transcripts, and not their transcription, could influence the formation of transcriptional and insulator condensates. To that end, we depleted Mtr4 and detected transcriptional condensates by Pol II immunofluorescence, and insulator condensates by Beaf32 immunofluorescence

(Figure 4A), using high-resolution confocal microscopy and a deep-learning based segmentation of foci. The quantification in over 500 cell per condition (3 biological replicates) shows that ncRNAs accumulation causes only a small increase in the number of Pol II condensates per cell ($p=0.066$), and does not increase their median area (Figure 4B). Opposite effects were observed for insulator condensates, with no changes in the number of condensates per cell, but a small increase in their size ($p=0.041$) (Figure 4C). Collectively, our results support that ncRNAs accumulation may have effects in the regulation of transcriptional condensates number per cells, and insulator condensates size, yet nuance the extent of such regulations to relatively slight global changes.

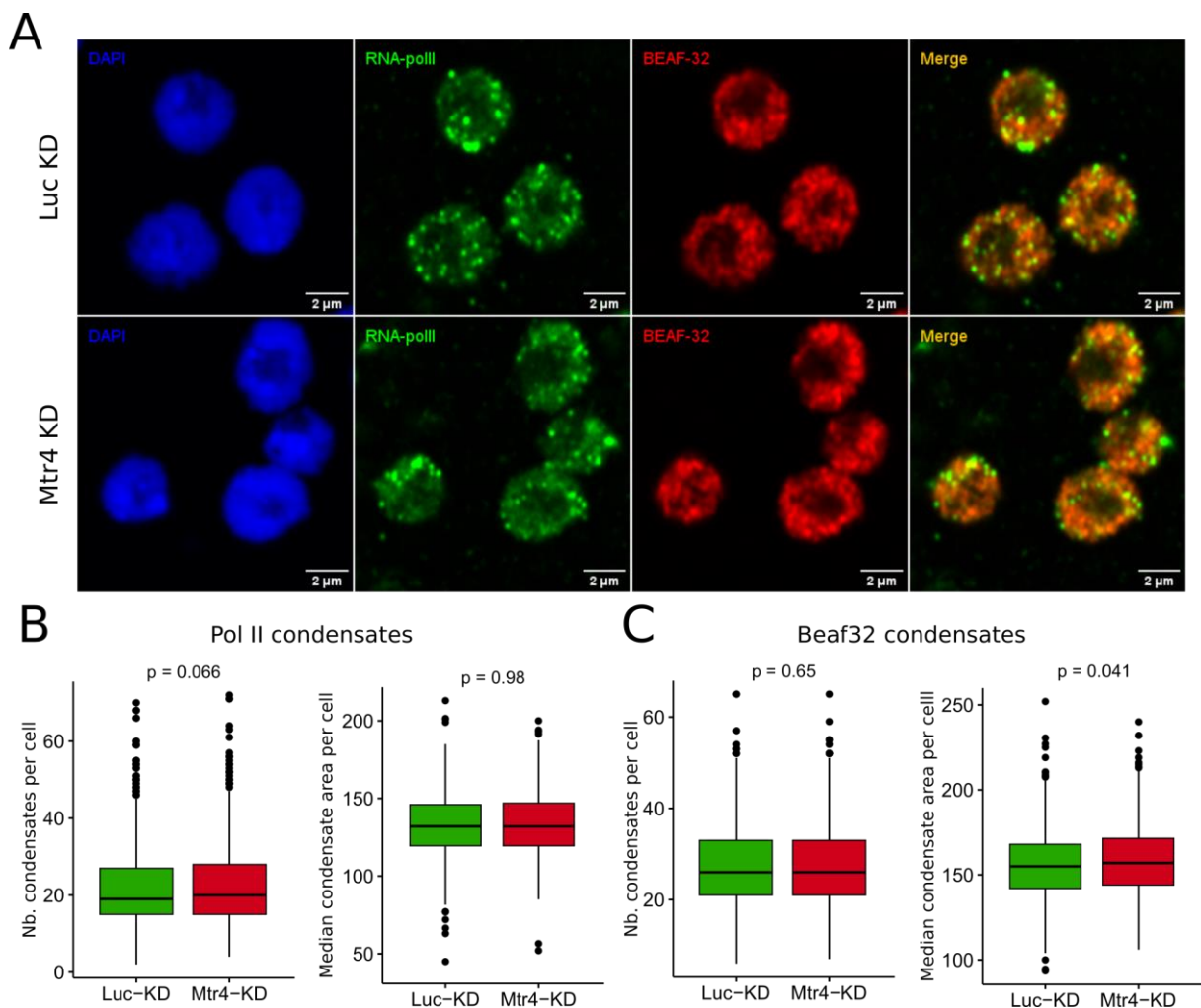


Figure 4: Transcriptional and insulator condensates show little effect of Mtr4 depletion and ncRNAs accumulation. **A.** Representative examples of Confocal Airyscan acquisitions of Beaf32 and Pol II immunofluorescence in both Control (top) and Mtr4 KD (bottom). **B.** Boxplots showing the number (left) and median area (right) of Pol II condensates per cell. p -values are measured by Wilcoxon test (biological replicates $n=3$; 597 cells for Luc KD, 484 cells for Mtr4 KD). **C.** Same as in B but for Beaf32 condensates (biological replicates $n=3$; 583 cells for Luc KD, 471 cells for Mtr4 KD).

Mtr4 does not affect insulation by insulator complexes

The main role of insulators is to act as constrainers of contacts, so that interactions are mainly kept inside TADs, and are limited between contiguous TADs. These TAD structures and their frontiers are readily seen from HiC data. To study the impact of Mtr4 on 3D genome organisation as a potential 3D regulator, we performed HiC in duplicates in control conditions, as well as in Mtr4, Beaf32 and dCTCF depletions. Having verified replicates reproducibility, we merged the reads of both replicates to obtain better coverage, and down-sampled all conditions to the condition lowest-depth, with ~67M reads, to obtain best inter-sample comparisons. First, we observed that no global disruption of the polymer behaviour of chromatin was observed when comparing HiC contact frequencies by genomic distances (*Figure 5A*). Hence, only changes specific to certain regions can be expected.

Keeping to the characterisation of our ChIPseq-derived clusters of insulator complexes, we used the Insulation Score as a measure of each clusters' capacity to insulate. The Insulation Score measures particularly the capacity for a given region (here of 10kb) to block contacts between the neighbouring regions. The lowest Insulation Scores correspond to the strongest insulation. Plotting the insulation score +/-100kb around groups of loci enables the visualisation of the "dip", which is proportional to the capacity of a locus to block contacts between the regions upstream and downstream. Plotting the peaks from each cluster separately clearly demonstrates that they are distinct in their capacity to insulate contacts (*Figure 5B*). Importantly, Su(Hw)-enriched Cluster C behaves as random and is incapable of insulation. On the other extreme, Cluster D and G (marked by M1BP and Beaf32) show the strongest levels of insulation, around -0.3 (*Figure 5B*). In between, cluster A, B and H show similar insulation with around -0.2, while cluster E and F are also comparable with an average 0.15 insulation score (*Figure 5B*).

Further comparisons of Insulation Scores in each cluster between depletion conditions for Mtr4, Beaf32 and CTCF with the control condition yields surprising insights in the different clusters' reactions to each Knock-down. First of all, Mtr4 does not seem to affect insulation in any cluster, except a marginal effect on cluster A (*Figure 5C & S4*). On the other hand, Cluster H is the most affected, with a relative and incomplete loss of insulation in both dCTCF and Beaf32 KD (*Figure 5C & S4*). Cluster B is comparably affected by dCTCF KD, but not by Beaf32 KD (*Figure 5C & S4*). Surprisingly, cluster D and G, where the most Beaf32 is found in ChIPseq show small increases in insulation upon both Beaf32 and dCTCF KD compared to control and Mtr4 KD. This would tend to show that at these loci, instead of causing insulation as expected, Beaf32 and CTCF would not impact, or even obstruct insulation. Finally, Cluster E and F show comparable changes, with particularly a small increase in insulation in cluster E for Beaf32 KD.

Despite the surprising results obtained for certain clusters upon Beaf32 and dCTCF depletion that seem to increase insulation, we can conclude that the clusters found using ChIP data are consistent in HiC with different insulating capacities. It is to be considered that compensations may be at play between M1BP and Beaf32 for example, in clusters D and G, in which M1BP would take over upon Beaf32 depletion as they bind very similar sites, accounting for the relative lack of effect of Beaf32 KD. Generally, however, Mtr4 KD causes no meaningful changes to insulation, despite the changes observed in Beaf32, CP190 and Cohesin. Therefore, Mtr4 and ncRNAs accumulation has apparently no impact on the insulating function of insulator complexes.

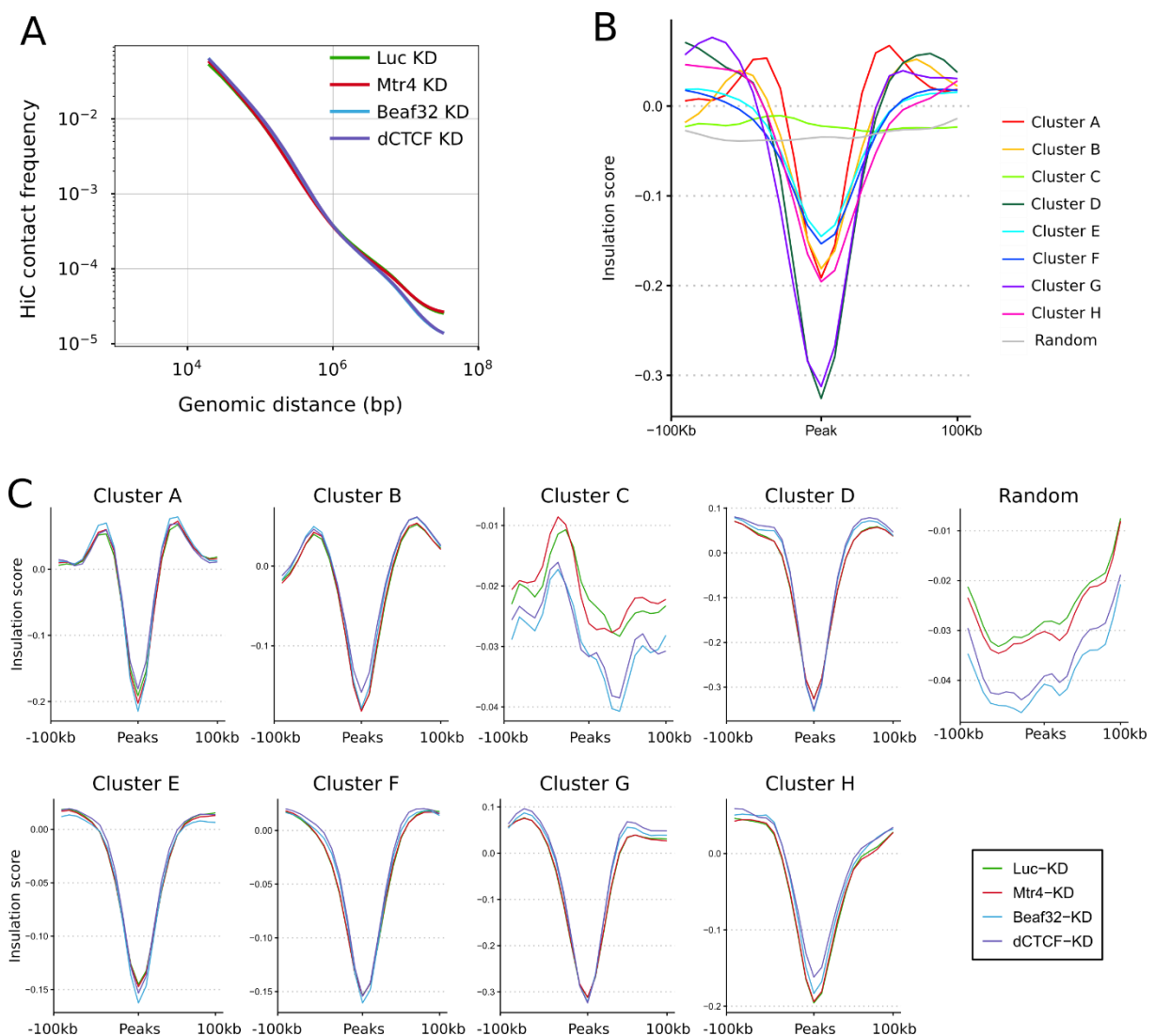


Figure 5: Mtr4 does not influence insulation by insulator complexes. A. Profile of the contact frequency as a function of genomic distance in all conditions. All HiC conditions are down-sampled before plotting. A smoothing from the coolpuppy package was applied. **B.** Profile of the insulation score around the peaks of each cluster +/-100kb, in control condition at a resolution of 10kb. **C.** Same as in B, but plotting each cluster separately, with all conditions.

Mtr4 influences long-range interactions

Associated with their capacity to form insulating frontiers, insulators have also been demonstrated to mediate specific LRIs. Also, in *Drosophila* in particular, gene promoters have been observed as recurrently interacting with other promoters. However, as insulator proteins are often found bound near promoters, it is yet unclear how much of promoter-promoter interactions are due to transcription dynamics, and how much is imputable to insulator-mediated contacts. To simplify the analysis of the effects of Mtr4 depletion on the formation of these LRIs, we focused on the observation of Mtr4 KD compared to control in each cluster.

Questioning once again the relevance of the ChIPseq-determined clusters in this context, we first plotted in Aggregated Peak Analysis (APA) the homotypic contacts, i.e. contacts between sites pertaining to the same clusters, ranging from 20kb to 500kb to probe the type of contacts they formed. Doing so, we exposed once again that these clusters also have characteristic behaviours in 3D, notably forming different LRIs (*Figure 6A*), which are not formed by random loci (*Figure 6B*). Strikingly, the heterochromatic cluster C forms low-intensity pinpoint loops with itself, without showing particularly enriched contacts in the bottom left quadrant, which would correspond to inside-TAD contacts. This result can be expected from the previously described formation of LRIs by Polycomb to spread and maintain inactive domains through development. Of interest, a small decrease in contacts is observed upon Mtr4 KD, suggesting that ncRNA accumulation might obstruct Polycomb loops. Secondly, the promoter-enriched Cluster D and G, marked by M1BP and Beaf32, show intense loops with themselves (*Figure 6A & C second panel*). They also exhibit clear accumulation of reads in the bottom left quadrant compared to the other quadrants, thus Beaf32 and M1BP promoter-frontiers expose both enriched contacts between themselves, and good insulation capacities, reminiscent of Human TAD-corner loops. Oppositely to cluster C however, loop-contacts are increased upon Mtr4 KD in both cluster D and G, without noticeable change in inside-TAD contacts. GAF-enriched cluster H show a similar pattern, with loops of lesser intensity but only with itself (*Figure 6A & C, both bottom panels*), which are increased in Mtr4 KD. Cluster A and B on the other hand expose no enriched loop, but do have strongly enriched intra-TAD contacts, supporting their role as TAD frontiers, with only little change in contacts in cluster A in Mtr4 KD (*Figure 6A & C top panel*). Finally, the larger cluster E and F show only little centre-loop and have enriched intra-TAD contacts which expose little to no change upon Mtr4 KD (*Figure 6A*). Thus, the different types of insulator complexes apparently form distinguishable 3D genome structure, and are differentially affected by Mtr4.

In order to examine the capacity of those clusters to interact with TSSs, we first examined the impact of gene expression levels on TSS-TSS interactions. Interestingly, on top of genes generally acting

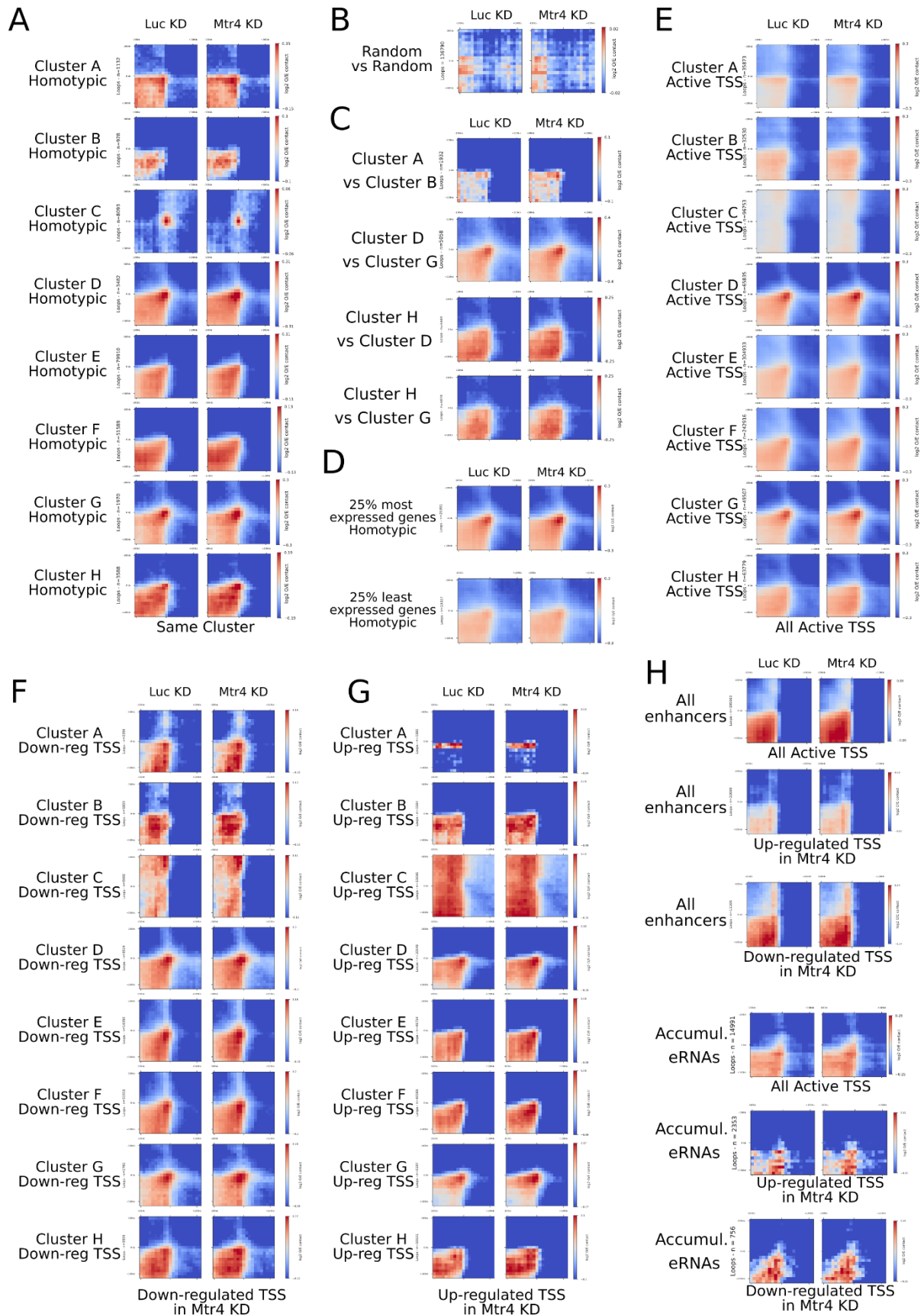


Figure 6: Mtr4 affects the strength of certain LRIs mediated by insulator complexes. A-H. Aggregated Peak Analysis (APA) derived from the coolpuppy package, at distances of min 20kb and max 500kb between anchor and bait. All plots show +/- 100kb around both anchor and bait. The plotted signal is a log₂ of observed over

(End of Figure 6 Legends) expected values in each condition. The random group is the same as in all previous figures (4630 random loci). **D, E, H.** "All Active TSSs" corresponds to genes that have at least 100 reads in RNAseq ($n=8875$). **H.** "All enhancers" are STARRseq called from Zabidi et al. 2015 that do not overlap with promoters; while "Accumul. eRNAs" are all the enhancers that have an accumulated eRNA in Mtr4 KD ($n=145$).

as contact barriers, we observed that strongly expressed genes interact significantly more than lowly-expressed genes (Figure 6D). This indicates that transcription levels might impact the formation of 3D LRIs. However, it is difficult to uncouple whether it is the formation of these loops that facilitate gene expression, or the opposite. Importantly, we can observe that Mtr4 depletion has only marginal effects on TSS-TSS interactions in general, with only a slight increase in contacts upon Mtr4 KD for the 25% most expressed genes (Figure 6D). When plotting all clusters separately against all active TSSs in the 20-500kb range, we observe similar patterns as in homotypic interactions, with the notable exception of cluster C that is under-enriched in contacts with active TSSs, and does not show any frontier activity (Figure 6E). Also, cluster E and F seem to have relatively stronger loops compared to their homotypic interactions, while cluster H does not exhibit such a loop with active TSSs, contrarily to its self-interactions (Figure 6E). Finally, the loop-forming cluster D and G seem to be the only ones to be noticeably affected by Mtr4 KD, with an increase in loop-contacts (Figure 6E). Altogether, despite distinguishable behaviours for each cluster in their interactions with active TSSs, the depletion of Mtr4 and ensuing accumulation of ncRNAs had only small effects on clusters' interactions with TSSs. However, strongly expressed genes and cluster D and G seem to interact somewhat more with active TSSs upon Mtr4 depletion.

The relatively small changes observed upon Mtr4 KD when considering interactions with all active TSSs led us to reason that separating down-regulated genes and up-regulated genes from non-deregulated ones may uncouple different behaviours in LRI changes. In fact, little to no changes are observed for clusters' interactions with down-regulated TSSs (Figure 6F). Conversely, for up-regulated TSSs, we can observe clear increases in loop-interactions with cluster D, E, F, G and H upon Mtr4 KD (Figure 6G). This correlation between gene upregulation and increased contacts may underline that Mtr4 degrades ncRNAs' accumulation to control the formation of regulatory LRIs and thus regulate gene expression. Hence, we tested whether the regulation of loop-contacts by Mtr4 KD and ncRNAs acted primarily on the regulation of long-range Enhancer-Promoter interactions, or on facilitating the formation of transcription-inducing phases. Plotting all Enhancer-Active TSSs' contacts and all enhancers versus upregulated and downregulated, we observed no apparent change upon Mtr4 depletion. Likewise, when considering only eRNA-accumulating enhancers, few modifications of contacts were observed in the absence of Mtr4. Therefore, while developing validating statistical tests, we hypothesize that Mtr4 chiefly serves to control ncRNAs' levels in order to regulate their enhancement of transcriptional phases, rather than on regulating long-range Enhancer-Promoter interactions.

Locus-specific artificial accumulation ncRNAs by dCas9 is not generally sufficient to influence Beaf32 recruitment and gene expression

As previously reported (Shechner *et al.*, 2015; Sigova *et al.*, 2015; Yang *et al.*, 2021; Hou and Kraus, 2022), the targeting of specific ncRNAs by the extension of the guide-scaffold sequence of a catalytic-dead Cas9 (dCas9) can be sufficient to modify the binding of architectural proteins to chromatin (Sigova *et al.*, 2015). As Mtr4 depletion and the subsequent accumulation of ncRNAs cause changes in Beaf32 recruitment in certain loci, we tried to set up a more direct measurement of the effect of ncRNAs accumulation on Beaf32. In other words, we tried to recapitulate the effects of Mtr4 KD on Beaf32 binding and gene expression through targeted accumulation of ncRNAs. Hence, we chose several loci in which eRNA accumulation was adjacent to Beaf32 binding sites that exposed differential binding upon Mtr4 KD. Then, we selected the ~100bp most accumulated in the eRNA, and fused it to a guide-scaffold RNA for dCas9 targeting adjacent to the Beaf32 peaks of interest, by transitory transfection and selection of transfected cells. As a control, we decided to use the exact same constructs but without insertions of an eRNA sequence.

To test different loci as proofs of concept, we chose 3 loci : (1) the upSET gene (hereafter referred to as upSET locus), that comprise an internal enhancer in its first intron that accumulates eRNAs upon Mtr4 KD, and has Beaf32 binding sites on its promoter and both internal enhancers, both of which have increased Beaf32 binding upon Mtr4 KD (*Figure S5A*); (2) a locus with 3 adjacent enhancers (hereafter called 3-Enhancer locus), two of which are active, bound by Beaf32, and show eRNA accumulation in Mtr4 KD (*Figure S7B*). Also, this entire locus is in long range interaction with Nolo 1,5Mb away, a gene de-regulated upon Mtr4 KD that is marked by several Beaf32 binding sites (*Figure S5C*); (3) a locus comprising a solitary enhancer and unexpressed gene (Lone Enhancer locus), marked by a small peak of Beaf32 accumulated upon Mtr4KD and a bi-directional accumulation of eRNAs (*Figure S5D*).

First, we confirmed good targeting of the dCas9 by these constructs using ChIP of the GFP fused to the dCas9, achieving enrichment ranging from 20 to 200 times the background (*Figure 7A*). Then, we tried to measure the effects of such a targeting of eRNAs on the surrounding peaks of Beaf32. In all cases, both locally and for the Nolo locus, although changes were observed, they were not reproducible in the two biological replicates produced (*Figure 7A*). This is probably due at least in part to the intrinsic variability of ChIP-qPCR, and would require several more biological replicates to enable the adequate measurement of changes as fine as those expected from Mtr4 KD ChIPseq data of Beaf32 (*Figure 7B*). Similarly, the measurements of gene expression by RT-qPCR for upSET and the surrounding

genes, as well as for Nolo, show ample variability, rendering impossible conclusions regarding the effects of this targeting of ncRNAs on gene regulation (Figure 7C).

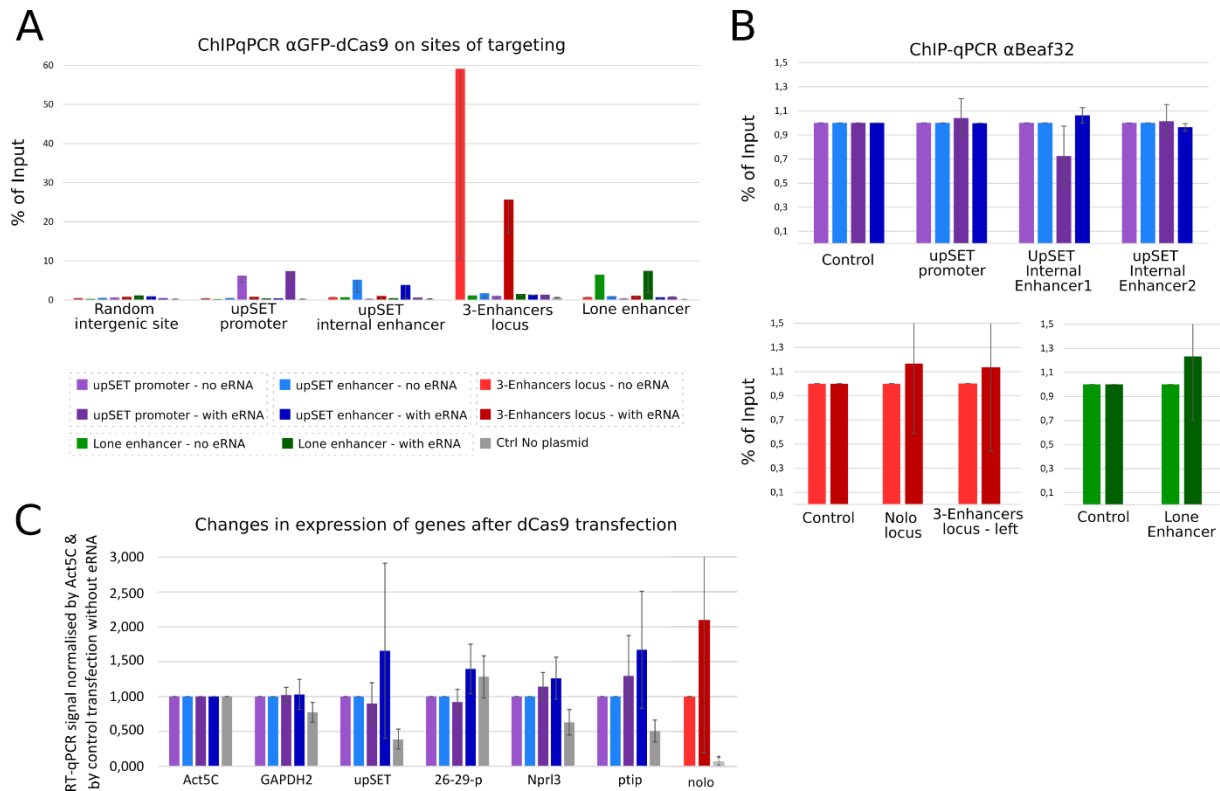


Figure 7: CRISPR-dCas9 targeting of ncRNAs is functional but is not sufficient to influence Beaf32 recruitment and gene expression. A. Barplot of ChIP-qPCR measurement of dCas9 targeting to the loci of interest. The random intergenic site used as control is a non-genic region (biological replicates, $n=2$). **B.** ChIP-qPCR barplot of Beaf32 on loci of interest, normalised by input and control sites. Control sites correspond to Beaf32 binding sites not targeted in the condition plotted. A normalisation is also made between “no eRNA” and “with eRNA” samples to ease comparisons (biological replicates, $n=2$). **C.** RTqPCR barplots of gene expression measurements, normalised by Act5C and between “no eRNA” and “with eRNA” samples to ease comparisons (biological replicates, $n=2$).

Chapter 2 – Rm62 & Abs impact on gene regulation, insulators & genome spatial organisation

Rm62 and Abs interact with insulators and regulate gene expression, but not ncRNAs

In the same Mass-spectrometry dataset that led to the identification of Mtr4 as a potential cofactor of Beaf32 insulators, we also identified Abs, another less studied RNA helicase, even though with less confidence due to low numbers of peptides. Rm62 on the other hand was not present in this dataset, but had been identified in older papers as a potential RNA helicase cofactor to CP190 in *Drosophila* (Lei and Corces, 2006) and CTCF in Human (Yao *et al.*, 2010), potentially in association with ncRNAs. In CoIP experiments, we reproducibly observed immunoprecipitation of Rm62 with Beaf32 (*Figure 1A*). Abs CoIP however were rather limited: although a band was indeed observed in both ways of CoIP, it was reproducibly tenuous in the case of Abs precipitation with Beaf32 antibodies (*Figure 1A*). Together, the modest percentage of both RNA helicase immunoprecipitation with Beaf32 may suggest that the interaction is either indirect, or that it may concern only a fraction of each factor.

After validation of both depletions by RTqPCR (*Figure S1A*), we probed the effect of each RNA helicase's depletion on gene expression regulation, using strand-specific RNAseq. We observed that, similarly to Mtr4 depletion, we had mainly upregulations (71,3% of the deregulated genes for Rm62 KD, and 67,2% for Abs KD) (*Figure 1B&C*). Also, Rm62 KD had larger effects than Abs KD, with a total of 1289 deregulated genes for the former, and 461 for the latter when using a threshold at an absolute Log2FC of 0,7 (~1,62 FC) (*Figure 1B*). In like manner to Mtr4 depletion, a tendency for deregulation of the same genes in the same manner was observed between Rm62 and Abs, as well as Beaf32 and dCTCF, yet the small numbers of genes with such propensities is too small to suggest a proper coregulation of genes. Comparisons of the enriched Gene Ontologies for up- and down-regulated genes in each condition did not yield any GO of relevance to the study (*Figure S1B*). This indicates that in S2 cells, these two factors may regulate a rather broad spectrum of genes, with a relative independence from the biological processes they are involved in.

Finally, a study of ncRNAs deregulated upon Rm62 and Abs depletions revealed only very few cases of accumulation or downregulation of both PROMPTs and eRNAs, with only tens of accumulation in particular for Rm62 KD (*Figure 1D*), and 5 to 10 deregulated non-coding transcripts for Abs KD (*Figure 1C*). Similar results were obtained when considering only annotated ncRNAs. This would tend to show

that Abs is not involved in regulation of ncRNAs, but Rm62 may be involved in very selective cases in the destabilisation or maturation of specific ncRNAs.

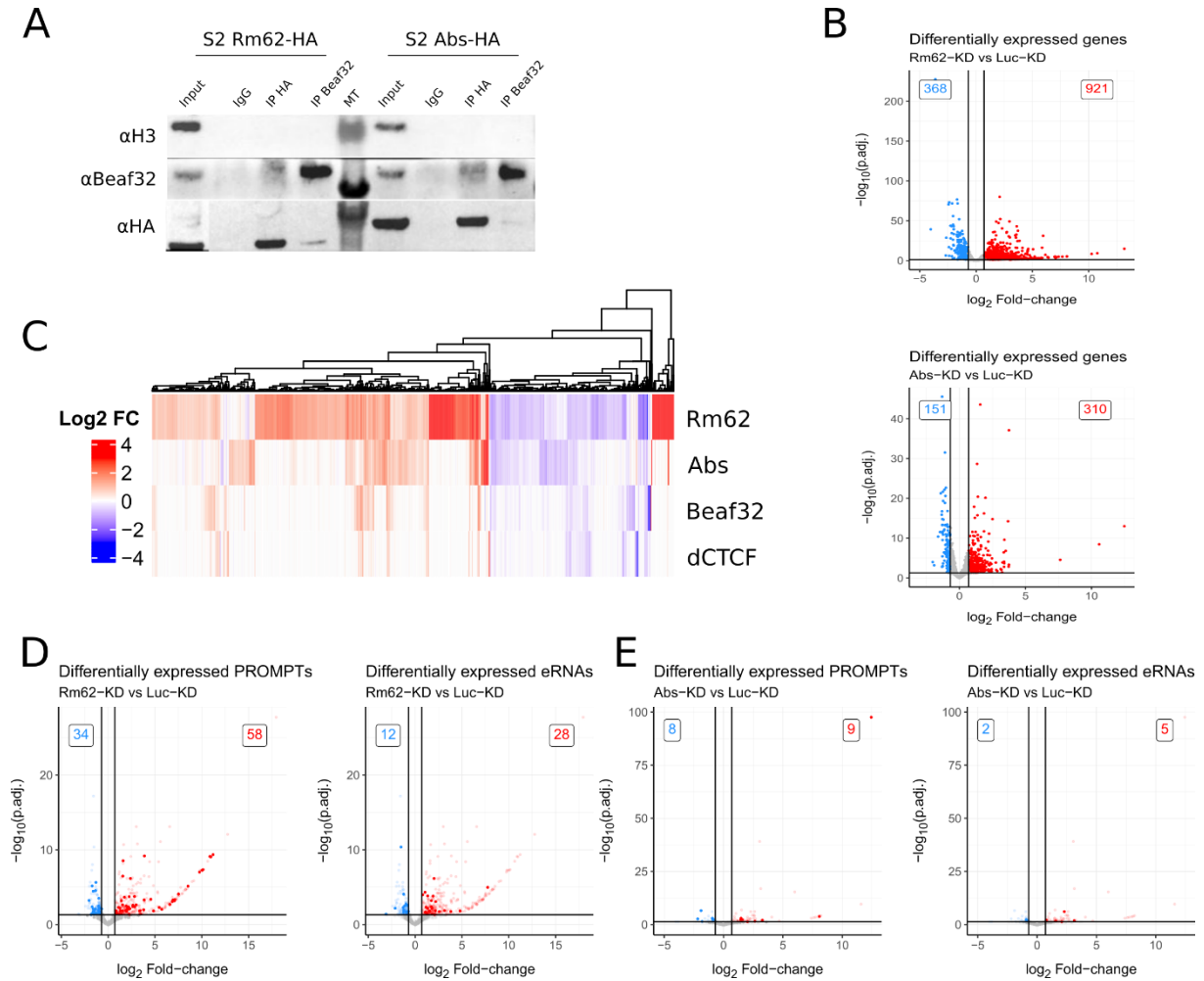


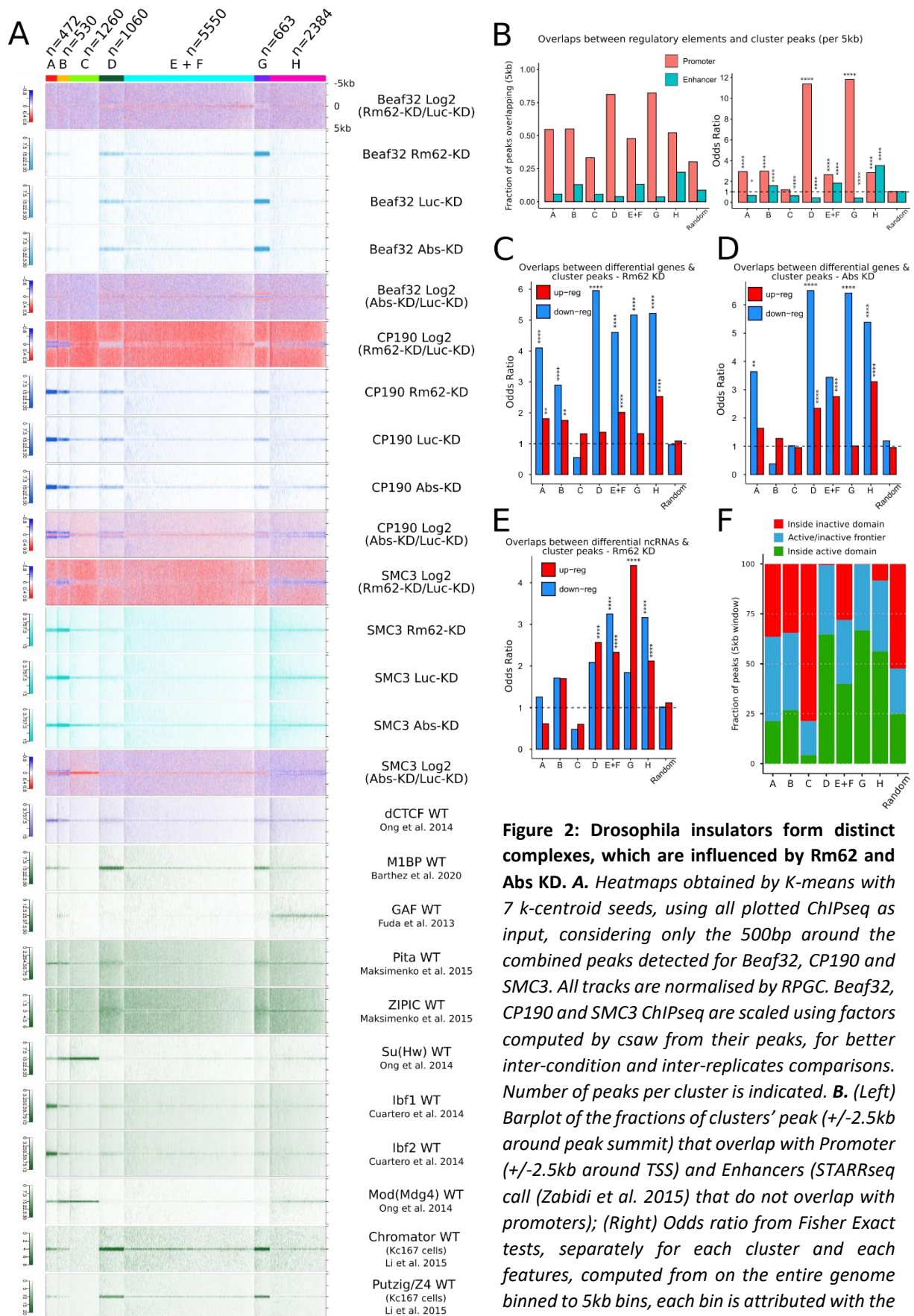
Figure 1: Rm62 and Abs interact with Beaf32 and regulate gene expression, and Rm62 regulated few ncRNAs. A. Western Blot of CoIP showing interaction Rm62-HA and Abs-HA with Beaf32 in stable S2 cell line expressing exogenous HA-tagged Rm62 and Abs in C-terminal. Input material is lysate before IP, 2/3 the amount is deposited compared to IP samples. Irrelevant Rabbit IgG serves as control. **B.** Volcano plots of the de-regulated genes detected by strand-specific RNAseq in Beaf32 KD, dCTCF KD, Rm62 KD and Abs KD. Coloured are the genes with an absolute log₂ fold change (FC) threshold of 0.7, corresponding to a FC of 1.62, and an adjusted p-value of 0.05. **C.** Clustered heatmaps of the FC of deregulated genes detected in all 4 conditions, coloured by the intensity of their FC. **D & E.** Volcano plots of the eRNAs and PROMPTs detected as differential upon Rm62 KD (**D**) and Abd KD (**E**).

Rm62 & Abs depletion cause changes in distinct insulator complexes

Applying the same method as for Mtr4 KD, we performed a k-means clustering to differentiate the insulator complexes, considering the changes observed upon Rm62 and Abs depletions. In these conditions, we observed best reproducibility using 7 k-seeds centroids, instead of 8. Primarily, these changes are due to a merge of cluster E and F (hereafter named cluster E+F, n= 5550) (*Figure 2A*). Also, a large part of these two clusters is transferred to the previously observed cluster H, marked by high GAF binding. This latter modification brings cluster H to a total of 2384 peaks (*Figure 2A*). We understand these modifications to be imputable to the lack of truly distinctive feature in the previous Clusters E & F observed in the Mtr4 chapter. As numerous sites in both clusters exposed GAF binding, albeit relatively low, it is also understandable then that they become more similar to cluster H and are clustered with it. Otherwise, cluster A, B, C, D and G are all correspond very well to their name-wise counterpart identified in the Mtr4 chapter, with only tens of peaks switching between clusters (*Figure 2A*). Shortly, cluster A & B are marked by high CP190, Cohesin, SMC3, dCTCF and Ibf1 & Ibf2 binding. Cluster is mainly marked by the highest levels of Su(Hw) and Mod(Mdg4). Finally, clusters D & G are chiefly bound by the strongest Beaf32 and M1BP peaks.

Upon Rm62 KD, we observe reproducible increases in CP190 recruitment in cluster E+F (see around centre bin (*Figure 2A*) and (*Figure S2, Middle*)). Cohesin on the other hand is mainly less recruited to cluster A, B and H (*Figure 2A & S3C, Right*). In the meantime, cluster E+F and G show slight yet reproducible increases in Cohesin binding (*Figure 2A & S3C, Right*). Strikingly, Abs KD showed markedly different changes, with a notable increase in Cohesin recruitment in all cluster except cluster H (*Figure 2A & S3C, Right*). However, cluster C shows a noticeably stronger increase, compared to other clusters, underlying a particular effect of Abs KD on this cluster (*Figure 2A*). Also, a reproducibly lower recruitment of CP190 is observed in cluster H upon Abs KD. As for clusters C and E+F, a large range of variations is observed in CP190 binding, causing a lack of significant changes in either direction, despite a generally higher binding in replica 1 in particular (*Figure 2A*). Collectively, we can summarise that both Rm62 and Abs impact the binding of CP190 and Cohesin, but not Beaf32. However, the two RNA helicases affect separate types of insulator complexes and in different manners.

Characterisation of the 1D chromatin context of the clusters identified upon Rm62 & Abs KD leads to nearly identical conclusions to the clusters observed in Chapter 1. Mainly, all cluster but in particular clusters D & G are enriched in Promoters (*Figure 2B*), while Enhancers are enriched in clusters H, E+F and to a lower extent cluster B, and are depleted in all other clusters (*Figure 2B*). Correlations of gene deregulations in Rm62 KD (*Figure 2C*) and Abs KD (*Figure 2D*) with the proximity of each cluster peaks yields little information, as all clusters except cluster C are enriched in de-



(End of Figure 2 Legends) promoter and peak (+/+). **C.** Odds ratio from Fisher Exact tests, separately for each cluster and up- and down-regulated genes in Rm62 KD, computed from on the entire genome binned to 5kb bins, each bin is attributed with the presence or absence of corresponding gene or peak, as in B right panel. **D.** Same as C, but with Abs-KD de-regulated genes. **E.** Odds ratio from Fisher Exact tests, separately for each cluster and up- and down-regulated ncRNAs upon Rm62 KD, computed from on the entire genome binned to 5kb bins, each bin is attributed with the presence or absence of corresponding gene or peak, as in C & D. **E.** Cumulative barplots of the fractions of peaks from each cluster (+/-2.5kb around peak summit) in heterochromatin, euchromatin and at their frontier. Heterochromatin/inactive domains correspond to H3K27me3 domains from Depierre et al. 2023 and exclude non-mapable regions. Frontiers are called from the edge of H3K27me3 domains, considering them as the surrounding 5kb. The remaining fraction is considered as active domains.

regulated genes. Thus, no clear link can be made between the changes observed in ChIPseq binding of CP190 & Cohesin and the deregulation of proximal genes in one way or another. Similarly, an enrichment for deregulations of ncRNAs accumulation upon Rm62 KD can be observed proximal to cluster D, E+F, G and H, despite the small number of de-regulated ncRNAs. However, these deregulations do not overlap well with the differential recruitments observed for Cohesin and CP190 in condition of Rm62 KD. Finally, we observe a distinguishable pattern of enrichment in heterochromatin, euchromatin and proximity to frontiers for each cluster, similar to the one observed for the clustering obtained with Mtr4 KD ChIPseq conditions. Shortly, clusters C peaks are primarily found in heterochromatin, while cluster D, E+F, G & H peaks are enriched in euchromatin and frontiers. Again, cluster A and B mark notably active-inactive frontiers.

Altogether, we note that all seven clusters are dissimilar in their recruitment of insulating binding proteins and cofactors recruitment, and mark different types of regulatory elements and chromatin transcriptional states. However, in spite of distinct changes in CP190 and Cohesin recruitment upon Rm62 and Abs KD between these clusters, they are not differentially associated with the local regulation of gene expression by Rm62 and Abs.

Rm62 & Abs may participate in regulating 3D genome organisation

Given the changes induced in the recruitment of CP190 and Cohesin by the depletion of Rm62 and Abs, we performed HiC to probe how these two RNA helicases may impact genome spatial organization. First, performing a down-sampling between all conditions and plotting the interactions frequency as a function of genomic distance demonstrated that no general effects are caused by either depletion (Figure 3A). Therefore, if these two RNA helicases impact genome 3D organisation, we can only expect changes at specific loci.

Similar to the clustering-derived types of insulator complexes of Chapter 1, the clusters obtained with Rm62 and Abs datasets had distinguishable proficiency in prohibiting contacts between upstream and downstream regions (*Figure 3B*). Cluster D and G, enriched in active promoters, Beaf32 and M1BP have the best insulation capacities, with Insulation Scores (IS) down to -0,3. Clusters A, B and H also expose good insulation, between -0,2 and -0,15. Cluster E+F on the other hand has relatively low insulation, down to -0,1. Finally, the mainly heterochromatic cluster C does not insulate contacts, and even seem to have a small tendency to favour contacts between surrounding regions. Surprisingly, comparisons of the insulation score for each cluster in conditions of Rm62 KD and Abs KD revealed that both depletions impact insulation. Whereas Abs depletion causes a global increase in insulation in all clusters but cluster C compared to random, only cluster A, B, E+F and H showed such an increase upon Rm62 depletion (*Figure 3C*). Importantly however, we were able to obtain statistical significance for these differences only in Cluster D and H in Abs KD. These results suggest that Rm62 and Abs somewhat impede the formation of insulating frontiers by insulator complexes. Concurring with the absence of changes of Beaf32 binding, the Beaf32-enriched clusters D and G are not impacted by Rm62 KD. On the other hand, the striking global impact of Abs KD in all insulating clusters may be related to the general increase in Cohesin recruitment observed previously (*Figure 2A*), but does not coincide well with the lack of changes in ChIPseq observed for cluster H, and with the relatively low correlation between Cohesin recruitment and insulation capacity. Hence it is still unclear from the changes observed in ChIPseq for Beaf32, CP190 and Cohesin what exactly causes the observed increases in insulation by the different insulator complexes.

Considering the formation of LRIs by the clusters, we first observed that the patterns observed for cluster homotypic interactions (i.e. between sites of the same cluster) were identical to those previously described in Chapter 1 (*Figure 3D*). Interestingly, these interactions were not impacted by the depletion of Rm62, with the exception of the pinpoint-loop form by cluster C which is lessened (*Figure 3D*). Surprisingly, same result is obtained for cluster C upon Abs KD (*Figure 3D*), in spite of the notably increased recruitment of Cohesin on those same loci (*Figure 2A*), underlining that Cohesin is not the mediator of these pinpoint loops. For all the other clusters, Abs KD causes a remarkable increase in loop contacts particularly (see centre of the heatmaps (*Figure 3D*)), but does not seem to impact the bottom-left corner contacts (i.e. contacts between the anchor and bait, often corresponding to TAD-like contacts). Similar effects are observed for both Rm62 KD and Abs KD when observing heterotypic interactions (*Figure S3B*) and contacts between the most expressed and least expressed promoters (*Figure S3C*). Accordingly, plotting interactions between clusters and de-regulated promoters (*Figure 3E*) in Rm62 KD exposed no changes of interactions, whereas in Abs

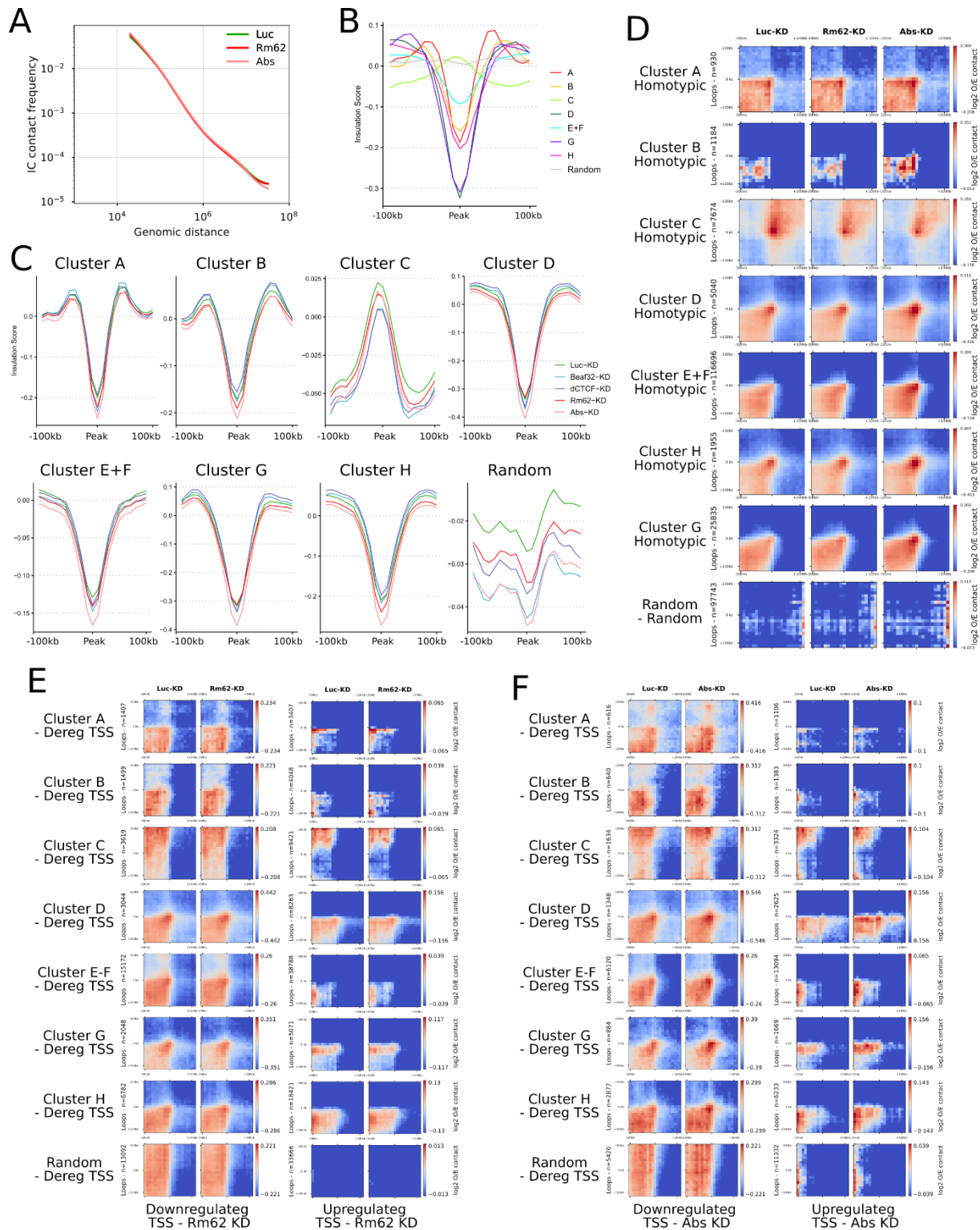


Figure 3: Modifications in insulation and LRIs observed by depletion of Rm62 and Abs. **A.** Profile of the contact frequency as a function of genomic distance in all conditions. All HiC conditions are down-sampled before plotting. A smoothing from the coolpuppy package was applied. **B.** Profile of the insulation score around the peaks of each cluster ± 100 kb, in control condition at a resolution of 10kb. **C.** Same as in B, but plotting each cluster separately and in all conditions to observed changes. **D, E & F.** Aggregated Plot Analysis (APA) derived from the coolpuppy package, at distances of min 20kb and max 500kb between anchor and bait. All plots show ± 100 kb around both anchor and bait. The plotted signal is a \log_2 of observed over expected values in each condition. The random group is the same as in all previous figures (5550 random loci);

(End of Figure 2 Legends) D. plots homotypic interactions (e.g. all potential interactions between all cluster A peaks); E & F. show interactions between clusters sites (anchor, mid-left) and de-regulated genes (bait, mid-bottom) in Rm62 KD (E) Abs KD (F).

depletion, clusters and deregulated genes tended to interact more, particularly in clusters D, E+F, G and H (*Figure 3F*). To confirm the modifications observed on these APAs, we are currently developing a statistical test that enables the proper validations of the changes.

Overall, although it is clear that both Rm62 and Abs impact gene expression, the recruitment of CP190 and Cohesin to chromatin and the formation insulating frontiers as well as the mediation LRIs, it is yet still uncertain how all three are related to each other. In fact, it is probable that other important factors in this mechanism have yet to be identified. Most of all, the way by which these two RNA helicases may cause all of these effects remains to be discovered, as they have only little to no effect on the accumulation of ncRNAs.

DISCUSSION

Limitations to the study of ncRNAs and their roles

In this study of insulators and their cofactors, we have described Mtr4 as a potential new cofactor, or regulator of insulator complexes. Notably, we have identified that a few hundreds of PROMPTs and eRNAs are regulated by this RNA helicase. Compared to the previously described thousands of potentially transcribed eRNAs and PROMPTs by Rennie *et al.*, 2018. Our view on this discrepancy is that Mtr4 may not be essential for the proper degradation of all eRNAs and PROMPTs. Indeed, as its enzymatic capacity is to unwind secondary structures, perhaps it is needed solely to facilitate the degradation of structured ncRNAs. However, Mtr4 functions as a complex, notably the Nuclear Exosome Targeting complex (NEXT)(Lubas *et al.*, 2015; Gerlach *et al.*, 2022), which also serves to target specific RNAs to the non-selective exosome for degradation. Nonetheless, multiple subcomplexes seem to be underexplored even in Human (Winczura *et al.*, 2018; Gerlach *et al.*, 2022), and remain to be studied in *Drosophila*, thus other exosome targeting complexes may exist to target non-structured ncRNAs for degradation. On the other hand, as Rennie *et al.* identified the transcription of PROMPTs and eRNAs using CAGEseq, it is plausible that the increased sensibility of CAGE for transcription initiation has identified rare transcripts, which would not be detected as enriched with our RNAseq coverage.

Also, our study was limited to unstable ncRNAs, and does not decipher the potential impacts of the 225 accumulated annotated ncRNAs. These ncRNAs may be stabilised and matured for specific purposes, and may also act not only locally, but be targeted to various loci across the genome, as it has been demonstrated in several cases (Oh *et al.*, 2021; Syed and Hon, 2021; Luo *et al.*, 2022). These more stable ncRNAs are currently seen in the field as selected eRNAs that have developed various roles through the course of evolution (Syed and Hon, 2021). However, the study of the effects of these potentially widely distributed ncRNAs is also arduous, as they may not bind strongly to chromatin. Also, the current techniques to study their localisation (such as Chromatin Isolation by RNA Precipitation (ChIRP), or R-loop mapping if they form R-loops (MapR, DRIPseq)) are error-prone, and are mostly semi-quantitative. In the same vein, the study of stable ncRNAs roles and or/structure is highly time-consuming, as it necessitates to focus on a single or few candidates. Nonetheless, the study of these more stable ncRNAs which are also regulated by Mtr4 hold great potential to develop our understanding of ncRNAs fine-tuning of genome regulation.

Strikingly however, linking stable or unstable ncRNAs and their fluctuation to chromatin changes and gene deregulation is challenging. Notably, they have the potential to act not only locally in 1 dimension along the chromatin fibre, but also to impact multiple loci in 3D through long-range contacts and functional condensates. Tracking and associating causes and effects thus becomes most complex, and is prone to be misled due to artefactual changes. Practically, at a genomic level, the small number of ncRNAs regulated by Mtr4 in particular also hinders many statistical analyses.

Another level of potential impacts of ncRNAs in epigenetic is the regulation of Transcriptional Pausing and of histone PTM deposition and maintenance, as it has been proposed by several studies (Syed and Hon, 2021; Gorbovytska *et al.*, 2022). These mechanisms remain mostly underexplored by this study and more generally in *Drosophila*, yet may also account for some of gene deregulation observed upon Mtr4 depletion.

The conundrum of *Drosophila* insulators

Regarding *Drosophila* insulators, we have identified several types of insulator complexes with distinctive features, both in terms of IBPs and cofactors bound but also of genomic context. However, these clusters do not overlap well with those obtained using only motifs found in insulating frontiers in Ramirez et al 2018. This is probably due to the base of the approach used, our study comprising many more loci that do not necessarily exhibit strong insulating capacities, and encompassing more insulating factors. However, we must note that clusters E and F, on top of being very large (respectively $n=4630$ and $n=3253$) do not exhibit strong binding by any known *Drosophila* insulator factor, perhaps exhibiting that yet more insulating-associated proteins are left to be described. Otherwise, it is possible that these loci only accumulate low-bindings of several IBPs, sufficiently to form functional insulator complexes.

Interestingly, we have observed that the mostly-heterochromatic Cluster C, enriched in Su(Hw), has no insulating capacity. Also, this cluster surprisingly shows opposite effects, with increases in CP190 and decreases in Cohesin recruitment upon Mtr4 depletion. As the LRI formed by this cluster resemble Polycomb loops, and are slightly decreased upon Mtr4 KD, we hypothesize that ncRNAs somewhat inhibit these LRI stabilisation by Cohesin. The puzzling simultaneous increase in CP190 might then result from an increased availability for Su(Hw) to recruit CP190. Although ncRNAs have previously been shown to participate in the regulation of H3K27me3 deposition by Polycomb

(Cifuentes-Rojas *et al.*, 2014), it remains to be explored whether the accumulation of ncRNAs by our depletion of Mtr4 is sufficient to cause changes in the dynamic of H3K27me3 maintenance.

Considering the other clusters where Cohesin recruitment modifications were observed upon Mtr4 KD, it is surprising that only decreases are observed. Actually, the literature to date has rather shown ncRNAs to promote Cohesin recruitment (Pan *et al.*, 2021). Also, in comparison to previously published dataset from Ramirez *et al.* (2018), we observe a strong recruitment of Cohesin to Su(Hw). This may be due to discrepancies in experimental method (ChIPseq of Rad21 compared to SMC3 ChIP in our case), but also to the previously discussed different method of identification of insulators, based on HiC in their case.

Of note, despite trying to produce ChIPseq for Mtr4, we were not able to produce a satisfying mapping of Mtr4 potential sites of recruitment, even though we observe reproducible co-immunoprecipitation of Mtr4 with Beaf32. This may be due to indirect and possibly feeble binding of Mtr4 to chromatin, through Beaf32. Alternatively, it is possible that our use of Formaldehyde crosslinks was either not potent enough, or that as it has been observed for other factors, Formaldehyde itself had disadvantageous effects on Mtr4 binding to chromatin.

More generally, our clustering of most IBPs and insulator cofactors identified to date shades new light on the variety of insulator complexes present in *Drosophila*. The various mechanisms, as well as the specific and redundant roles of each factor is still subject to much debate to date. Notably, our observation that depletion of insulator factors has the surprising effect of both decreasing insulation in some loci, and increasing insulation in other loci, has also been observed in other studies, such as for dCTCF by Kaushal *et al.*, 2022. It remains mysterious how DNA sequences and factors binding them are capable of insulation in reporter assays, yet their removal is not always sufficient to abolish insulation. Much arduous investigation remains to be done to understand the combinatorial effects of all of *Drosophila* insulator complexes factors. In fact, such studies have to date been focused on the main factors, such CP190, Beaf32 and dCTCF. Future studies will probably reveal further understanding of insulators by examining the separate and combinatorial roles of the other less-studied insulator binding proteins and their cofactors.

Evidently, associating insulator complexes with their role in gene regulation, on top of their influence on 3D genome organisation, is difficult due to the incompatibility between the two scales. Also, 3D genome organisation seems more important for the regulation of gene expression variability and throughout development, rather than for transcription factor-like up- or down-regulation of transcription (Rowley *et al.*, 2017). In the same vein, regulatory contacts between enhancers and promoters may occur in a dynamic timescale of a few seconds (Downen *et al.*, 2014), thus being virtually

impossible to capture using genome-wide techniques, despite having important effects on gene expression. Hence, it has been previously shown that most genes de-regulation upon deletion of insulating sequences and factors could not be directly linked to changes in domain-boundary changes in insulation (Cavalheiro *et al.*, 2023). Also, in complete agreement with our results, transcription itself has a major influence on insulation. In fact, it has been reported that transcriptional arrest at a boundary had a greater impact than the deletion of the region bound by IBPs itself (Cavalheiro *et al.*, 2023). In the context of the packed *Drosophila* genome, where a large part of insulators and promoters overlap, the exact importance of insulators in the formation of insulating frontiers remains relatively obscure. On that point, our study has shown that the role of transcription in insulation would rather be due to the act of transcription itself, as the accumulation of ncRNAs has demonstrated little effects on insulation, contrarily to previous propositions (Lei and Corces, 2006; Matzat, Dale and Lei, 2013; Saldaña-Meyer *et al.*, 2019). However, we cannot exclude that certain specific ncRNAs, particularly stable ncRNAs, may have such capacities.

Multiple interplays between ncRNAs and genome organisation

Furthermore, our identification of changes in the formation of LRIs upon Mtr4 KD raises several questions. Chiefly, as Mtr4 does not seem to impact insulation, exactly how Mtr4 may influence LRIs remains to be investigated further. One of our hypotheses was that some IBPs and insulator cofactors may bind ncRNAs regulated by Mtr4. However, our attempts at iCLIP for Beaf32 remain inconclusive to date, and confirm neither Beaf32 capacity or incapacity to bind RNA. Alternatively, other and not all proteins pertaining to insulator complexes may have the specific capacity to bind RNAs. Nonetheless, to study further how Mtr4 KD may impact LRIs, it would be interesting to separate potential contacts between those occurring within TADs, compared to those between different TADs. Despite the still unreliable calling of TADs in *Drosophila* using HiC due to their small size, such an analysis may yield a better view of exactly which contacts are impacted by Mtr4 KD and ncRNAs accumulation.

Our results show that interactions between active TSS and between certain types of insulators and upregulated genes in particular seem to be increased upon Mtr4 KD. We believe that such an increase in contacts may be linked to our observation of increased number of transcriptional condensates and increased size of Beaf32 condensates upon Mtr4 depletion. In fact, as it has been proposed that eRNAs in particular participate in the formation of transcriptional condensates (Henninger *et al.*, 2021; Hamamoto *et al.*, 2023), we suggest that the ncRNAs accumulated due to Mtr4 depletion somewhat facilitate the condensates initiation. This would conceptually mainly explain the

increases in gene expression, which concerns about three quarter of the deregulated genes upon Mtr4 KD. Regarding down-regulated genes, which also exhibit less intense deregulation, we may expect that the disequilibrium caused by facilitated transcriptional phases and burst is at the expense of genes that typically readily initiate transcriptional bursts, independently of ncRNAs. In fact, if one supposes as the literature suggests that a threshold of concentration of factors, including RNAs, is necessary for transcriptional condensate formation and burst initiation, then the accumulation of ncRNAs due to their lack of degradation would lower that threshold. Thus, the facilitated initiation of transcriptional bursts would rather favour genes that are usually at the lower end of the limit and for which initiation of a transcriptional burst is challenging, making them more reliant on ncRNAs. On the other hand, in this global increase in transcriptional bursting, transcriptional factors may become limiting. In this case, genes that do not depend on ncRNAs to form transcriptional condensates and initiate a burst of transcription would be generally disadvantaged, hence reducing their level of expression. However, our data does not enable the measurement of transcriptional burst induction as it was done by Hamamoto et al., 2023. Hence, we must infer from their observation of eRNAs transcription impact on transcriptional burst that our observed accumulation of ncRNAs is of the same effect, at least to some extent.

Regarding our design of dCAS9-targeted eRNAs to influence insulators and transcription, although yet inconclusive, we are still hopeful that different loci and measurement methods may yield better exploitability. Importantly, the designs for these 3 sites were produced before our bioinformatic analysis led us to consider the different clusters of insulator complexes exposed above. Thus, having a proof of concept that the targeting of eRNAs by dCas9 is functional, it will be of more interest to focus the targeting on the most impacted loci for each cluster. Also, this proof of concept focused on ChIP of Beaf32, and future experiments should also probe for changes in CP190 and Cohesin recruitment changes, as they are observed for clusters A, B and C. Finally, to solidify the reliability of the measurements of IBPs recruitments, it may be interesting to use sequencing instead of qPCR as an output. To enhance practicality and reduce cost, it is also possible to target multiple loci on different chromosomes or separated by sufficient distances, instead of performing time-consuming separate transfection and ChIP for each construct as it was done here.

Rm62 and Abs as potential insulator cofactors

Admittedly, our study of Rm62 and Abs as cofactors to insulator complexes is still somewhat unrefined. Chiefly, we do lack a model of how these two under-studied RNA helicases may exert the extensive effects we have observed, ranging from gene expression regulation to 3D genome organisation, and including modifications of Cohesin and CP190 recruitment to chromatin. In fact, although for Mtr4 we observed large accumulations of ncRNAs, themselves backed by the literature to have multiple roles in chromatin regulation, that is not true of Rm62 and Abs. For one thing, we did not observe accumulation of ncRNAs in depletion of Abs, and very few when depleting Rm62, thus we can exclude that ncRNAs levels are the mean through which Rm62 and Abs can regulate chromatin. However, they may yet regulate the structure of specific, stable ncRNAs, themselves potentially influential in the regulation of other proteins, as it has been described in several cases notably for CTCF (Yao *et al.*, 2010; Hansen *et al.*, 2019; Saldaña-Meyer *et al.*, 2019; Oh *et al.*, 2021; Harris *et al.*, 2023). Rm62 Human homolog DDX5 in particular is already thought to be a regulator of ncRNA conformation to facilitate the formation of specific large protein complexes (Giraud, Terrone and Bourgeois, 2018).

Alternatively, it is conceivable for both Rm62 and Abs that their capacity to resolve R-loops (Yu *et al.*, 2020; Mosler *et al.*, 2021) is conserved in *Drosophila*. In fact, R-loops have been proposed to be regulators of various factors' binding to chromatin, such as CTCF (Oh *et al.*, 2021; Luo *et al.*, 2022; Wulfridge *et al.*, 2023), and Cohesin (Pan *et al.*, 2021). Also, as previously described, R-Loops may directly enhance transcription (Tan-Wong, Dhir and Proudfoot, 2019), and cause increased DNA double strand breaks went unproperly regulated. As the study of R-loops as genomic regulator is still booming at the time of writing, it is also likely that they will be associated with other roles in the future, including but not confining to regulations of *Drosophila* insulator complexes and 3D genome organisation in general.

Regarding the previous studies of Rm62 and its human homolog, DDX5, it is probable that at least some of the gene deregulations observed in this study can be accounted for by its other demonstrated roles in the recruitment of chromatin modifiers and remodelers (Giraud, Terrone and Bourgeois, 2018). Our results however seem somewhat contradictory with the previously observed role for Rm62 in the insulating capacity of CTCF in human (Yao *et al.*, 2010). This may be expected, as indeed even Beaf32 and dCTCF depletion does not suffice in *Drosophila* to observe broad effects in dysregulations of insulating frontiers, probably due to the previously discussed redundancies and intricacies of the various *Drosophila* insulating factors. We do confirm nonetheless that Rm62 interacts with insulator complexes, as described by Lei and Corces in 2006. In the same line of thought, it remains mysterious from our results, as much as it was in those two studies, exactly how this RNA helicase can

mediate changes in insulator complexes and 3D genome organisation. Regardless, we did provide much information in the way of what types of changes are observed, and in what context, hopefully setting the ground for future identification of the mechanisms which are at play.

Collectively, our study of Mtr4, Rm62 and Abs as potential new insulator cofactors has been fruitful in the way of regulation of insulator-associated proteins recruitment to chromatin, as well as in the apparent regulation of several features of genome spatial organisation. Despite our relative lack of understanding of the underlying mechanisms, we have identified that these 3 RNA helicases have contextualized effects that depend on the type of insulator complexes that are considered. We propose that all 3 RNA helicases exert different effects and at different sites, possibly through separate mechanisms, thus we support that RNA helicases may be an important branch of insulator complexes regulators that account for some of the previously described loci-dependent behaviour of insulators. Along this course, the global accumulation of ncRNAs upon depletion of Mtr4 has also permitted us to study the effects of ncRNAs accumulation when uncoupled from their transcription, a current subject of intense debate in the field. Further, we have set up in our system a method to try and assess directly the effect of ncRNAs accumulation on insulator complexes and transcription at specific loci, hopefully laying the ground for local chromatin regulations by ncRNAs to be exemplified in *Drosophila*. Finally, we aspire for this study to add good wood to the fire that is currently the field of RNA roles in chromatin regulations.

METHODS

Wet-lab experiments

Cell culture, dsRNA depletions & transfection

Drosophila melanogaster Schneider 2 (S2) cells were cultured at 25°C in Schneider's medium (11590576, Fisher) supplemented with 10% Foetal Bovine Serum (FBS) (11550356, Life Technologies) and 1% Penicillin/Streptomycin (P-4333, Sigma). For double-stranded RNAs (dsRNA), PCR templates were amplified using the PrimeSTAR Max DNA Polymerase Premix 2X (TAKR045A, Takara). After PCR product purification using the Wizard SV Gel and PCR Clean-up kit (A9282, Promega), single-stranded RNAs (ssRNA) were transcribed *in vitro* with T7 promoters on both 5' ends using the TranscriptAid T7 High Yield Transcription Kit (K0441, ThermoFisher Scientific) according to the manufacturer's instructions. ssRNA were then purified with phenol-chloroform, precipitated, and paired. dsRNA against luciferase was used as a control. For each condition, cells were passed and after 24 hours were transferred to media with no FBS, added with 10µg dsRNA per million cells. After a 2 hours incubation at 25°C, media supplemented with FBS was added to a final concentration of 10% FBS and 1% antibiotic and 1 million cells per mL. Cells were harvested 3 days post-treatment. For plasmid transfection, 3M cells were seeded in 35mm plates, and left overnight (O/N) to grow. Cells were then transfected using a dropwise-mix of solution A (per plate: 18µL 2M CaCl₂, 9,5µg plasmid of interest, added with 1µg selection co-vector for stable transfection, complemented to 150µL with sterile H₂O) and solution B (2X HEPES-Buffered Saline; 50mM HEPES, 1.5 mM Na₂HPO₄, 280mM NaCl, pH 7.1), left for incubation at Room Temperature (RT) for 30 minutes. After 24 hours incubation, the cells were washed once with complete medium and diluted into 3mL medium per plate to incubate 2-5 days. For transitory transfection, cells were then harvested. To select for stable cell lines, cells were added with the proper selecting antibiotics at day 5, and then passed until stabilisation of the cells and proliferation (2-3 weeks), replacing selecting media every 4-5 days.

Cloning

To build HA-tagged RNA helicases, a Getaway protocol was followed. Shortly, genomic DNA was PCR amplified for each protein with AttB extension on the primers, using PrimeSTAR Max DNA polymerase (Takara Bio). Amplified product was then cloned into pDONR221 using a BP reaction. After transformation and purification of the obtained plasmids, they were subjected to LR reaction with pDEST plasmids with Actin promoters. For Mtr4, a pAHW (DGRC barcode 1095) was used, to tag the helicase in N-terminal, whereas pAWH (DGRC barcode 1096) were used for Rm62 & Abs to tag in C-

terminal. For CTCF, a pAWF was used (DGRC barcode 1112), for a Flag in C-terminal. All minipreps were performed using Wizard Plus SV Minipreps DNA purification systems (Promega), all midipreps were performed using QIAGEN Plasmid Plus Midi Kit. All plasmids were fully sequenced to confirm the absence of mutations.

FACS

2M S2 cells were centrifuged and washed into 10mL of cold 1X PBS, centrifuged again resuspended into 0.9mL cold 1X PBS. 2.1mL cold Ethanol absolute was added dropwise on low vortex, and incubated 1h at -20°C for fixation. After centrifugation, cells were washed with 10mL cold PBS 1X 0.5% Tween (PBST), centrifuged again, and resuspended into 0.8mL PBST. 400µL were kept as control, and the rest was centrifuged to be resuspended into 400µL of PBS 1X, RNaseA 0.2mg/mL and Propidium Iodide 50µg/mL, and incubated 1h at 37°C or O/N at 4°C. A minimum of 50 000 cells were counted per condition.

Reverse Transcription & qPCR

Total RNA extraction was performed using RNeasy Plus Mini kit (QIAGEN Cat#74134). Reverse transcription (RT) of 1µg of total RNA was done using the Maxima H Minus First Strand cDNA Synthesis kit, using both Oligo(dT) and random hexamer. For qPCR, samples were diluted to 1/20, or 1/10 for rare RNAs such as PROMPTs & eRNAs, and a 4-points scale was produced using a mix of all sample. qPCR acquisition was performed on a ViiA 7 Real-Time PCR System.

RNA-seq

Total RNA extraction was performed using TRIzol, and sent to BGI genomics for RT and library preparation, using rRNA depletion and strand-specific library preparation. Sequencing was aimed for ~20M 100bp PE reads per sample.

Western Blot & Co-Immunoprecipitation

Nuclear extracts or whole cell extracts were denatured and stained with the NuPAGE MOPS SDS Buffer Kit (NP0050, ThermoFisher), and run on Nu-PAGE NOVEX 4-12% Bis-Tris Gels (10247002, Fisher). Proteins were transferred from the gel to two nitrocellulose membranes by passive diffusion O/N at 55°C in Transfer Buffer (50mM NaCl, 1mM EDTA, 10 mM Tris-HCl pH 7.4, 0.1mM DTT). All washing of nitrocellulose membranes were performed with PBS + 0,15% Tween (PBST). After washing thrice, membranes were blocked for ~1 hour with PBST + 5% milk. After blocking, the membranes were washed twice and incubated for 2 hours at RT with the primary antibody diluted in PBST + 2% BSA. After three more washes, membranes were incubated 1 hour at RT with IgG HRP secondary antibody

likewise diluted. Membranes were imaged using ECL Primer Western Blotting System on a ChemiDoc Imaging System machine (BioRad).

For Co-immunoprecipitation (CoIP), 20M cells were centrifuged and washed twice with PBS 1X, and lysed into 400 μ L ES2 lysis buffer (20mM HEPES pH 7.5, 100mM KCl, 0.05% Triton X-100, 2.5mM EDTA, 5mM DTT, 5% Glycerol, 1X protease inhibitor cocktail (Roche)) for 2h at RT. Sepharose beads were washed twice with PBS 1X, and kept in a 1:1 (v/v) slurry into PBS1X. After centrifugation at max speed for 10 mins at 4°C of the lysates, 20 μ L was kept as input, and the rest was incubated with 50 μ L bead slurry for 30 mins at 4°C. Quick spinning of the beads enabled transfer of supernatant to a new tube. In the case of DNase I or MNase treatment, the samples were incubated 30 mins at RT with 1 μ L of stock solutions of either enzymes. Subsequently, lysates were incubated with 5 μ g antibody per IP for 1h at 4°C. 50 μ L of beads were then added and incubated for 1h at 4°C. Beads were then washed twice with 1mL cold ES2 buffer, once with 1mL TBS (35mM Tris-HCl, 140mM NaCl), and once with 0.05mM Tris-HCl pH 6.8. After removal of supernatant, beads were eluted with 50 μ L 0.05M Tris-HCl pH 6.8 added with 25 μ L 4X NuPage and 10 μ L DTT 1M, and incubation for 5 mins at 95°C. Supernatant was then run on Western Blot gels as any other sample.

Immunofluorescence & microscopy acquisition

Cells were harvested, centrifuged and permeabilized in 1mL PBS 0.3% Triton X-100. After 15 mins incubation at RT, 50 μ L paraformaldehyde (PFA) 16% was added and incubated for 15 mins. 1M nuclei were deposited dropwise in each well of 24-wells plate, previously prepared with 1mL glycerol cushions (30% Glycerol, 0.05% Triton W-100 into PBS 1X). Plates were centrifuged at 2500g for 15 mins, 4°C to plate the nuclei on the coverslips. Nuclei were crosslinked onto the coverslips by incubation with 300 μ L of PBS-PFA 0.16% for 4 mins at RT. After 2 washings using PBS 1X 0.1% Triton X-100 (PBST), a Blocking solution (PBST 3% BSA) was incubated for 1 hour at RT or O/N at 4°C. Nuclei were antibody-labelled with 500 μ L of 1/500-1/1000 dilutions of the relevant antibodies into Blocking solution for 2 hours at RT, or O/N at 4°C. After 10 washing using PBST, 500 μ L secondary antibodies into Blocking solution were added and incubated for 1h at RT. After 5 PBST washings, 5 μ g/mL DAPI into PBST was added for 7 mins, and then washed twice. Coverslips were then mounted onto glass slides using Mowiol. Cells were imaged on a LSM 880 Confocal Microscope with an Airyscan module to obtain high-resolution images.

Beaf32 & Pol II cluster detection

For Beaf32 and Pol II cluster detection, we made use of Cellpose(2) (Stringer *et al.*, 2021), a deep learning-based segmentation method developed from a diverse training dataset of cells and

nuclei. By combining the horizontal and vertical gradients predicted by a U-Net-shaped neural network, Cellpose generates vector fields from a topological map. In its first version, the authors propose full built-in models: Cytoplasm model ('cyto'), Nucleus model ('nuclei') and Cytoplasm 2.0 model ('cyto2'). In addition, in the last Version 2.2 users can train his own model from the base of one of these built-in models. We trained a custom model to detect in-nucleus clusters above 16 pixels. We used Fiji (Schindelin *et al.*, 2012) to develop our own process, in ImageJ1 macro language, applying the SCF plugin (<https://www.mpi-cbg.de/research/scientific-cores-support/scientific-services/scientific-computing-facility>) and Cellpose Segmentation, followed by the BIOP plugin (<https://github.com/BIOP>) to call Cellpose segmentation processing. The processing consists of a protein-cluster segmentation (Cellpose: "custom model") under the nuclei mask segmentation (Cellpose v2.2 model "nuclei"), and finally an area calculation for each cluster.

ChIP

5M cells per IP were crosslinked for 10 minutes with at RT 1/45V Formaldehyde (FA) 37% (Sigma F1635) for a final concentration of 0,8% FA. Crosslinking was stopped with 1/15 glycine 2M for 5 minutes at RT. After 2 washes with PBS 1X, cells were frozen at -80°C until ready to perform the following permeabilization for 20 minutes at RT with 500µL PBS supplemented with 0,2% Triton X-100 and 10mM NaBu. After centrifugation, pellets were washed with 1mL of Lysis buffer (LB: NaCl 140mM, HEPES pH 7,6 15mM, EDTA pH 8 1mM, EGTA 0,5mM, Triton X-100 1%, Sodium Desoxycholate 0,1%, DTT 0,5mM, sodium butyrate 10mM, protease inhibitor 1X (04693124001, Roche)) and resuspended in 100µL LB added with 1% SDS and 0,5% N-lauroylsarcosyl for 30 minutes at 4°C. Samples were then transferred to sonication tubes (C30010016, Diagenode) and sonicated for 5 cycles of 30 seconds ON and 30 seconds OFF at 4°C with the Bioruptor Pico (Diagenode). After centrifugation, the lysates were henceforth kept on ice at 4°C. To ensure proper chromatin fragmentation, a chromatin aliquot of each sample was added with RNase A and incubated for 30 minutes at 37°C, and then with Proteinase K for 2 hours at 55°C. After phenol-chloroform purification, loading dye was added and samples were run on a 1,5% agarose gel. Protein A or protein G beads were prepared by washing twice the required volume of beads with 1mL of LB supplemented with 0,1% SDS and 0,5% N-lauroylsarcosyl (hereafter named LB 0,1% SDS). Beads were then resuspended in 10V LB 0,1% SDS and incubated with 0,1mg/mL BSA NEB for 2 hours at 4°C. After two washings with LB 0,1% SDS, antibodies were linked to the beads by adding 20µL of prepared beads to 10µg of antibody, and completing to 400µL with LB 0,1% SDS for an incubation O/N at 4°C. Before immunoprecipitation, chromatin samples were pre-cleared O/N at 4°C in 10V LB with 10µL of prepared beads without antibody. Antibody-linked beads were washed four times with LB 0,1% SDS before use. 10% of the volume of each pre-cleared sample was taken for input. Pre-cleared chromatin samples were incubated with 10µL of antibody-linked beads for 4 hours at 4°C.

The beads were then washed sequentially with LB 0,1% SDS (four times) and TE (twice). Elution was carried out at 70°C and 1000rpm, first for 20 minutes with EDTA pH 8 10mM, 1% SDS and 50mM of Tris-HCl, and secondly with SDS 0,67% in TE 1X. Crosslinking was reversed O/N at 65°C. IP samples and inputs were then incubated at 37°C for 30 minutes with RNase A, and then at 55°C for 2 hours with 250µL of TE supplemented with 140mg/mL of glycogen and 400µg/mL of proteinase K. DNA was purified by phenol-chloroform purification, and then precipitated by incubation with 1,3mL of ethanol absolute per sample for 30 minutes at -80°C. After 30 minutes of centrifugation at max speed at 4°C, DNA pellets were washed with 1mL of ethanol 70%, centrifuged for 10 minutes at max speed at 4°C, and dried at RT to be resuspended in water. For analysis by sequencing, libraries were prepared using the Colibri PS DNA Library Prep Kit with PCR amplification for Illumina following manufacturer instructions (ThermoFisher Cat.A38613024) and sequenced by BGI Genomics at around 15M 100bp Paired-End reads per sample. For qPCR analysis, inputs were diluted at 1/100 and IP were diluted at 1/20, and measured using SYBR Green (1725124, BioRad) and a Viia 7 Real-Time PCR System (ThermoFisher Scientific) machine. Primers used for amplifications were selected for amplicons between 80 & 150bp.

HiC

HiC was performed on 2M cells crosslinked as in ChIP, using the Arima Genomics HiC kit, following manufacturer instructions. After quality control of the purified DNA, sonication was performed by 7-8 cycles of 15 secs ON and 90 secs OFF on Bioruptor Pico (Diagenode). After verification of fragment-size homogeneity between samples, libraries were prepared using the KAPA Hyper prep Kit (Roche Cat#KK8500) following manufacturer instruction. Sequencing was performed by BGI Genomics, aiming for around 100M 100bp Paired-End reads per sample.

iCLIP

iCLIP was performed as described in Lee et al., 2021, Biorxiv, using 20M S2 cell per IP as input material, and performing UV crosslinks at 254nm, at either 100, 200 or 400mJ/cm². For primer labelling, IR-dye-800CW-DBCO was replaced by ATTO647.

CRISPR-dCas9

Cloning of CRISPR-dCas9 plasmids with gRNAs added with ncRNAs was adapted from Sigova et al. 2015. Shortly, a dCas9-EGFP amplified from another plasmid was inserted by Gibson protocol (Pro Ligation-free Cloning Kit, ABM) into a pAc-sgRNA-Cas9 (Addgene p49330). gRNA insertion site & scaffold cassette were removed by PCR, to allow for ligation-free insertion of gBlocks containing the gRNA, Cas9 scaffold and eRNA, or only gRNA and scaffold as control. After amplification and complete

sequencing of each plasmid, they were transfected into S2 cells. Selection using 5mg/mL Puromycin on cells diluted to 1M/mL was performed starting 48h after transfection. After a 48h selection, proper transfection and selection was confirmed by observation of GFP fluorescence. Cells were then harvested and crosslinked for ChIP, and non-crosslinked aliquots were kept for RNA extraction and analysis.

Antibodies

For α Beaf32 and α CP190, homemade polyclonal antibodies were used, taking 5 μ g affinity purified antibody per IP for Beaf32 and 1 μ L of unpurified serum for CP190 (Lhoumaud *et al.*, 2014; Liang *et al.*, 2014; Vogelmann *et al.*, 2014). For α HA, two antibodies were used: Monoclonal Mouse α HA tag produced in mouse, clone GT423 (MilliporeSigma, Cat. SAB2702196); and polyclonal α HA produced in rabbit (MilliporeSigma Cat.H6908). For Cohesin ChIP, a rabbit α SMC3 antibody (Abcam Cat. Ab9263) was used. Pol2 immunofluorescence were performed using monoclonal mouse antibody clone 8WG16 (MerckMillipore Cat. 05-952-I). For H3 and Act controls, α H3 rabbit polyclonal (Abcam Cat. Ab1791) and α Act mouse monoclonal clone C4 (MerckMillipore Cat. MAB1501) were used. For GFP-dCas9 ChIP, α GFP rabbit polyclonal (Cliniscience, Cat. TP-401) was used. Finally, the Mtr4 antibody used was the α Mtr4(SKIV2L2) rabbit polyclonal antibody (Bethyl laboratories Cat. A300-614A).

Bioinformatic analysis

RNAseq

Adapter sequences trimming and quality filtering of reads was performed using fastp (v 0.22.0 (Chen *et al.*, 2018)), with a quality threshold of 30 (-q 30). Reads from ribosomal RNAs were also dropped using SortMeRNA (v. 4.3.2 (Kopylova, Noé and Touzet, 2012)) and the default databases (https://github.com/sortmerna/sortmerna/tree/master/data/rRNA_databases). Reads were then aligned to the *Drosophila melanogaster* genome (v. dm6_r6.13) using STAR (v. 2.6.0c (Dobin *et al.*, 2013)). PCR duplicates, unmapped reads and reads aligned with low quality were then removed (-q20 -F 1548). Read quantification on known genes was performed using htseq-count (v. 2.0.2 (Putri *et al.*, 2022)) by requiring that reads should be in a reverse stranded configuration (- -stranded reverse). With this parameter, first reads (R1) are expected to be on the complementary strand and second reads (R2) are expected to be on the forward strand of genes.

In order to detect non-coding RNAs, exonic reads were first removed. To do this, coordinates of all known exons were extracted. Bedtools (v. 2.29.0 (Quinlan and Hall, 2010)) was then used to keep only

non-exonic reads. Non-exonic reads were required to either overlap non-coding DNA or have first reads on the forward strand of exons and second reads on the complementary strand. Bam files containing non-exonic reads were then used to identify non-coding RNAs using the R package *icetea* (v. 1.16.0 (Bhardwaj *et al.*, 2019)). The function `detectTSS` was run using the following parameters: read-counts over background should have a fold-change over 5 and window size of 200. The non-coding RNAs detected in all samples were then merged in to the same bed file for further usage.

Bed file containing all the detected non-coding RNAs was then imported into R and read-quantification per non-coding RNAs was then performed using feature counts from *Rsubread* package (v. 2.12.3 (Liao, Smyth and Shi, 2019)). Reads were required to be in the reverse stranded configuration (`--strandSpecific 2`), similar to the mRNAs read quantification. A *DESeq2* object was then generated using these quantifications for differential expression analysis.

Differential expression for both mRNAs and non-coding RNAs was performed using *DESeq2* (v 1.38.0 (Love, Huber and Anders, 2014)), and *lfcShrink* was used to obtain differentially expressed genes (Zhu, Ibrahim and Love, 2019). Genes and ncRNAs were considered significantly differentially expressed if the absolute values of their log₂ fold-changes were over 0.7 and BH adjusted p. values were less than 0.05. The R package *clusterProfiler* (Wu *et al.*, 2021) was used for over-representation analysis of differentially expressed genes.

ChIPseq

cutadapt (v. 2.1) was used to trim adapter sequences and remove bad quality reads (-q 20 minimum-length 30). *Bwa* was then used to align reads to dm6 genome, then duplicated reads were removed using *picard* (v. 2.18.2). *BigWig* coverage files, with a bin size of 10 bases, were generated using *deeptools* by normalizing with RPGC (Reads per Genome Content) and a scaling factor calculated by *csaw* R package (Lun and Smyth, 2015). Scale factors for each protein of interest were calculated by counting reads in all samples with window sizes of 10 kb (kilo bases) and 300 bp (base pairs). The 10 kb window sizes quantified reads in the background, while the second window size estimated read counts in peaks, as the fragment lengths had a median size of 300 bases. Normalization factors were then calculated within windows showing at least 5 times more read counts than the background signal. *BigWigCompare* was used to generate differential ChIP-Seq tracks between conditions. Peaks were called using *macs2* with a threshold of either p value 0.05 or a q value of 0.05 depending on the data. Peaks were then centered to a 500 bp wide window around their summits. Heatmaps and average profiles were performed using *seqplots* (v. 1.23.3 (Stempor and Ahringer, 2016)). K-means clustering

on the ChIP-Seq heatmaps were performed by using the values in the central 500 bp window around the summits of peaks. The optimal k was found by performing different k values and settling on the one that do not result in redundant clusters with regard to enrichment with the different proteins considered and their combinations.

HiC

HiC files were generated using the Juicer pipeline (Juicer version 1.6; BWA 0.7.17-r1188; Juicer Tools Version 1.19.02 (Durand *et al.*, 2016); openjdk version "1.8.0_312"), genome: dm6 (r6.32). Replicate-merged data in .hic format were converted to mcool files using hicexplorer (v. 3.7.2), before being imported as cooler objects (Cooler v 0.9.1). The 10k resolution contacts maps were used to calculate insulation scores but also to perform Aggregated Peak Analysis.

Data from the different conditions were first imported and down-sampled using cooltools (v. 0.5.4), so that all samples could have the same total number of cis contacts. The down-sampling was performed to meet the lowest number of cis contacts observed among the samples considered. After down-sampling, observed over expected values were calculated for each sample. The matrices were then balanced using cooltools' balance_cooler function, by ignoring the first 2 bins of the diagonal in the contact matrix, meaning the value for bin self-interaction and the contacts between neighbouring bins.

For contacts versus distance plots, the extracted values are balanced, averaged and smoothed to decrease the variability at very long distances.

Insulation scores were calculated using the diamond insulation score method implemented in cooltools with a sliding diamond window of 50 kb. Scores per bin were exported as csv files for further statistical analyses and plotting in R. Insulation capacity (or Insulation Score Loss) correspond to the lagged differences between the central (Peak summit) bin and the surrounding bins outside of the valley, namely at +/- 4 to 6 bins corresponding to the edges of the windows used to calculate the insulation scores. Wilcoxon test was performed to evaluate the significance of differences in these values between conditions.

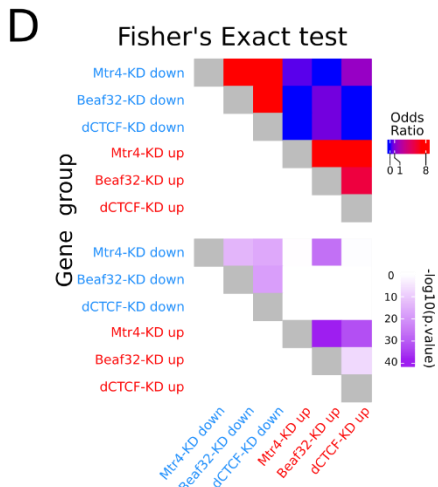
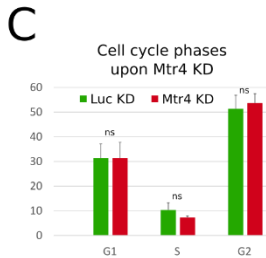
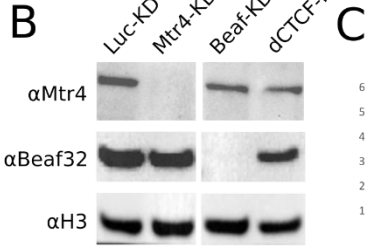
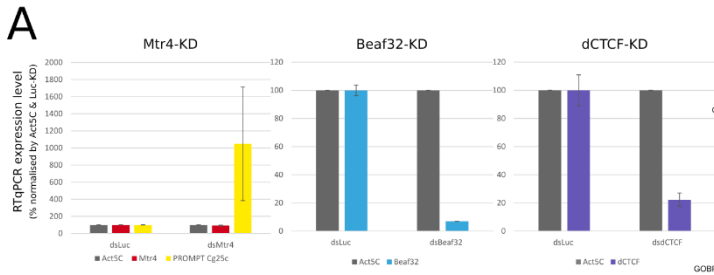
Aggregated peak analysis was performed by using coolpuppy (v1.1.0) and cooltools (Flyamer, Illingworth and Bickmore, 2020) with a minimum distance of 20kb and a maximum distance of 500kb.

APPENDIX

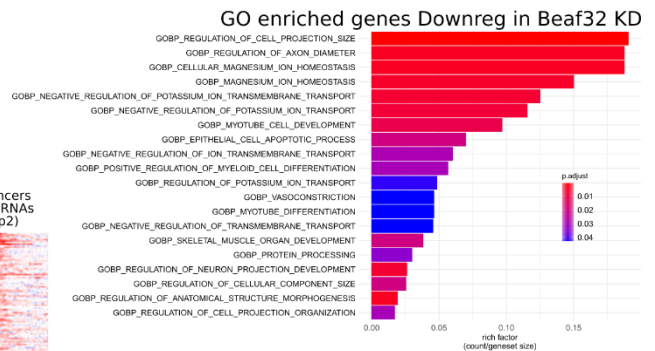
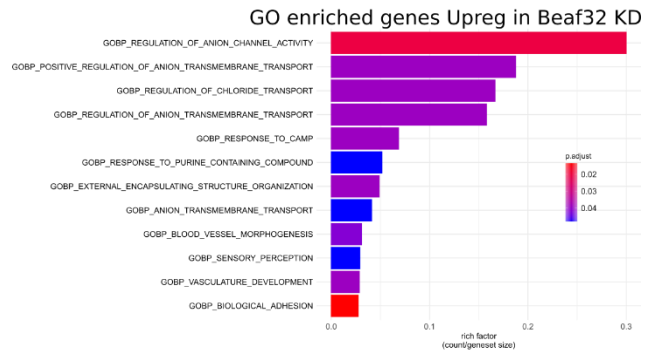
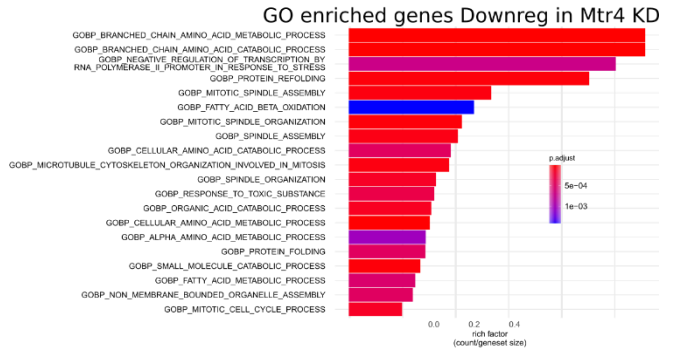
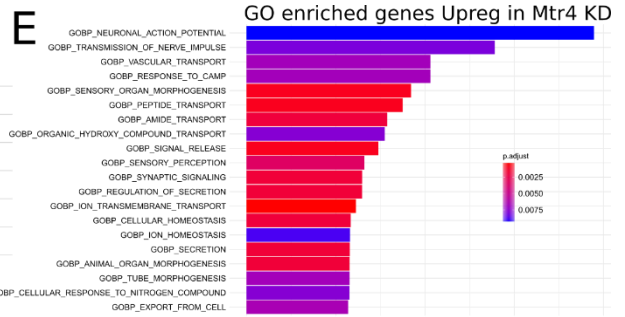
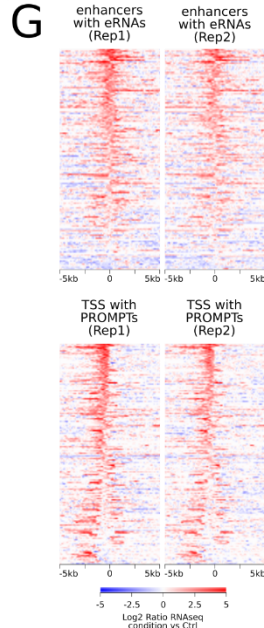
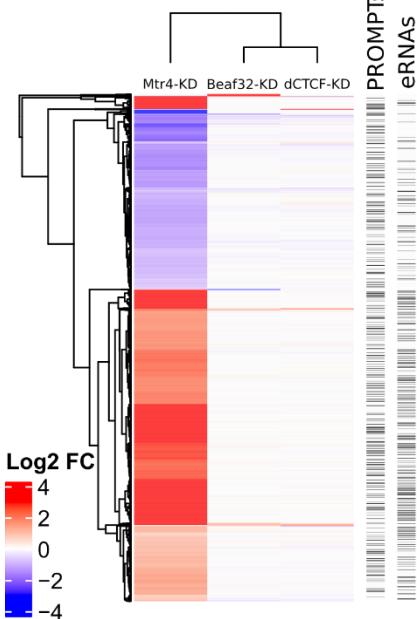
Supplementary figures - Chapter 1

Figure S1: Depletion controls and deregulated genes GO analysis

- A) RTqPCR validations of depletions, normalised by Act5C.
- B) Western Blot validations of depletion.
- C) FACS detections of cell-cycle phases for Mtr4 depleted cells, biological replicates n=3
- D) Fisher exact test between genes up- or down-regulated upon Mtr4Kd, Beaf32 KD and dCTCF KD, taking all expressed genes as base set.
- E) Enriched GO of biological processes (BP) for Mtr4 KD and Beaf32 up- and down-regulated genes. No enriched GO was found for dCTCF KD.
- F) Clustered heatmaps of the FC of deregulated ncRNAs detected in all 3 conditions, coloured by the intensity of their FC.
- G) Heatmaps plotting the Log₂ ratio of RNAseq non-exonic reads centred on Promoters with accumulated PROMPTs and enhancers with accumulated eRNAs.
- H) Volcano plot of the differentially accumulated annotated ncRNAs upon Mtr4 KD.



F DE ncRNAs all conditions vs Ctrl (n = 1541)



H Differential expressed stable ncRNAs Mtr4 KD vs Luc KD

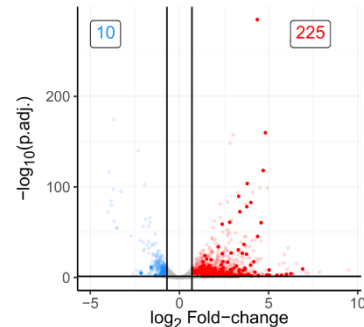


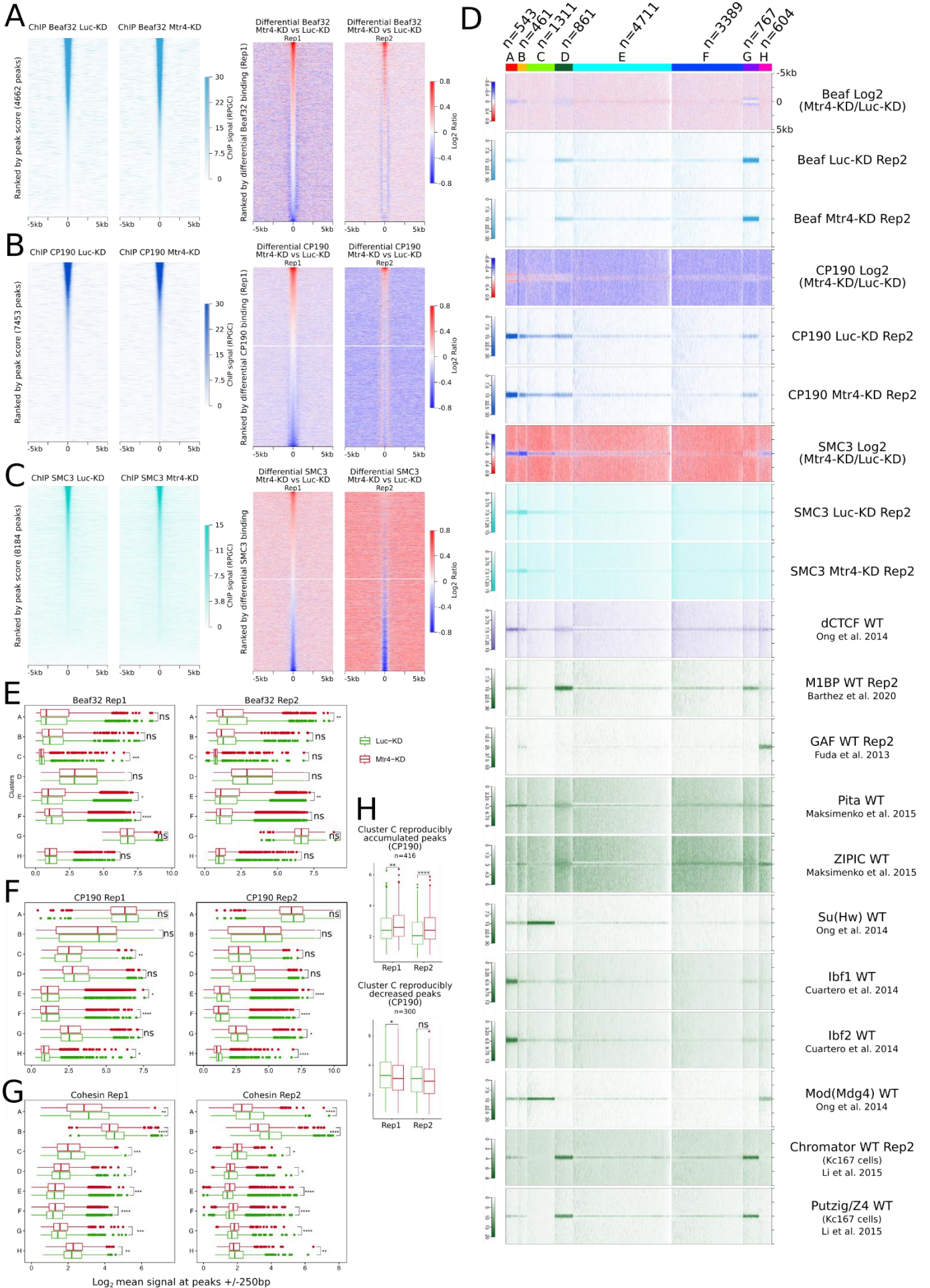
Figure S2: ChIPseq validations and reproducibility of differences

A, B, and C) Heatmaps plotting ChIPseq normalised as in clustered heatmap (fig2A), ranked by peak score (left panels) and by differential binding in Rep 1 (right panels).

D) Same clustered heatmaps as in Fig2A, but made with all available replicates (Beaf32, CP190, SMC3, M1BP, GAF and Chromator).

E, F, and G) Boxplots of Beaf32 (E), CP190(F) and Cohesin (G) separated by clusters in control and Mtr4 KD in both replicates, plotting the 500bp around peak summits in Log2. P-values are calculated by Wilcoxon two-sided tests.

I) Boxplots showing only Cluster C peaks that are above 1 in FC in both replicates (top), and beneath 1 in FC in both replicates (bottom).



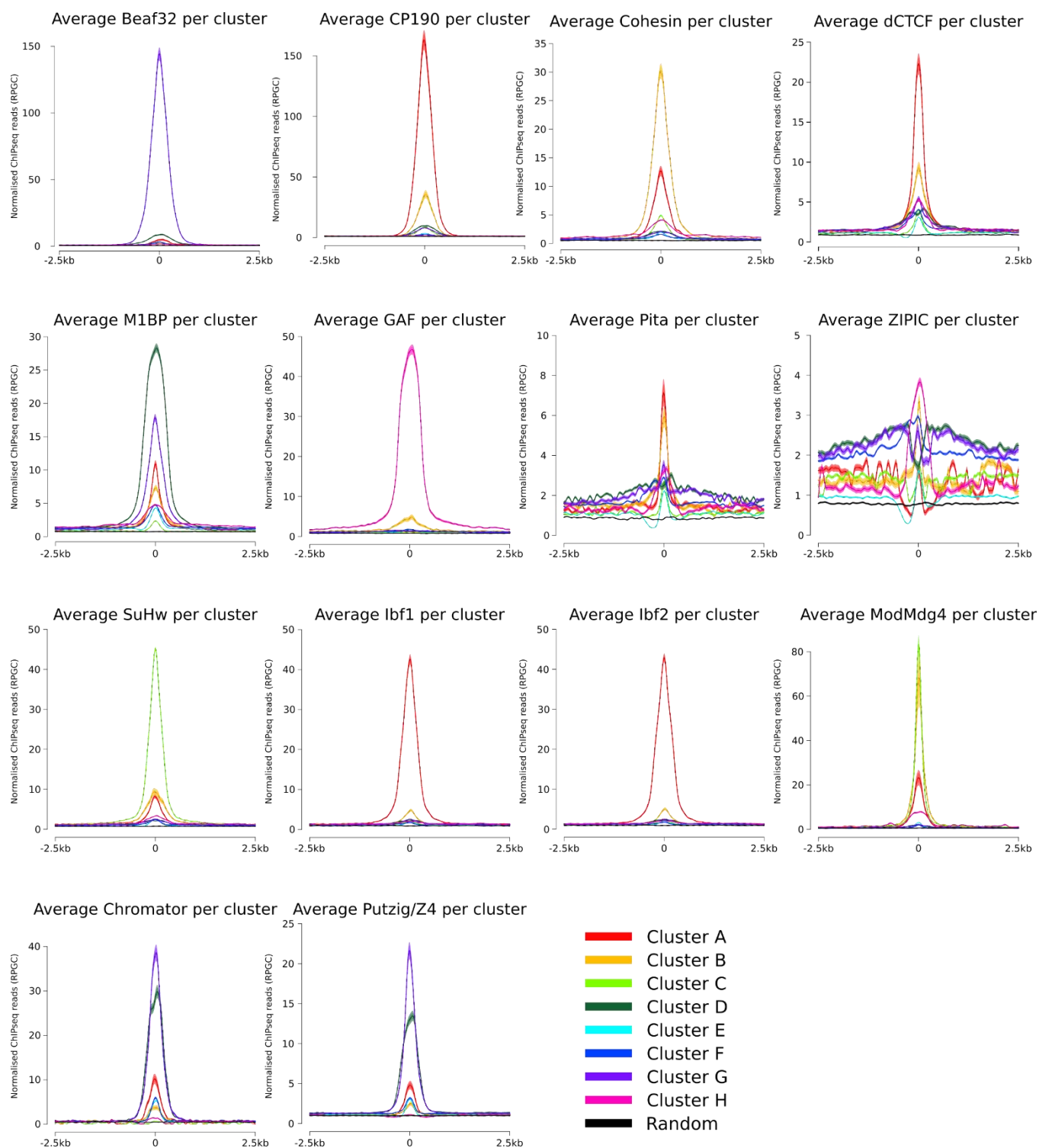


Figure S3: Profile of each IBP in each cluster

Profiles of ChIPseq reads normalised by RPGC for all peaks from each cluster compared to a control of random sites (same number of sites as the biggest cluster, namely E, with 4630 peaks).

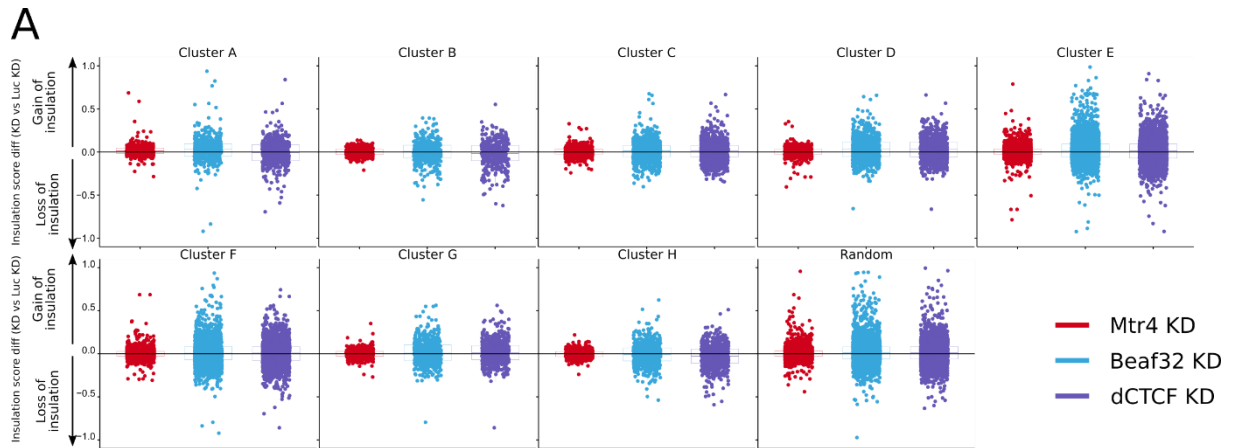


Figure S4: Quantification of Insulation Score

Boxplots equivalents to figure 5C showing the differential Insulation Score between conditions vs control, on the central bin. All Wilcoxon test compared to Control are non-significant.

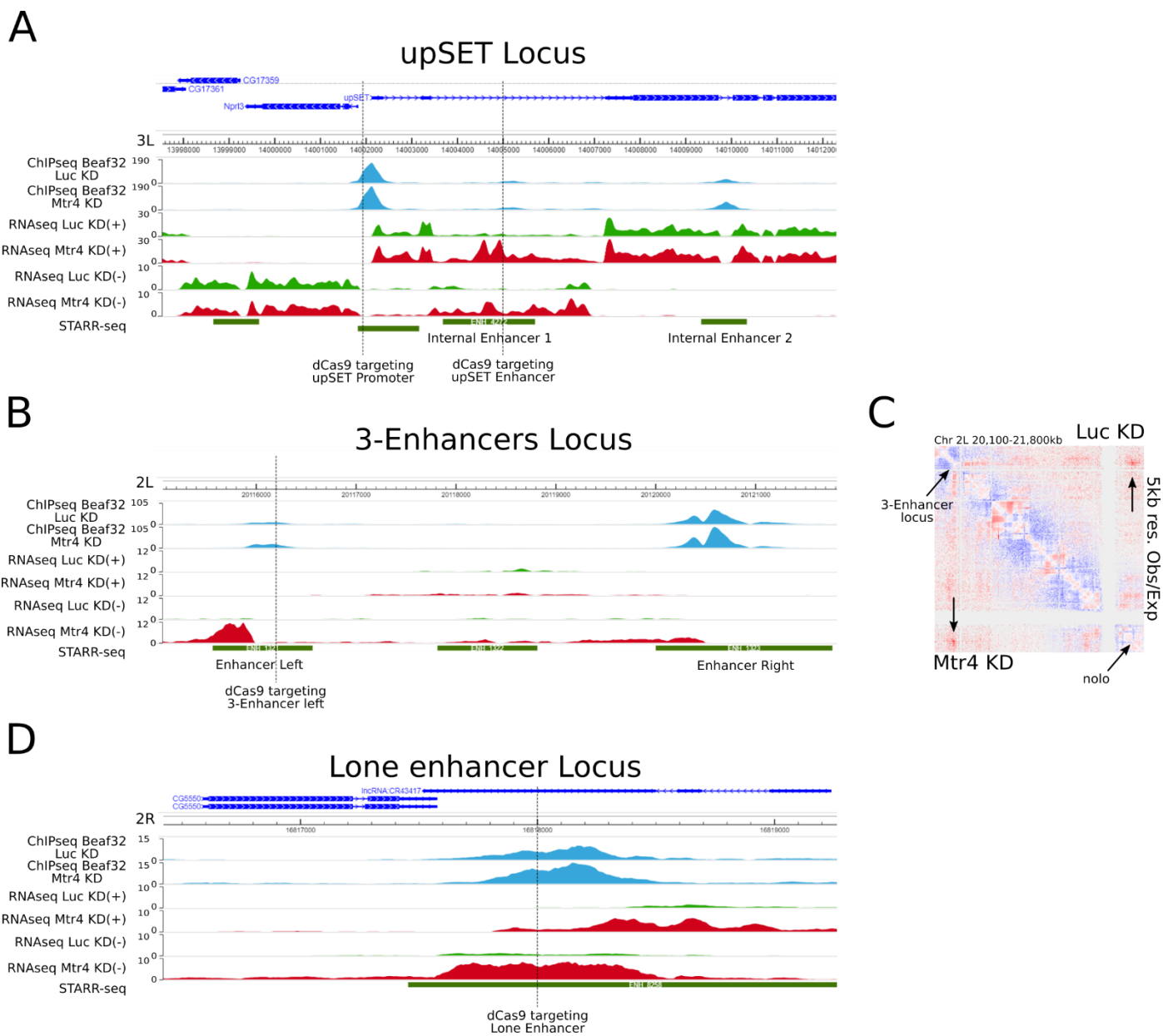


Figure S5: CRISPR-dCas9 loci

A, B and D) browser shots obtained from the WashU browser, respectively of the upset locus, the 3-Enhancers Locus, and the Lone Enhancer locus. C) HiC snapshot from Juicebox comparing Luc KD (top right) and Mtr4 KD (bottom left) at a 5kb resolution, Observed/Expected, SquareRoot normalised.

Supplementary figures - Chapter 2

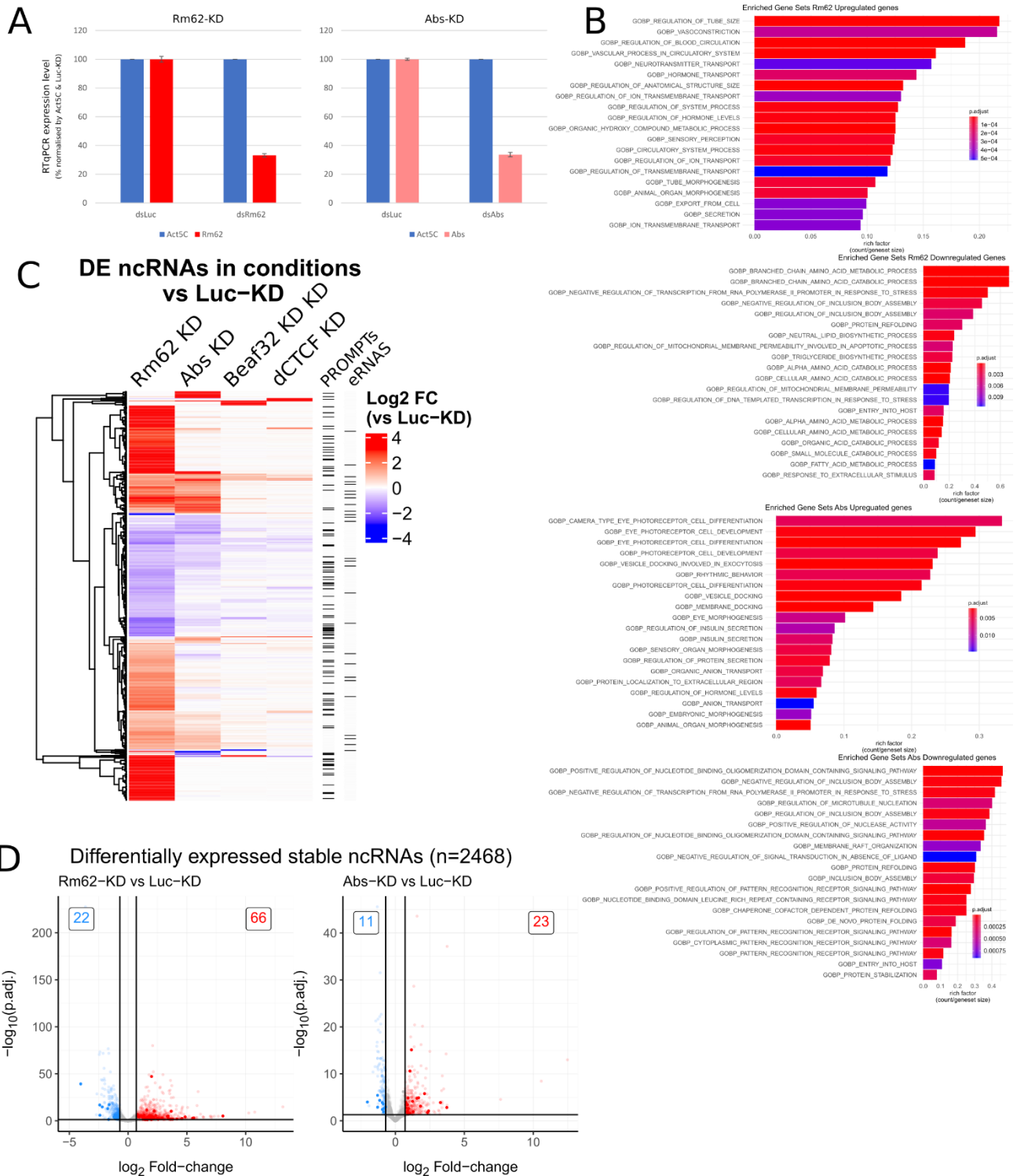


Figure S1: Depletion controls and deregulated genes GO analysis

- RTqPCR validations of depletions, normalised by Act5C.
- Enriched GO of biological processes (BP) for Rm62 KD and Abs up- and down-regulated genes.
- Clustered heatmaps of the FC of deregulated ncRNAs detected in all 4 conditions, coloured by the intensity of their FC.
- Volcano plot of the differentially accumulated annotated ncRNAs upon Rm62 KD (*Left*) and Abs KD (*Right*).

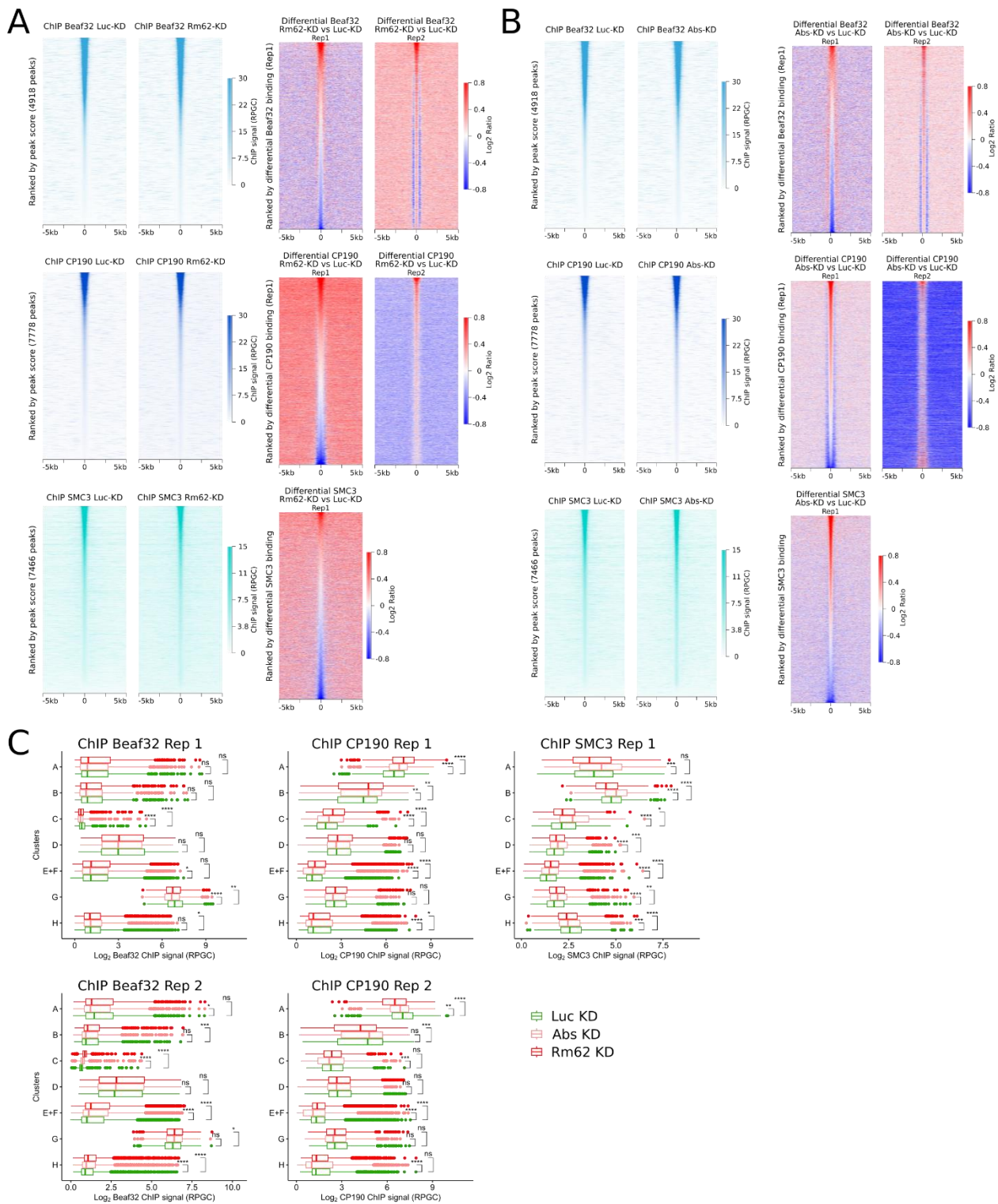


Figure S2: CHIPseq validations and reproducibility of differences

A) & B) Heatmaps plotting CHIPseq normalised as in clustered heatmap (fig2A), ranked by peak score (left panels) and by differential binding in Rep 1 (right panels). In A) are shown all CHIPseq in Rm62 KD, while B) corresponds to Abs KD.

C) Boxplots of Beaf32 (*Left*), CP190 (*Middle*) and Cohesin (*Right*) separated by clusters in control, Rm62 KD and Abs KD in both replicates, plotting the 500bp around peak summits in Log₂. P-values are calculated by Wilcoxon two-sided tests.

Bibliography

Abdalla, M.O.A. *et al.* (2019) 'The Eleanor ncRNAs activate the topological domain of the ESR1 locus to balance against apoptosis', *Nature Communications*, 10(1). Available at: <https://doi.org/10.1038/s41467-019-11378-4>.

Allison, D.F. and Wang, G.G. (2019) 'R-loops: Formation, function, and relevance to cell stress', *Cell Stress*. Shared Science Publishers OG, pp. 38–46. Available at: <https://doi.org/10.15698/cst2019.02.175>.

Almada, A.E. *et al.* (2013) 'Promoter directionality is controlled by U1 snRNP and polyadenylation signals', *Nature*, 499(7458). Available at: <https://doi.org/10.1038/nature12349>.

Alvarez-Dominguez, J.R. *et al.* (2017) 'The Super-Enhancer-Derived alncRNA-EC7/Bloodline Potentiates Red Blood Cell Development in trans', *Cell Reports*, 19(12), pp. 2503–2514. Available at: <https://doi.org/10.1016/j.celrep.2017.05.082>.

Amankwaa, B., Schoborg, T. and Labrador, M. (2022) 'Drosophila insulator proteins exhibit in vivo liquid-liquid phase separation properties', *Life Science Alliance*, 5(12). Available at: <https://doi.org/10.26508/lsa.202201536>.

Amy C. Seila *et al.* (2008) 'Divergent Transcription from Active Promoters', *Science*, 322(5909), pp. 1845–1848. Available at: <https://doi.org/10.1126/science.1162228>.

Andersson, R. *et al.* (2014) 'An atlas of active enhancers across human cell types and tissues', *Nature*, 507(7493). Available at: <https://doi.org/10.1038/nature12787>.

Andersson, R. *et al.* (2015) 'Human Gene Promoters Are Intrinsically Bidirectional', *Molecular Cell*. Available at: <https://doi.org/10.1016/j.molcel.2015.10.015>.

Andrisani, O. *et al.* (2022) 'Biological functions of DEAD/DEAH-box RNA helicases in health and disease', *Nature Immunology*. Nature Research, pp. 354–357. Available at: <https://doi.org/10.1038/s41590-022-01149-7>.

Armache, A. *et al.* (2020) 'Histone H3.3 phosphorylation amplifies stimulation-induced transcription', *Nature*, 583(7818), pp. 852–857. Available at: <https://doi.org/10.1038/s41586-020-2533-0>.

Arnould, C. *et al.* (no date) 'ATM-dependent formation of a novel chromatin compartment regulates the Response to DNA Double Strand Breaks and the biogenesis of translocations'. Available at: <https://doi.org/10.1101/2021.11.07.467654>.

Atianand, M.K. *et al.* (2016) 'A Long Noncoding RNA lincRNA-EPS Acts as a Transcriptional Brake to Restrain Inflammation', *Cell*, 165(7), pp. 1672–1685. Available at: <https://doi.org/10.1016/j.cell.2016.05.075>.

- Bag, I. *et al.* (2021) 'M1BP cooperates with CP190 to activate transcription at TAD borders and promote chromatin insulator activity', *Nature Communications*, 12(1). Available at: <https://doi.org/10.1038/s41467-021-24407-y>.
- Banerji, J., Rusconi, S. and Schaffner, W. (1981) *Expression of a β -Globin Gene Is Enhanced by Remote SV40 DNA Sequences*, *Cell*.
- Banigan, E.J. *et al.* (2023) 'Transcription shapes 3D chromatin organization by interacting with loop extrusion', *Proceedings of the National Academy of Sciences*, 120(11). Available at: <https://doi.org/10.1073/pnas.2210480120>.
- Barisic, D. *et al.* (2019) 'Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors', *Nature*, 569(7754), pp. 136–140. Available at: <https://doi.org/10.1038/s41586-019-1115-5>.
- Batut, P.J. *et al.* (2022) 'Genome organization controls transcriptional dynamics during development', *Science* [Preprint]. Available at: <https://www.science.org>.
- Beagan, J.A. and Phillips-Cremins, J.E. (2020) 'On the existence and functionality of topologically associating domains', *Nature Genetics*, 52(1), pp. 8–16. Available at: <https://doi.org/10.1038/s41588-019-0561-1>.
- Beltran, M. *et al.* (2016) 'The interaction of PRC2 with RNA or chromatin is mutually antagonistic', *Genome Research*, 26(7), pp. 896–907. Available at: <https://doi.org/10.1101/gr.197632.115>.
- Bhardwaj, V. *et al.* (2019) 'MAPCap allows high-resolution detection and differential expression analysis of transcription start sites', *Nature Communications*, 10(1). Available at: <https://doi.org/10.1038/s41467-019-11115-x>.
- Bickmore, W.A. and Van Steensel, B. (2013) 'Genome architecture: Domain organization of interphase chromosomes', *Cell*. Available at: <https://doi.org/10.1016/j.cell.2013.02.001>.
- Blank-Giwojna, A., Postepska-Igielska, A. and Grummt, I. (2019) 'lncRNA KHPS1 Activates a Poised Enhancer by Triplex-Dependent Recruitment of Epigenomic Regulators', *Cell Reports*, 26(11), pp. 2904-2915.e4. Available at: <https://doi.org/10.1016/j.celrep.2019.02.059>.
- Boeke, J. *et al.* (2011) 'The RNA helicase Rm62 cooperates with SU(VAR)3-9 to re-silence active transcription in *Drosophila melanogaster*', *PLoS ONE*, 6(6). Available at: <https://doi.org/10.1371/journal.pone.0020761>.
- Boeynaems, S. *et al.* (2019) 'Spontaneous driving forces give rise to protein–RNA condensates with coexisting phases and complex material properties', *Proceedings of the National Academy of Sciences of the United States of America*, 116(16), pp. 7889–7898. Available at: <https://doi.org/10.1073/pnas.1821038116>.
- Boija, A. *et al.* (2018) 'Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains', *Cell*, 175(7), pp. 1842-1855.e16. Available at: <https://doi.org/10.1016/j.cell.2018.10.042>.

Bose, D.A. *et al.* (2017) 'RNA Binding to CBP Stimulates Histone Acetylation and Transcription', *Cell*, 168(1–2), pp. 135–149.e22. Available at: <https://doi.org/10.1016/j.cell.2016.12.020>.

Brenner, S., Jacob, F. and Meselson, M. (1961) 'An unstable intermediate carrying information from genes to ribosomes for protein synthesis', *Nature*, 190(4776). Available at: <https://doi.org/10.1038/190576a0>.

Buszczak, M. and Spradling, A.C. (2006) 'The Drosophila P68 RNA helicase regulates transcriptional deactivation by promoting RNA release from chromatin', *Genes and Development*, 20(8), pp. 977–989. Available at: <https://doi.org/10.1101/gad.1396306>.

Büttner, K., Nehring, S. and Hopfner, K.P. (2007) 'Structural basis for DNA duplex separation by a superfamily-2 helicase', *Nature Structural and Molecular Biology*, 14(7), pp. 647–652. Available at: <https://doi.org/10.1038/nsmb1246>.

Cai, H. and Levine, M. (1995) 'Modulation of enhancer-promoter interactions by insulators in the Drosophila embryo', *Nature* [Preprint].

Cai, Y. *et al.* (2021) 'H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions', *Nature Communications*, 12(1). Available at: <https://doi.org/10.1038/s41467-021-20940-y>.

Camats, M. *et al.* (2008) 'P68 RNA helicase (DDX5) alters activity of Cis- and trans-acting factors of the alternative splicing of H-Ras', *PLoS ONE*, 3(8). Available at: <https://doi.org/10.1371/journal.pone.0002926>.

Cardozo Gizzi, A.M. *et al.* (2019) 'Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms', *Molecular Cell*, 74(1), pp. 212–222.e5. Available at: <https://doi.org/10.1016/j.molcel.2019.01.011>.

Cargill, M., Venkataraman, R. and Lee, S. (2021) 'DEAD-Box RNA Helicases and Genome Stability', *Genes*, p. 12. Available at: <https://doi.org/10.3390/genes>.

Carrozza, M.J. *et al.* (2005) 'Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription', *Cell*, 123(4), pp. 581–592. Available at: <https://doi.org/10.1016/j.cell.2005.10.023>.

Carullo, N.V.N. *et al.* (2020) 'Enhancer RNAs predict enhancer–gene regulatory links and are critical for enhancer function in neuronal systems', *Nucleic Acids Research*, 48(17), pp. 9550–9570. Available at: <https://doi.org/10.1093/nar/gkaa671>.

Cattoni, D.I. *et al.* (2017) 'Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions', *Nature Communications*, 8(1). Available at: <https://doi.org/10.1038/s41467-017-01962-x>.

Cavalheiro, G.R. *et al.* (2023) 'CTCF, BEAF-32, and CP190 are not required for the establishment of TADs in early Drosophila embryos but have locus-specific roles', *SCIENCE ADVANCES* [Preprint]. Available at: <https://www.science.org>.

- Chen, Q. *et al.* (2023) 'Enhancer RNAs in transcriptional regulation: recent insights', *Frontiers in Cell and Developmental Biology*. Frontiers Media S.A. Available at: <https://doi.org/10.3389/fcell.2023.1205540>.
- Chen, S. *et al.* (2018) 'Fastp: An ultra-fast all-in-one FASTQ preprocessor', in *Bioinformatics*. Available at: <https://doi.org/10.1093/bioinformatics/bty560>.
- Chen, X. *et al.* (2019) 'Key role for CTCF in establishing chromatin structure in human embryos', *Nature*, 576(7786). Available at: <https://doi.org/10.1038/s41586-019-1812-0>.
- Chen, Y. *et al.* (2016) 'Principles for RNA metabolism and alternative transcription initiation within closely spaced promoters', *Nature Genetics*, 48(9), pp. 984–994. Available at: <https://doi.org/10.1038/ng.3616>.
- Cheng, W. *et al.* (2018) 'DDX5 RNA helicases: Emerging roles in viral infection', *International Journal of Molecular Sciences*. Available at: <https://doi.org/10.3390/ijms19041122>.
- Cho, W.-K. *et al.* (2018) 'Mediator and RNA polymerase II clusters associate in transcription-dependent condensates', *Science*, 361, pp. 412–415. Available at: <https://www.science.org>.
- Churchman, L.S. and Weissman, J.S. (2011) 'Nascent transcript sequencing visualizes transcription at nucleotide resolution', *Nature*, 469(7330). Available at: <https://doi.org/10.1038/nature09652>.
- Cifuentes-Rojas, C. *et al.* (2014) 'Regulatory Interactions between RNA and Polycomb Repressive Complex 2', *Molecular Cell*, 55(2), pp. 171–185. Available at: <https://doi.org/10.1016/j.molcel.2014.05.009>.
- Cobb, M. (2017) '60 years ago, Francis Crick changed the logic of biology', *PLoS Biology*, 15(9). Available at: <https://doi.org/10.1371/journal.pbio.2003243>.
- Core, L.J. *et al.* (2014) 'Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers', *Nature Genetics*, 46(12), pp. 1311–1320. Available at: <https://doi.org/10.1038/ng.3142>.
- Core, L.J., Waterfall, J.J. and Lis, J.T. (2008) 'Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters', *Science*, 322(5909), pp. 1845–1848. Available at: <https://doi.org/10.1126/science.1162228>.
- Cramer, P. (2019) 'Organization and regulation of gene transcription', *Nature*. Nature Publishing Group, pp. 45–54. Available at: <https://doi.org/10.1038/s41586-019-1517-4>.
- Cremer, T. and Cremer, C. (2001) *CHROMOSOME TERRITORIES, NUCLEAR ARCHITECTURE AND GENE REGULATION IN MAMMALIAN CELLS*. Available at: www.nature.com/reviews/genetics.
- Cuartero, S. *et al.* (2014) 'Ibf1 and Ibf2 are novel CP190-interacting proteins required for insulator function', *EMBO Journal*, 33(6), pp. 637–647. Available at: <https://doi.org/10.1002/embj.201386001>.

Dahm, R. (2005) 'Friedrich Miescher and the discovery of DNA', *Developmental Biology*. Academic Press Inc., pp. 274–288. Available at: <https://doi.org/10.1016/j.ydbio.2004.11.028>.

Dardenne, E. *et al.* (2012) 'Splicing switch of an epigenetic regulator by RNA helicases promotes tumor-cell invasiveness', *Nature Structural and Molecular Biology*, 19(11). Available at: <https://doi.org/10.1038/nsmb.2390>.

Dardenne, E. *et al.* (2014) 'RNA Helicases DDX5 and DDX17 Dynamically Orchestrate Transcription, miRNA, and Splicing Programs in Cell Differentiation', *Cell Reports*, 7(6). Available at: <https://doi.org/10.1016/j.celrep.2014.05.010>.

Davidson, I.F. *et al.* (2019) 'DNA loop extrusion by human cohesin', *Science* [Preprint]. Available at: <https://www.science.org>.

Davidson, I.F. *et al.* (2023) 'CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion', *Nature*, 616(7958), pp. 822–827. Available at: <https://doi.org/10.1038/s41586-023-05961-5>.

Davidson, I.F. and Peters, J.M. (2021) 'Genome folding through loop extrusion by SMC complexes', *Nature Reviews Molecular Cell Biology*. Nature Research, pp. 445–464. Available at: <https://doi.org/10.1038/s41580-021-00349-7>.

Dekker, J. *et al.* (2002) 'Capturing Chromosome Conformation', *Science* [Preprint]. Available at: <https://www.science.org>.

Depierre, D. *et al.* (2023) 'Chromatin in 3D distinguishes dMes-4/NSD and Hyph/dSet2 in protecting genes from H3K27me3 silencing', *Life Science Alliance*, 6(11), p. e202302038. Available at: <https://doi.org/10.26508/lsa.202302038>.

Dietzel, S. *et al.* (1999) *The 3D Positioning of ANT2 and ANT3 Genes within Female X Chromosome Territories Correlates with Gene Activity*. Available at: <http://www.idealibrary.com>.

Djebali, S. *et al.* (2012) 'Landscape of transcription in human cells', *Nature*, 489(7414). Available at: <https://doi.org/10.1038/nature11233>.

Dobin, A. *et al.* (2013) 'STAR: Ultrafast universal RNA-seq aligner', *Bioinformatics*, 29(1). Available at: <https://doi.org/10.1093/bioinformatics/bts635>.

Donsbach, P. and Klostermeier, D. (2021) 'Regulation of RNA helicase activity: Principles and examples', *Biological Chemistry*. De Gruyter Open Ltd, pp. 529–559. Available at: <https://doi.org/10.1515/hsz-2020-0362>.

Dorsett, D. (2019) 'The Many Roles of Cohesin in Drosophila Gene Transcription', *Trends in Genetics*. Elsevier Ltd, pp. 542–551. Available at: <https://doi.org/10.1016/j.tig.2019.04.002>.

Dorsett, D. and Merckenschlager, M. (2013) 'Cohesin at active genes: A unifying theme for cohesin and gene expression from model organisms to humans', *Current Opinion in Cell Biology*, pp. 327–333. Available at: <https://doi.org/10.1016/j.ceb.2013.02.003>.

- Downen, J.M. *et al.* (2014) 'Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes', *Cell*, 159(2), pp. 374–387. Available at: <https://doi.org/10.1016/j.cell.2014.09.030>.
- Dunham, I. *et al.* (2012) 'An integrated encyclopedia of DNA elements in the human genome', *Nature*, 489(7414), pp. 57–74. Available at: <https://doi.org/10.1038/nature11247>.
- Durand, N.C. *et al.* (2016) 'Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments', *Cell Systems*, 3(1). Available at: <https://doi.org/10.1016/j.cels.2016.07.002>.
- Emberly, E. *et al.* (2008) 'BEAF regulates cell-cycle genes through the controlled deposition of H3K9 methylation marks into its conserved dual-core binding sites.', *PLoS biology*, 6(12). Available at: <https://doi.org/10.1371/journal.pbio.0060327>.
- Engreitz, J.M. *et al.* (2016) 'Local regulation of gene expression by lncRNA promoters, transcription and splicing', *Nature*, 539(7629), pp. 452–455. Available at: <https://doi.org/10.1038/nature20149>.
- Erez Lieberman-Aiden *et al.* (2009) 'Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome', *Science*, 326(5950), pp. 285–289. Available at: <https://doi.org/10.1126/science.1178746>.
- Ernst, J. and Kellis, M. (2017) 'Chromatin-state discovery and genome annotation with ChromHMM', *Nature Protocols*, 12(12), pp. 2478–2492. Available at: <https://doi.org/10.1038/nprot.2017.124>.
- Espinola, S.M. *et al.* (2021) 'Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early Drosophila development', *Nature Genetics*, 53(4), pp. 477–486. Available at: <https://doi.org/10.1038/s41588-021-00816-z>.
- Espinosa, J.M. (2017) 'On the Origin of lncRNAs: Missing Link Found', *Trends in Genetics*. Available at: <https://doi.org/10.1016/j.tig.2017.07.005>.
- Falk, S. *et al.* (2016) 'Structure of the RBM7-ZCCHC8 core of the NEXT complex reveals connections to splicing factors', *Nature Communications*, 7. Available at: <https://doi.org/10.1038/ncomms13573>.
- Falk, S. *et al.* (2017) 'Structural insights into the interaction of the nuclear exosome helicase Mtr4 with the preribosomal protein Nop53', *RNA*, 23(12). Available at: <https://doi.org/10.1261/rna.062901.117>.
- Ferrie, J.J. *et al.* (2022) "'Structure"-function relationships in eukaryotic transcription factors: The role of intrinsically disordered regions in gene regulation', *Molecular Cell*. Cell Press, pp. 3970–3984. Available at: <https://doi.org/10.1016/j.molcel.2022.09.021>.
- Filion, G.J. *et al.* (2010) 'Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells', *Cell*, 143(2), pp. 212–224. Available at: <https://doi.org/10.1016/j.cell.2010.09.009>.
- Flyamer, I.M., Illingworth, R.S. and Bickmore, W.A. (2020) 'Coolpup.py: Versatile pile-up analysis of Hi-C data', *Bioinformatics*, 36(10). Available at: <https://doi.org/10.1093/bioinformatics/btaa073>.

- Franke, M. *et al.* (2021) 'CTCF knockout in zebrafish induces alterations in regulatory landscapes and developmental gene expression', *Nature Communications*, 12(1). Available at: <https://doi.org/10.1038/s41467-021-25604-5>.
- Fukaya, T., Lim, B. and Levine, M. (2016) 'Enhancer Control of Transcriptional Bursting', *Cell*, 166(2), pp. 358–368. Available at: <https://doi.org/10.1016/j.cell.2016.05.025>.
- Fuller-Pace, F. V. and Moore, H.C. (2011) 'RNA helicases p68 and p72: Multifunctional proteins with important implications for cancer development', *Future Oncology*, pp. 239–251. Available at: <https://doi.org/10.2217/fon.11.1>.
- García-Jove Navarro, M. *et al.* (2019) 'RNA is a critical element for the sizing and the composition of phase-separated RNA–protein condensates', *Nature Communications*, 10(1). Available at: <https://doi.org/10.1038/s41467-019-11241-6>.
- García-Muse, T. and Aguilera, A. (2019) 'R Loops: From Physiological to Pathological Roles', *Cell*. Cell Press, pp. 604–618. Available at: <https://doi.org/10.1016/j.cell.2019.08.055>.
- Gaszner, M. and Felsenfeld, G. (2006) 'Insulators: Exploiting transcriptional and epigenetic mechanisms', *Nature Reviews Genetics*, pp. 703–713. Available at: <https://doi.org/10.1038/nrg1925>.
- Gerlach, P. *et al.* (2022) 'Structure and regulation of the nuclear exosome targeting complex guides RNA substrates to the exosome', *Molecular Cell*, 82(13), pp. 2505-2518.e7. Available at: <https://doi.org/10.1016/j.molcel.2022.04.011>.
- Ghavi-Helm, Y. *et al.* (2014) 'Enhancer loops appear stable during development and are associated with paused polymerase', *Nature*, 512(1), pp. 96–100. Available at: <https://doi.org/10.1038/nature13417>.
- Gil, N. and Ulitsky, I. (2018) 'Production of Spliced Long Noncoding RNAs Specifies Regions with Increased Enhancer Activity', *Cell Systems*, 7(5), pp. 537-547.e3. Available at: <https://doi.org/10.1016/j.cels.2018.10.009>.
- Giraud, G., Terrone, S. and Bourgeois, C.F. (2018) 'Functions of DEAD box RNA helicases DDX5 and DDX17 in chromatin organization and transcriptional regulation', *BMB Reports*. The Biochemical Society of the Republic of Korea, pp. 613–622. Available at: <https://doi.org/10.5483/BMBRep.2018.51.12.234>.
- Gorbovytska, V. *et al.* (2022) 'Enhancer RNAs stimulate Pol II pause release by harnessing multivalent interactions to NELF', *Nature Communications*, 13(1). Available at: <https://doi.org/10.1038/s41467-022-29934-w>.
- Gros, F. *et al.* (1961) 'Unstable ribonucleic acid revealed by pulse labelling of *Escherichia coli*', *Nature*, 190(4776). Available at: <https://doi.org/10.1038/190581a0>.
- Grote, P. *et al.* (2013) 'The Tissue-Specific lncRNA Fendrr Is an Essential Regulator of Heart and Body Wall Development in the Mouse', *Developmental Cell*, 24(2), pp. 206–214. Available at: <https://doi.org/10.1016/j.devcel.2012.12.012>.

- Guo, Y.E. *et al.* (2019) 'Pol II phosphorylation regulates a switch between transcriptional and splicing condensates', *Nature*, 572(7770), pp. 543–548. Available at: <https://doi.org/10.1038/s41586-019-1464-0>.
- Haarhuis, J.H.I. *et al.* (2017) 'The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension', *Cell*, 169(4), pp. 693–707.e14. Available at: <https://doi.org/10.1016/j.cell.2017.04.013>.
- Hamamoto, K. *et al.* (2023) 'Dynamic interplay between non-coding enhancer transcription and gene activity in development', *Nature Communications*, 14(1). Available at: <https://doi.org/10.1038/s41467-023-36485-1>.
- Hansen, A.S. *et al.* (2019) 'Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF', *Molecular Cell*, 76(3), pp. 395–411.e13. Available at: <https://doi.org/10.1016/j.molcel.2019.07.039>.
- Hao Tan, S. *et al.* (2019) 'The enhancer RNA ARIEL activates the oncogenic transcriptional program in T-cell acute lymphoblastic leukemia', *Blood* [Preprint]. Available at: <http://ashpublications.org/blood/article-pdf/134/3/239/1557714/blood874503.pdf>.
- Harris, H.L. *et al.* (2023) 'Chromatin alternates between A and B compartments at kilobase scale for subgenic organization', *Nature communications*, 14(1), p. 3303. Available at: <https://doi.org/10.1038/s41467-023-38429-1>.
- Harrison, L.J. and Bose, D. (2022) 'Enhancer RNAs step forward: New insights into enhancer function', *Development (Cambridge)*, 149(16). Available at: <https://doi.org/10.1242/dev.200398>.
- Heitz, E. (1928) 'Das Heterochromatin der Moose', *Jahrbücher für wissenschaftliche Botanik*, 69.
- Henninger, J.E. *et al.* (2021) 'RNA-Mediated Feedback Control of Transcriptional Condensates', *Cell*, 184(1), pp. 207–225.e24. Available at: <https://doi.org/10.1016/j.cell.2020.11.030>.
- Heurteau, A. *et al.* (2020) 'Insulator-based loops mediate the spreading of H3K27me3 over distant micro-domains repressing euchromatin genes', *Genome Biology*, 21(1). Available at: <https://doi.org/10.1186/s13059-020-02106-z>.
- Hildebrand, E.M. and Dekker, J. (2020) 'Mechanisms and Functions of Chromosome Compartmentalization', *Trends in Biochemical Sciences*. Elsevier Ltd, pp. 385–396. Available at: <https://doi.org/10.1016/j.tibs.2020.01.002>.
- Hiragami-Hamada, K. *et al.* (2016) 'Dynamic and flexible H3K9me3 bridging via HP1 β dimerization establishes a plastic state of condensed chromatin', *Nature Communications*, 7. Available at: <https://doi.org/10.1038/ncomms11310>.
- Hirling, H. *et al.* (1989) 'RNA helicase activity associated with the human p68 protein', *Nature*, 339(6225). Available at: <https://doi.org/10.1038/339562a0>.
- Hon, C.C. *et al.* (2017) 'An atlas of human long non-coding RNAs with accurate 5' ends', *Nature*, 543(7644), pp. 199–204. Available at: <https://doi.org/10.1038/nature21374>.

- Hondele, M. *et al.* (2019) 'DEAD-box ATPases are global regulators of phase-separated organelles', *Nature*, 573(7772), pp. 144–148. Available at: <https://doi.org/10.1038/s41586-019-1502-y>.
- Hou, T.Y. and Kraus, W.L. (2022) 'Analysis of estrogen-regulated enhancer RNAs identifies a functional motif required for enhancer assembly and gene expression', *Cell Reports*, 39(11). Available at: <https://doi.org/10.1016/j.celrep.2022.110944>.
- Hsieh, C.L. *et al.* (2014) 'Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation', *Proceedings of the National Academy of Sciences of the United States of America*, 111(20), pp. 7319–7324. Available at: <https://doi.org/10.1073/pnas.1324151111>.
- Hsieh, T.H.S. *et al.* (2015) 'Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C', *Cell*, 162(1), pp. 108–119. Available at: <https://doi.org/10.1016/j.cell.2015.05.048>.
- Hsieh, T.H.S. *et al.* (2022) 'Enhancer–promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1', *Nature Genetics*, 54(12), pp. 1919–1932. Available at: <https://doi.org/10.1038/s41588-022-01223-8>.
- Huang, Z. *et al.* (2018) 'The enhancer RNA Inc-SLC4A1-1 epigenetically regulates unexplained recurrent pregnancy loss (URPL) by activating CXCL8 and NF- κ B pathway', *EBioMedicine*, 38. Available at: <https://doi.org/10.1016/j.ebiom.2018.11.015>.
- Ing-Simmons, E. *et al.* (2021) 'Independence of chromatin conformation and gene regulation during *Drosophila* dorsoventral patterning', *Nature Genetics*, 53(4), pp. 487–499. Available at: <https://doi.org/10.1038/s41588-021-00799-x>.
- Irion, U. *et al.* (2004) 'Abstrakt, a DEAD Box Protein, Regulates Insc Levels and Asymmetric Division of Neural and Mesodermal Progenitors', *Current Biology*, 14(2), pp. 138–144. Available at: <https://doi.org/10.1016/j.cub.2004.01.002>.
- Irion, U. and Leptin, M. (1999) 'Developmental and cell biological functions of the *Drosophila* DEAD-box protein Abstrakt', *Current Biology*, 9(23). Available at: [https://doi.org/10.1016/S0960-9822\(00\)80082-2](https://doi.org/10.1016/S0960-9822(00)80082-2).
- Jarroux, J., Morillon, A. and Pinskaya, M. (2017) 'History, discovery, and classification of lncRNAs', in *Advances in Experimental Medicine and Biology*. Available at: https://doi.org/10.1007/978-981-10-5203-3_1.
- Jerkovic', I. and Cavalli, G. (2021) 'Understanding 3D genome organization by multidisciplinary methods', *Nature Reviews Molecular Cell Biology*. Nature Research, pp. 511–528. Available at: <https://doi.org/10.1038/s41580-021-00362-w>.
- Jiang, N. *et al.* (2009) 'Genome-Wide Mapping of Boundary Element-Associated Factor (BEAF) Binding Sites in *Drosophila melanogaster* Links BEAF to Transcription', *Molecular and Cellular Biology*, 29(13), pp. 3556–3568. Available at: <https://doi.org/10.1128/mcb.01748-08>.
- Jiang, Y. *et al.* (2017) 'The emerging roles of the DDX41 protein in immunity and diseases', *Protein and Cell*. Available at: <https://doi.org/10.1007/s13238-016-0303-4>.

- Jin, Y. *et al.* (2017) 'The Ground State and Evolution of Promoter Region Directionality', *Cell*, 170(5), pp. 889-898.e10. Available at: <https://doi.org/10.1016/j.cell.2017.07.006>.
- Johnson, S.J. and Jackson, R.N. (2013) 'Ski2-like RNA helicase structures: Common themes and complex assemblies', *RNA Biology*. Taylor and Francis Inc., pp. 33–43. Available at: <https://doi.org/10.4161/rna.22101>.
- Kagey, M.H. *et al.* (2010) 'Mediator and cohesin connect gene expression and chromatin architecture', *Nature*, 467(7314), pp. 430–435. Available at: <https://doi.org/10.1038/nature09380>.
- Kahn, T.G. *et al.* (2023) 'Topological screen identifies hundreds of Cp190-and CTCF-dependent Drosophila chromatin insulator elements', *Science Advances* [Preprint]. Available at: <https://doi.org/10.1126/sciadv.ade0090>.
- Kaikkonen, M.U. *et al.* (2013) 'Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription', *Molecular Cell*, 51(3), pp. 310–325. Available at: <https://doi.org/10.1016/j.molcel.2013.07.010>.
- Kaushal, A. *et al.* (2022) 'Essential role of Cp190 in physical and regulatory boundary formation', *Sci. Adv*, 8, p. 8834. Available at: <https://www.science.org>.
- Kawasaki, K. and Fukaya, T. (2023) 'Functional coordination between transcription factor clustering and gene activity', *Molecular Cell*, 83(10), pp. 1605-1622.e9. Available at: <https://doi.org/10.1016/j.molcel.2023.04.018>.
- Kellum', R. and Schedl, P. (1991) *A Position-Effect Assay for Boundaries of Higher Order Chromosomal Domains*, *Cell*.
- El Khattabi, L. *et al.* (2019) 'A Pliable Mediator Acts as a Functional Rather Than an Architectural Bridge between Promoters and Enhancers', *Cell*, 178(5), pp. 1145-1158.e20. Available at: <https://doi.org/10.1016/j.cell.2019.07.011>.
- Kilchert, C., Wittmann, S. and Vasiljeva, L. (2016) 'The regulation and functions of the nuclear RNA exosome complex', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, pp. 227–239. Available at: <https://doi.org/10.1038/nrm.2015.15>.
- Kim, J.J. and Kingston, R.E. (2022) 'Context-specific Polycomb mechanisms in development', *Nature Reviews Genetics*. Available at: <https://doi.org/10.1038/s41576-022-00499-0>.
- Kim, T.K. *et al.* (2010) 'Widespread transcription at neuronal activity-regulated enhancers', *Nature*, 465(7295), pp. 182–187. Available at: <https://doi.org/10.1038/nature09033>.
- Kim, Y.J. *et al.* (2018) 'Global transcriptional activity dynamics reveal functional enhancer RNAs', *Genome Research*, 28(12), pp. 1799–1811. Available at: <https://doi.org/10.1101/gr.233486.117>.
- Kopylova, E., Noé, L. and Touzet, H. (2012) 'SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data', *Bioinformatics*, 28(24). Available at: <https://doi.org/10.1093/bioinformatics/bts611>.

- Kraft, K. *et al.* (2022) 'Polycomb-mediated genome architecture enables long-range spreading of H3K27 methylation', *PNAS* [Preprint]. Available at: <https://doi.org/10.1073/pnas>.
- Kurz, A. *et al.* (1996) 'Active and Inactive Genes Localize Preferentially in the Periphery of Chromosome Territories', *The Journal of Cell Biology* [Preprint]. Available at: <http://rupress.org/jcb/article-pdf/135/5/1195/1266105/1195.pdf>.
- Lai, F. *et al.* (2013) 'Activating RNAs associate with Mediator to enhance chromatin architecture and transcription', *Nature*, 494(7438), pp. 497–501. Available at: <https://doi.org/10.1038/nature11884>.
- Lai, F. *et al.* (2015) 'Integrator mediates the biogenesis of enhancer RNAs', *Nature*, 525(7569), pp. 399–403. Available at: <https://doi.org/10.1038/nature14906>.
- Lam, M.T.Y. *et al.* (2013) 'Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription', *Nature*, 498(7455), pp. 511–515. Available at: <https://doi.org/10.1038/nature12209>.
- Lee, J.H. *et al.* (2021) 'Enhancer RNA m6A methylation facilitates transcriptional condensate formation and gene activation', *Molecular Cell*, 81(16), pp. 3368–3385.e9. Available at: <https://doi.org/10.1016/j.molcel.2021.07.024>.
- Lee, Y.J., Wang, Q. and Rio, D.C. (2018) 'Coordinate regulation of alternative pre-mRNA splicing events by the human RNA chaperone proteins hnRNPA1 and DDX5', *Genes and Development*, 32(15–16). Available at: <https://doi.org/10.1101/gad.316034.118>.
- Lei, E.P. and Corces, V.G. (2006) 'RNA interference machinery influences the nuclear organization of a chromatin insulator', *Nature Genetics*, 38(8), pp. 936–941. Available at: <https://doi.org/10.1038/ng1850>.
- Levo, M. *et al.* (2022) 'Transcriptional coupling of distant regulatory genes in living embryos', *Nature*, 605(7911), pp. 754–760. Available at: <https://doi.org/10.1038/s41586-022-04680-7>.
- Lhoumaud, P. *et al.* (2014) 'Insulators recruit histone methyltransferase dMes4 to regulate chromatin of flanking genes', *The EMBO Journal*, 33(14), pp. 1599–1613. Available at: <https://doi.org/10.15252/embj.201385965>.
- Li, L. *et al.* (2015) 'Widespread Rearrangement of 3D Chromatin Organization Underlies Polycomb-Mediated Stress-Induced Silencing', *Molecular Cell*, 58(2), pp. 216–231. Available at: <https://doi.org/10.1016/j.molcel.2015.02.023>.
- Li, W. *et al.* (2013) 'Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation', *Nature*, 498(7455), pp. 516–520. Available at: <https://doi.org/10.1038/nature12210>.
- Li, X. *et al.* (2023) 'GAGA-associated factor fosters loop formation in the Drosophila genome.', *Molecular cell* [Preprint]. Available at: <https://doi.org/10.1016/j.molcel.2023.03.011>.

- Li, Z. *et al.* (2014) 'The long noncoding RNA THRIL regulates TNF α expression through its interaction with hnRNPL', *Proceedings of the National Academy of Sciences of the United States of America*, 111(3), pp. 1002–1007. Available at: <https://doi.org/10.1073/pnas.1313768111>.
- Liang, J. *et al.* (2014) 'Chromatin Immunoprecipitation Indirect Peaks Highlight Long-Range Interactions of Insulator Proteins and Pol II Pausing', *Molecular Cell*, 53(4), pp. 672–681. Available at: <https://doi.org/10.1016/j.molcel.2013.12.029>.
- Liao, Y., Smyth, G.K. and Shi, W. (2019) 'The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads', *Nucleic Acids Research*, 47(8). Available at: <https://doi.org/10.1093/nar/gkz114>.
- Lim, S.J. *et al.* (2013) 'Genome-wide localization of exosome components to active promoters and chromatin insulators in Drosophila', *Nucleic Acids Research*, 41(5), pp. 2963–2980. Available at: <https://doi.org/10.1093/nar/gkt037>.
- Linder, P. and Jankowsky, E. (2011) 'From unwinding to clamping - the DEAD box RNA helicase family', *Nature Reviews Molecular Cell Biology*, pp. 505–516. Available at: <https://doi.org/10.1038/nrm3154>.
- Lingaraju, M. *et al.* (2019) 'The MTR4 helicase recruits nuclear adaptors of the human RNA exosome using distinct arch-interacting motifs', *Nature Communications*, 10(1). Available at: <https://doi.org/10.1038/s41467-019-11339-x>.
- Liu, J. *et al.* (2020) 'N6-methyladenosine of chromosome-associated regulatory RNA regulates chromatin state and transcription', *Science* [Preprint]. Available at: <https://www.science.org>.
- Liu, Y. and Imai, R. (2018) 'Function of plant DExD/H-Box RNA helicases associated with ribosomal RNA biogenesis', *Frontiers in Plant Science*. Available at: <https://doi.org/10.3389/fpls.2018.00125>.
- Loda, A. and Heard, E. (2019) 'Xist RNA in action: Past, present, and future', *PLoS Genetics*. Available at: <https://doi.org/10.1371/journal.pgen.1008333>.
- Long, Y. *et al.* (2020) 'RNA is essential for PRC2 chromatin occupancy and function in human pluripotent stem cells', *Nature Genetics*, 52(9), pp. 931–938. Available at: <https://doi.org/10.1038/s41588-020-0662-x>.
- Love, M.I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12). Available at: <https://doi.org/10.1186/s13059-014-0550-8>.
- Lubas, M. *et al.* (2011) 'Interaction Profiling Identifies the Human Nuclear Exosome Targeting Complex', *Molecular Cell*, 43(4), pp. 624–637. Available at: <https://doi.org/10.1016/j.molcel.2011.06.028>.
- Lubas, M. *et al.* (2015) 'The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis', *Cell Reports*, 10(2), pp. 178–192. Available at: <https://doi.org/10.1016/j.celrep.2014.12.026>.

- Lun, A.T.L. and Smyth, G.K. (2015) 'Cis: A Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows', *Nucleic Acids Research*, 44(5). Available at: <https://doi.org/10.1093/nar/gkv1191>.
- Luo, H. *et al.* (2019) 'HOTTIP lncRNA Promotes Hematopoietic Stem Cell Self-Renewal Leading to AML-like Disease in Mice', *Cancer Cell*, 36(6), pp. 645-659.e8. Available at: <https://doi.org/10.1016/j.ccell.2019.10.011>.
- Luo, H. *et al.* (2022) 'HOTTIP-dependent R-loop formation regulates CTCF boundary activity and TAD integrity in leukemia', *Molecular Cell*, 82(4), pp. 833-851.e11. Available at: <https://doi.org/10.1016/j.molcel.2022.01.014>.
- Ma, J. *et al.* (2022) 'DDX41 is needed for pre- and postnatal hematopoietic stem cell differentiation in mice', *Stem Cell Reports*, 17(4). Available at: <https://doi.org/10.1016/j.stemcr.2022.02.010>.
- Mach, P. *et al.* (2022) 'Cohesin and CTCF control the dynamics of chromosome folding', *Nature Genetics*, 54(12), pp. 1907–1918. Available at: <https://doi.org/10.1038/s41588-022-01232-7>.
- Machida, S. *et al.* (2018) 'Structural Basis of Heterochromatin Formation by Human HP1', *Molecular Cell*, 69(3), pp. 385-397.e8. Available at: <https://doi.org/10.1016/j.molcel.2017.12.011>.
- Maharana, S. *et al.* (2018) 'RNA buffers the phase separation behavior of prion-like RNA binding proteins', *Science*, 360(6391). Available at: <https://doi.org/10.1126/science.aar7366>.
- Maksimenko, O. *et al.* (2015) 'Two new insulator proteins, Pita and ZIPIC, target CP190 to chromatin', *Genome Research*, 25(1), pp. 89–99. Available at: <https://doi.org/10.1101/gr.174169.114>.
- Mangiavacchi, A., Morelli, G. and Orlando, V. (2023) 'Behind the scenes: How RNA orchestrates the epigenetic regulation of gene expression', *Frontiers in Cell and Developmental Biology*. Frontiers Media S.A. Available at: <https://doi.org/10.3389/fcell.2023.1123975>.
- Maruyama, A., Mimura, J. and Itoh, K. (2014) 'Non-coding RNA derived from the region adjacent to the human HO-1 E2 enhancer selectively regulates HO-1 gene induction by modulating Pol II binding', *Nucleic Acids Research*, 42(22), pp. 13599–13614. Available at: <https://doi.org/10.1093/nar/gku1169>.
- Matzat, L.H., Dale, R.K. and Lei, E.P. (2013) 'Messenger RNA is a functional component of a chromatin insulator complex', *EMBO Reports*, 14(10), pp. 916–922. Available at: <https://doi.org/10.1038/embor.2013.118>.
- Melo, C.A. *et al.* (2013) 'lncRNAs Are Required for p53-Dependent Enhancer Activity and Gene Transcription', *Molecular Cell*, 49(3), pp. 524–535. Available at: <https://doi.org/10.1016/j.molcel.2012.11.021>.
- Meola, N. *et al.* (2016) 'Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts', *Molecular Cell*, 64(3), pp. 520–533. Available at: <https://doi.org/10.1016/j.molcel.2016.09.025>.

- Mersaoui, S.Y. *et al.* (2019) 'Arginine methylation of the DDX 5 helicase RGG / RG motif by PRMT 5 regulates resolution of RNA:DNA hybrids', *The EMBO Journal*, 38(15). Available at: <https://doi.org/10.15252/embj.2018100986>.
- Mikhaylichenko, O. *et al.* (2018) 'The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription', *Genes and Development*, 32(1), pp. 42–57. Available at: <https://doi.org/10.1101/gad.308619.117>.
- Mohana, G. *et al.* (2023) 'Chromosome-level organization of the regulatory genome in the Drosophila nervous system', *Cell* [Preprint]. Available at: <https://doi.org/10.1016/j.cell.2023.07.008>.
- Moon, H. *et al.* (2005) 'CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator', *EMBO Reports*, 6(2), pp. 165–170. Available at: <https://doi.org/10.1038/sj.embor.7400334>.
- Morgan, T.H. (1910) 'SEX LIMITED INHERITANCE IN DROSOPHILA', *Science* [Preprint]. Available at: <http://www.esp.org>.
- Mosler, T. *et al.* (2021) 'R-loop proximity proteomics identifies a role of DDX41 in transcription-associated genomic instability', *Nature Communications*, 12(1). Available at: <https://doi.org/10.1038/s41467-021-27530-y>.
- Mousavi, K. *et al.* (2013) 'ERNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci', *Molecular Cell*, 51(5), pp. 606–617. Available at: <https://doi.org/10.1016/j.molcel.2013.07.022>.
- Muller, H.J. (1930) 'Types of visible variations induced by X-rays in Drosophila', *Journal of Genetics*, 22(3). Available at: <https://doi.org/10.1007/BF02984195>.
- Mumbach, M.R. *et al.* (2016) 'HiChIP: Efficient and sensitive analysis of protein-directed genome architecture', *Nature Methods*, 13(11), pp. 919–922. Available at: <https://doi.org/10.1038/nmeth.3999>.
- Nagano, T. *et al.* (2013) 'Single-cell Hi-C reveals cell-to-cell variability in chromosome structure', *Nature*, 502(7469), pp. 59–64. Available at: <https://doi.org/10.1038/nature12593>.
- Nagano, T. *et al.* (2017) 'Cell-cycle dynamics of chromosomal organization at single-cell resolution', *Nature*, 547(7661), pp. 61–67. Available at: <https://doi.org/10.1038/nature23001>.
- Nègre, N. *et al.* (2010) 'A comprehensive map of insulator elements for the Drosophila genome', *PLoS Genetics*, 6(1). Available at: <https://doi.org/10.1371/journal.pgen.1000814>.
- Nguyen, H.Q. *et al.* (2020) '3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing', *Nature Methods*, 17(8), pp. 822–832. Available at: <https://doi.org/10.1038/s41592-020-0890-0>.

- Nojima, T. *et al.* (2018) 'Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence', *Molecular Cell*, 72(6), pp. 970-984.e7. Available at: <https://doi.org/10.1016/j.molcel.2018.10.011>.
- Ntini, E. *et al.* (2013) 'Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality', *Nature Structural and Molecular Biology*, 20(8). Available at: <https://doi.org/10.1038/nsmb.2640>.
- Oh, H.J. *et al.* (2021) 'Jpx RNA regulates CTCF anchor site selection and formation of chromosome loops', *Cell*, 184(25), pp. 6157-6173.e24. Available at: <https://doi.org/10.1016/j.cell.2021.11.012>.
- Ohno, S. (1972) 'So much "junk" DNA in our genome.', *Brookhaven symposia in biology*, 23.
- Ohtsuki, S. and Levine, M. (1998) 'GAGA mediates the enhancer blocking activity of the eve promoter in the Drosophila embryo', *Genes & Development* [Preprint]. Available at: www.genesdev.org.
- Olsen, K.J. and Johnson, S.J. (2021) 'Mtr4 RNA helicase structures and interactions', *Biological Chemistry*. De Gruyter Open Ltd, pp. 605–616. Available at: <https://doi.org/10.1515/hsz-2020-0329>.
- Ong, C.T. *et al.* (2013) 'Poly(ADP-ribosyl)ation regulates insulator function and intrachromosomal interactions in drosophila', *Cell*, 155(1), pp. 148–159. Available at: <https://doi.org/10.1016/j.cell.2013.08.052>.
- Ørom, U.A. *et al.* (2010) 'Long noncoding RNAs with enhancer-like function in human cells', *Cell*, 143(1), pp. 46–58. Available at: <https://doi.org/10.1016/j.cell.2010.09.001>.
- Pan, H. *et al.* (2021) 'Cohesin SA1 and SA2 are RNA binding proteins that localize to RNA containing regions on DNA', *Nucleic Acids Research*, 48(10), pp. 5639–5655. Available at: <https://doi.org/10.1093/NAR/GKAA284>.
- Pascal Preker *et al.* (2008) 'Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters', *Science*, 322(5909), pp. 1845–1848. Available at: <https://doi.org/10.1126/science.1162228>.
- Patrick, E.M. *et al.* (2017) 'The RNA helicase Mtr4p is a duplex-sensing translocase', *Nature Chemical Biology*, 13(1), pp. 99–104. Available at: <https://doi.org/10.1038/nchembio.2234>.
- Pefanis, E. *et al.* (2015) 'RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity', *Cell*, 161(4), pp. 774–789. Available at: <https://doi.org/10.1016/j.cell.2015.04.034>.
- Pherson, M. *et al.* (2019) 'Cohesin occupancy and composition at enhancers and promoters are linked to DNA replication origin proximity in Drosophila', *Genome Research*, 29(4), pp. 602–612. Available at: <https://doi.org/10.1101/gr.243832.118>.
- Pnueli, L. *et al.* (2015) 'RNA transcribed from a distal enhancer is required for activating the chromatin at the promoter of the gonadotropin α -subunit gene', *Proceedings of the National Academy of Sciences of the United States of America*, 112(14), pp. 4369–4374. Available at: <https://doi.org/10.1073/pnas.1414841112>.

- Polprasert, C. *et al.* (2015) 'Inherited and Somatic Defects in DDX41 in Myeloid Neoplasms', *Cancer Cell*, 27(5). Available at: <https://doi.org/10.1016/j.ccell.2015.03.017>.
- Przanowska, R.K. *et al.* (2022) 'Distinct MUNC lncRNA structural domains regulate transcription of different promyogenic factors', *Cell Reports*, 38(7). Available at: <https://doi.org/10.1016/j.celrep.2022.110361>.
- Putri, G.H. *et al.* (2022) 'Analysing high-throughput sequencing data in Python with HTSeq 2.0', *Bioinformatics*, 38(10). Available at: <https://doi.org/10.1093/bioinformatics/btac166>.
- Quinlan, A.R. and Hall, I.M. (2010) 'BEDTools: A flexible suite of utilities for comparing genomic features', *Bioinformatics*, 26(6). Available at: <https://doi.org/10.1093/bioinformatics/btq033>.
- Quinodoz, S.A. *et al.* (2018) 'Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus', *Cell*, 174(3), pp. 744-757.e24. Available at: <https://doi.org/10.1016/j.cell.2018.05.024>.
- Quinodoz, S.A. *et al.* (2021) 'RNA promotes the formation of spatial compartments in the nucleus', *Cell*, 184(23), pp. 5775-5790.e30. Available at: <https://doi.org/10.1016/j.cell.2021.10.014>.
- Rahnamoun, H. *et al.* (2018) 'RNAs interact with BRD4 to promote enhanced chromatin engagement and transcription activation', *Nature Structural and Molecular Biology*, 25(8), pp. 687–697. Available at: <https://doi.org/10.1038/s41594-018-0102-0>.
- Ramasamy, S. *et al.* (2023) 'The Mediator complex regulates enhancer-promoter interactions', *Nature Structural & Molecular Biology*, 30(7), pp. 991–1000. Available at: <https://doi.org/10.1038/s41594-023-01027-2>.
- Ramírez, F. *et al.* (2018) 'High-resolution TADs reveal DNA sequences underlying genome organization in flies', *Nature Communications*, 9(1). Available at: <https://doi.org/10.1038/s41467-017-02525-w>.
- Rao, S.S.P. *et al.* (2014) 'A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping', *Cell*, 159(7), pp. 1665–1680. Available at: <https://doi.org/10.1016/j.cell.2014.11.021>.
- Rao, S.S.P. *et al.* (2017) 'Cohesin Loss Eliminates All Loop Domains', *Cell*, 171(2), pp. 305-320.e24. Available at: <https://doi.org/10.1016/j.cell.2017.09.026>.
- Ren, G. *et al.* (2017) 'CTCF-Mediated Enhancer-Promoter Interaction Is a Critical Regulator of Cell-to-Cell Variation of Gene Expression', *Molecular Cell*, 67(6), pp. 1049-1058.e6. Available at: <https://doi.org/10.1016/j.molcel.2017.08.026>.
- Rennie, S. *et al.* (2018) 'Transcription start site analysis reveals widespread divergent transcription in *D. melanogaster* and core promoter-encoded enhancer activities', *Nucleic Acids Research*, 46(11), pp. 5455–5469. Available at: <https://doi.org/10.1093/nar/gky244>.
- Rinn, J.L. *et al.* (2007) 'Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs', *Cell*, 129(7). Available at: <https://doi.org/10.1016/j.cell.2007.05.022>.

- Roden, C. and Gladfelter, A.S. (2021) 'RNA contributions to the form and function of biomolecular condensates', *Nature Reviews Molecular Cell Biology*. Nature Research, pp. 183–195. Available at: <https://doi.org/10.1038/s41580-020-0264-6>.
- Rollins, R.A., Morcillo, P. and Dorsett, D. (1999) 'Nipped-B, a Drosophila Homologue of Chromosomal Adherins, Participates in Activation by Remote Enhancers in the cut and Ultrabithorax Genes', *Genetics* [Preprint]. Available at: <https://academic.oup.com/genetics/article/152/2/577/6034922>.
- Roseman, R.R., Pirrotta, V. and Geyer, P.K. (1993) 'The su(Hw) protein insulates expression of the Drosophila melanogaster white gene from chromosomal position-effects', *EMBO Journal*, 12(2), pp. 435–442. Available at: <https://doi.org/10.1002/j.1460-2075.1993.tb05675.x>.
- Rössler, O.G., Straka, A. and Stahl, H. (2001) 'Rearrangement of structured RNA via branch migration structures catalysed by the highly related DEAD-box proteins p68 and p72', *Nucleic Acids Research*, 29(10). Available at: <https://doi.org/10.1093/nar/29.10.2088>.
- Rosow, K.L. and Janknecht, R. (2003) 'Synergism between p68 RNA helicase and the transcriptional coactivators CBP and p300', *Oncogene*, 22(1). Available at: <https://doi.org/10.1038/sj.onc.1206067>.
- Rowley, M.J. *et al.* (2017) 'Evolutionarily Conserved Principles Predict 3D Chromatin Organization', *Molecular Cell*, 67(5), pp. 837-852.e7. Available at: <https://doi.org/10.1016/j.molcel.2017.07.022>.
- Rowley, M.J. *et al.* (2019) 'Condensin II Counteracts Cohesin and RNA Polymerase II in the Establishment of 3D Chromatin Organization', *Cell Reports*, 26(11), pp. 2890-2903.e3. Available at: <https://doi.org/10.1016/j.celrep.2019.01.116>.
- Rudolph, M.G. and Klostermeier, D. (2015) 'When core competence is not enough: Functional interplay of the DEAD-box helicase core with ancillary domains and auxiliary factors in RNA binding and unwinding', *Biological Chemistry*. Available at: <https://doi.org/10.1515/hsz-2014-0277>.
- Sabari, B.R. *et al.* (2018) 'Coactivator condensation at super-enhancers links phase separation and gene control', *Science*, 361(6400). Available at: <https://doi.org/10.1126/science.aar3958>.
- Saksouk, N., Simboeck, E. and Déjardin, J. (2015) 'Constitutive heterochromatin formation and transcription in mammals', *Epigenetics and Chromatin*. Available at: <https://doi.org/10.1186/1756-8935-8-3>.
- Saldaña-Meyer, R. *et al.* (2014) 'CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53', *Genes and Development*, 28(7), pp. 723–734. Available at: <https://doi.org/10.1101/gad.236869.113>.
- Saldaña-Meyer, R. *et al.* (2019) 'RNA Interactions Are Essential for CTCF-Mediated Genome Organization', *Molecular Cell*, 76(3), pp. 412-422.e5. Available at: <https://doi.org/10.1016/j.molcel.2019.08.015>.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) 'DNA sequencing with chain-terminating inhibitors.', *Proceedings of the National Academy of Sciences of the United States of America*, 74(12). Available at: <https://doi.org/10.1073/pnas.74.12.5463>.

- de Santa, F. *et al.* (2010) 'A large fraction of extragenic RNA Pol II transcription sites overlap enhancers', *PLoS Biology*, 8(5). Available at: <https://doi.org/10.1371/journal.pbio.1000384>.
- Schaaf, C.A. *et al.* (2013) 'Genome-Wide Control of RNA Polymerase II Activity by Cohesin', *PLoS Genetics*, 9(3). Available at: <https://doi.org/10.1371/journal.pgen.1003382>.
- Schaukowitch, K. *et al.* (2014) 'Enhancer RNA facilitates NELF release from immediate early genes', *Molecular Cell*, 56(1), pp. 29–42. Available at: <https://doi.org/10.1016/j.molcel.2014.08.023>.
- Schindelin, J. *et al.* (2012) 'Fiji: An open-source platform for biological-image analysis', *Nature Methods*. Available at: <https://doi.org/10.1038/nmeth.2019>.
- Schmid, M. and Jensen, T.H. (2019) 'The Nuclear RNA Exosome and Its Cofactors', in *Advances in Experimental Medicine and Biology*. Springer, pp. 113–132. Available at: https://doi.org/10.1007/978-3-030-31434-7_4.
- Schmidt, D. *et al.* (2010) 'A CTCF-independent role for cohesin in tissue-specific transcription', *Genome Research*, 20(5), pp. 578–588. Available at: <https://doi.org/10.1101/gr.100479.109>.
- Schmucker, D. *et al.* (1999) 'The Drosophila gene *abstrakt*, required for visual system development, encodes a putative RNA helicase of the DEAD box protein family', *Mechanisms of Development* [Preprint]. Available at: www.elsevier.com/locate/modo.
- Schoenfelder, S. and Fraser, P. (2019) 'Long-range enhancer–promoter contacts in gene expression control', *Nature Reviews Genetics*. Nature Research, pp. 437–455. Available at: <https://doi.org/10.1038/s41576-019-0128-0>.
- Schubert, T. *et al.* (2012) 'Df31 Protein and snoRNAs Maintain Accessible Higher-Order Structures of Chromatin', *Molecular Cell*, 48(3), pp. 434–444. Available at: <https://doi.org/10.1016/j.molcel.2012.08.021>.
- Schultz, J. (1936) 'Variegation in Drosophila and the Inert Chromosome Regions', *Proceedings of the National Academy of Sciences*, 22(1). Available at: <https://doi.org/10.1073/pnas.22.1.27>.
- Schwalb, B. *et al.* (2016) 'TT-seq maps the human transient transcriptome', *Science*, 352(6290). Available at: <https://doi.org/10.1126/science.aad9841>.
- Schwarzer, W. *et al.* (2017) 'Two independent modes of chromatin organization revealed by cohesin removal', *Nature*, 551(7678), pp. 51–56. Available at: <https://doi.org/10.1038/nature24281>.
- Sengoku, T. *et al.* (2006) 'Structural Basis for RNA Unwinding by the DEAD-Box Protein Drosophila Vasa', *Cell*, 125(2), pp. 287–300. Available at: <https://doi.org/10.1016/j.cell.2006.01.054>.
- Sexton, T. *et al.* (2012) 'Three-dimensional folding and functional organization principles of the Drosophila genome', *Cell*, 148(3), pp. 458–472. Available at: <https://doi.org/10.1016/j.cell.2012.01.010>.

- Sharp, P.A. *et al.* (2022) 'RNA in formation and regulation of transcriptional condensates', *RNA* [Preprint]. Available at: <https://doi.org/10.1261/rna>.
- Shechner, D.M. *et al.* (2015) 'Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display', *Nature Methods*, 12(7). Available at: <https://doi.org/10.1038/nmeth.3433>.
- Shii, L. *et al.* (2017) 'SERPINB2 is regulated by dynamic interactions with pause-release proteins and enhancer RNAs', *Molecular Immunology*, 88, pp. 20–31. Available at: <https://doi.org/10.1016/j.molimm.2017.05.005>.
- Shrinivas, K. *et al.* (2019) 'Enhancer Features that Drive Formation of Transcriptional Condensates', *Molecular Cell*, 75(3), pp. 549–561.e7. Available at: <https://doi.org/10.1016/j.molcel.2019.07.009>.
- Sigova, A.A. *et al.* (2015) 'Transcription factor trapping by RNA in gene regulatory elements', *Science*, 350(6263), pp. 978–991. Available at: <https://doi.org/10.1126/science.aad3346>.
- Skourti-Stathaki, K. and Proudfoot, N.J. (2014) 'A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression', *Genes and Development*. Cold Spring Harbor Laboratory Press, pp. 1384–1396. Available at: <https://doi.org/10.1101/gad.242990.114>.
- Spurlock, C.F. *et al.* (2017) 'Profiles of Long Noncoding RNAs in Human Naive and Memory T Cells', *The Journal of Immunology*, 199(2). Available at: <https://doi.org/10.4049/jimmunol.1700232>.
- Srivastava, D. and DeWitt, N. (2016) 'In Vivo Cellular Reprogramming: The Next Generation', *Cell*. Available at: <https://doi.org/10.1016/j.cell.2016.08.055>.
- Stack, S.M., Brown, D.B. and Deweyf, W.C. (1977) *VISUALIZATION OF INTERPHASE CHROMOSOMES*, *J. Cell Set.*
- Stempor, P. and Ahringer, J. (2016) 'SeqPlots - Interactive software for exploratory data analyses, pattern discovery and visualization in genomics', *Wellcome Open Research*, 1. Available at: <https://doi.org/10.12688/wellcomeopenres.10004.1>.
- St.Laurent, G., Wahlestedt, C. and Kapranov, P. (2015) 'The Landscape of long noncoding RNA classification', *Trends in Genetics*. Available at: <https://doi.org/10.1016/j.tig.2015.03.007>.
- Stringer, C. *et al.* (2021) 'Cellpose: a generalist algorithm for cellular segmentation', *Nature Methods*, 18(1). Available at: <https://doi.org/10.1038/s41592-020-01018-x>.
- Strom, A.R. *et al.* (2017) 'Phase separation drives heterochromatin domain formation', *Nature*, 547(7662), pp. 241–245. Available at: <https://doi.org/10.1038/nature22989>.
- Sun, S. *et al.* (2013) 'Xjpx RNA activates xist by evicting CTCF', *Cell*, 153(7), p. 1537. Available at: <https://doi.org/10.1016/j.cell.2013.05.028>.
- Suzuki, H.I. *et al.* (2009) 'Modulation of microRNA processing by p53', *Nature*, 460(7254). Available at: <https://doi.org/10.1038/nature08199>.

- Syed, K.M. and Hon, C.C. (2021) 'Heterogeneity among enhancer RNAs: Origins, consequences and perspectives', *Essays in Biochemistry*. Portland Press Ltd, pp. 709–721. Available at: <https://doi.org/10.1042/EBC20200064>.
- Szabo, Q. *et al.* (2020) 'Regulation of single-cell genome organization into TADs and chromatin nanodomains', *Nature Genetics*, 52(11), pp. 1151–1157. Available at: <https://doi.org/10.1038/s41588-020-00716-8>.
- Szabo, Q., Bantignies, F. and Cavalli, G. (2019) 'Principles of genome folding into topologically associating domains', *Science Advances* [Preprint]. Available at: <http://advances.sciencemag.org/>.
- Takahashi, K. and Yamanaka, S. (2006) 'Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors', *Cell*, 126(4), pp. 663–676. Available at: <https://doi.org/10.1016/j.cell.2006.07.024>.
- Tan-Wong, S.M., Dhir, S. and Proudfoot, N.J. (2019) 'R-Loops Promote Antisense Transcription across the Mammalian Genome', *Molecular Cell*, 76(4), pp. 600-616.e6. Available at: <https://doi.org/10.1016/j.molcel.2019.10.002>.
- Tatomer, D.C. *et al.* (2019) 'The Integrator complex cleaves nascent mRNAs to attenuate transcription', *Genes & development*, 33(21–22), pp. 1525–1538. Available at: <https://doi.org/10.1101/gad.330167.119>.
- Teo, W.W. *et al.* (2022) 'Non-coding RNA LEVER sequestration of PRC2 can mediate long range gene regulation', *Communications Biology*, 5(1). Available at: <https://doi.org/10.1038/s42003-022-03250-x>.
- Terrone, S. *et al.* (2022) 'RNA helicase-dependent gene looping impacts messenger RNA processing', *Nucleic Acids Research*, 50(16), pp. 9226–9246. Available at: <https://doi.org/10.1093/nar/gkac717>.
- Thoms, M. *et al.* (2015) 'The Exosome Is Recruited to RNA Substrates through Specific Adaptor Proteins', *Cell*, 162(5), pp. 1029–1038. Available at: <https://doi.org/10.1016/j.cell.2015.07.060>.
- Tsai, P.F. *et al.* (2018) 'A Muscle-Specific Enhancer RNA Mediates Cohesin Recruitment and Regulates Transcription In trans', *Molecular Cell*, 71(1), pp. 129-141.e8. Available at: <https://doi.org/10.1016/j.molcel.2018.06.008>.
- Tsukamoto, T. *et al.* (2020) 'Insights into the Involvement of Spliceosomal Mutations in Myelodysplastic Disorders from Analysis of SACY-1/DDX41 in *Caenorhabditis elegans*', *Genetics*, 214(4). Available at: <https://doi.org/10.1534/GENETICS.119.302973>.
- Tunnacliffe, E. and Chubb, J.R. (2020) 'What Is a Transcriptional Burst?', *Trends in Genetics*. Elsevier Ltd, pp. 288–297. Available at: <https://doi.org/10.1016/j.tig.2020.01.003>.
- Venkatesh, S. and Workman, J.L. (2015) 'Histone exchange, chromatin structure and the regulation of transcription', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, pp. 178–189. Available at: <https://doi.org/10.1038/nrm3941>.

- Vian, L. *et al.* (2018) 'The Energetics and Physiological Impact of Cohesin Extrusion', *Cell*, 173(5), pp. 1165-1178.e20. Available at: <https://doi.org/10.1016/j.cell.2018.03.072>.
- Vogelmann, J. *et al.* (2014) 'Chromatin Insulator Factors Involved in Long-Range DNA Interactions and Their Role in the Folding of the Drosophila Genome', *PLoS Genetics*, 10(8). Available at: <https://doi.org/10.1371/journal.pgen.1004544>.
- Waddington, C. (1942) 'Canalization of Development and the Inheritance of Acquired Characters', *Nature* [Preprint].
- Waddington, C. (1953) 'EXPERIMENTS IN ACQUIRED CHARACTERISTICS', *Scientific American*, 189(6), pp. 92–99. Available at: <https://doi.org/10.2307/24944428>.
- Wang, J. *et al.* (2019) 'NRDE2 negatively regulates exosome functions by inhibiting MTR4 recruitment and exosome interaction', *Genes and Development*, 33(9–10). Available at: <https://doi.org/10.1101/gad.322602.118>.
- Wang, Yunzhe *et al.* (2022) 'Enhancer RNA (eRNA) in Human Diseases', *International Journal of Molecular Sciences*. Available at: <https://doi.org/10.3390/ijms231911582>.
- Warner, D.R. *et al.* (2004) 'Functional interaction between Smad, CREB binding protein, and p68 RNA helicase', *Biochemical and Biophysical Research Communications*, 324(1). Available at: <https://doi.org/10.1016/j.bbrc.2004.09.017>.
- Weber, C.M., Ramachandran, S. and Henikoff, S. (2014) 'Nucleosomes are context-specific, H2A.Z-Modulated barriers to RNA polymerase', *Molecular Cell*, 53(5), pp. 819–830. Available at: <https://doi.org/10.1016/j.molcel.2014.02.014>.
- Weinreb, J.T. *et al.* (2021) 'Excessive R-loops trigger an inflammatory cascade leading to increased HSPC production', *Developmental Cell*, 56(5), pp. 627-640.e5. Available at: <https://doi.org/10.1016/j.devcel.2021.02.006>.
- Weinreb, J.T. *et al.* (2022) 'Ddx41 inhibition of DNA damage signaling permits erythroid progenitor expansion in zebrafish', *Haematologica*, 107(3). Available at: <https://doi.org/10.3324/haematol.2020.257246>.
- Weintraub, A.S. *et al.* (2017) 'YY1 Is a Structural Regulator of Enhancer-Promoter Loops', *Cell*, 171(7), pp. 1573-1588.e28. Available at: <https://doi.org/10.1016/j.cell.2017.11.008>.
- Wilson, B.J. *et al.* (2004) 'The p68 and p72 DEAD box RNA helicases interact with HDAC1 and repress transcription in a promoter-specific manner', *BMC Molecular Biology*, 5. Available at: <https://doi.org/10.1186/1471-2199-5-11>.
- Winczura, K. *et al.* (2018) 'Characterizing ZC3H18, a Multi-domain Protein at the Interface of RNA Production and Destruction Decisions', *Cell Reports*, 22(1), pp. 44–58. Available at: <https://doi.org/10.1016/j.celrep.2017.12.037>.

- Wu, G. *et al.* (2019) 'DDX5 helicase resolves G-quadruplex and is involved in MYC gene transcriptional activation', *Proceedings of the National Academy of Sciences of the United States of America*, 116(41), pp. 20453–20461. Available at: <https://doi.org/10.1073/pnas.1909047116>.
- Wu, T. *et al.* (2021) 'clusterProfiler 4.0: A universal enrichment tool for interpreting omics data', *Innovation*, 2(3). Available at: <https://doi.org/10.1016/j.xinn.2021.100141>.
- Wulfridge, P. *et al.* (2023) 'G-quadruplexes associated with R-loops promote CTCF binding.', *Molecular cell* [Preprint]. Available at: <https://doi.org/10.1016/j.molcel.2023.07.009>.
- Wutz, G. *et al.* (2017) 'Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins', *The EMBO Journal*, 36(24), pp. 3573–3599. Available at: <https://doi.org/10.15252/embj.201798004>.
- Xie, L. *et al.* (2022) 'BRD2 compartmentalizes the accessible genome', *Nature Genetics*, 54(4), pp. 481–491. Available at: <https://doi.org/10.1038/s41588-022-01044-9>.
- Xing, Z., Ma, W.K. and Tran, E.J. (2019) 'The DDX5/Dbp2 subfamily of DEAD-box RNA helicases', *Wiley Interdisciplinary Reviews: RNA*. Blackwell Publishing Ltd. Available at: <https://doi.org/10.1002/wrna.1519>.
- Xu, W. *et al.* (2022) 'Dynamic control of chromatin-associated m6A methylation regulates nascent RNA synthesis', *Molecular Cell*, 82(6), pp. 1156–1168.e7. Available at: <https://doi.org/10.1016/j.molcel.2022.02.006>.
- Yamaguchi, Y. *et al.* (2002) 'Evidence that Negative Elongation Factor Represses Transcription Elongation through Binding to a DRB Sensitivity-Inducing Factor/RNA Polymerase II Complex and RNA', *Molecular and Cellular Biology*, 22(9). Available at: <https://doi.org/10.1128/mcb.22.9.2918-2927.2002>.
- Yang, F. *et al.* (2021) 'Shape of promoter antisense RNAs regulates ligand-induced transcription activation', *Nature*, 595(7867), pp. 444–449. Available at: <https://doi.org/10.1038/s41586-021-03589-x>.
- Yang, F. (2022) 'Promoter antisense RNAs: beyond transcription by-products of active promoters', *RNA Biology*. Taylor and Francis Ltd., pp. 533–540. Available at: <https://doi.org/10.1080/15476286.2022.2062177>.
- Yao, H. *et al.* (2010) 'Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA', *Genes and Development*, 24(22), pp. 2543–2555. Available at: <https://doi.org/10.1101/gad.1967810>.
- Yu, Z. *et al.* (2020) 'DDX5 resolves R-loops at DNA double-strand breaks to promote DNA repair and avoid chromosomal deletions', *NAR Cancer*, 2(3). Available at: <https://doi.org/10.1093/narcan/zcaa028>.

Zabidi, M.A. *et al.* (2015) 'Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation', *Nature*, 518(7540). Available at: <https://doi.org/10.1038/nature13994>.

Zenk, F. *et al.* (2021) 'HP1 drives de novo 3D genome reorganization in early *Drosophila* embryos', *Nature*, 593(7858), pp. 289–293. Available at: <https://doi.org/10.1038/s41586-021-03460-z>.

Zhang, H. *et al.* (2016) 'RNA helicase DEAD box protein 5 regulates Polycomb repressive complex 2/Hox transcript antisense intergenic RNA function in hepatitis B virus infection and hepatocarcinogenesis', *Hepatology*, 64(4). Available at: <https://doi.org/10.1002/hep.28698>.

Zhang, S. *et al.* (2023) 'Enhancer–promoter contact formation requires RNAPII and antagonizes loop extrusion', *Nature Genetics*, 55(5), pp. 832–840. Available at: <https://doi.org/10.1038/s41588-023-01364-4>.

Zhao, K., Hart, C.M. and Laemmli, U.K. (1995) *Visualization of Chromosomal Domains with Boundary Element-Associated Factor BEAF-32*, *Cell*.

Zhao, Y. *et al.* (2016) 'Activation of P-TEFb by Androgen Receptor-Regulated Enhancer RNAs in Castration-Resistant Prostate Cancer', *Cell Reports*, 15(3), pp. 599–610. Available at: <https://doi.org/10.1016/j.celrep.2016.03.038>.

Zhao, Y. *et al.* (2019) 'MyoD induced enhancer RNA interacts with hnRNPL to activate target gene transcription during myogenic differentiation', *Nature Communications*, 10(1). Available at: <https://doi.org/10.1038/s41467-019-13598-0>.

Zhu, A., Ibrahim, J.G. and Love, M.I. (2019) 'Heavy-Tailed prior distributions for sequence count data: Removing the noise and preserving large differences', *Bioinformatics*, 35(12). Available at: <https://doi.org/10.1093/bioinformatics/bty895>.

Zouaz, A. *et al.* (2017) 'The Hox proteins Ubx and AbdA collaborate with the transcription pausing factor M1 BP to regulate gene transcription', *The EMBO Journal*, 36(19), pp. 2887–2906. Available at: <https://doi.org/10.15252/embj.201695751>.

Zuin, J. *et al.* (2022) 'Nonlinear control of transcription through enhancer–promoter interactions', *Nature*, 604(7906), pp. 571–577. Available at: <https://doi.org/10.1038/s41586-022-04570-y>.

Titre : Rôles des Hélicases d'ARN et des ARN non-codants dans l'organisation spatiale et fonctionnelle du génome

Mots clés : Chromatine, ncRNAs, Insulateurs, Organisation des Chromosomes en 3D, Régulation des gènes

Résumé : La transcription des gènes est un processus nucléaire majeur qui influence les organismes et les cellules à toutes les étapes de leur développement. La transcription est donc précisément contrôlée pour assurer le bon fonctionnement des processus transcriptionnels. Au cours des dernières décennies, les domaines de la chromatine et de l'épigénétique sont devenus prévalents dans l'étude de la régulation de l'expression des gènes. Des nucléosomes et leurs modifications jusqu'aux larges structures 3D formées pour médier et contraindre les interactions des éléments génomiques, l'importance de la chromatine dans le contrôle de la transcription a été démontré à toutes ses étapes. Par exemple, les insulateurs limitent les contacts entre régions voisines pour empêcher certaines interactions régulatrices délétères, formant ainsi les Domaines Topologiquement Associés (TADs). Plus récemment, il a été découvert que la transcription n'a pas lieu seulement sur des gènes bien définis, mais de manière plus pervasive. Ainsi, les Promoteurs et les Enhancers, éléments principaux permettant la transcription des gènes, sont transcrits en ARN non-codants appelés PROMPTs et eRNAs. Ces derniers sont également proposés comme régulateurs de l'expression des gènes. La recherche des divers facteurs qui composent et influencent l'organisation du génome, à la fois spatialement et fonctionnellement, est encore en cours, tandis que de nouveaux facteurs et rôles sont régulièrement découverts. Parmi ceux-ci, les insulateurs sont composés d'environ 15 protéines chez la Drosophile, et au moins 4 chez l'Humain. Récemment, les ARNs, et en particulier les ARNs non-codants, ont été proposés comme régulateurs des protéines insulatrices, influençant leur liaison et leur oligomérisation à certains loci. Dans ce contexte, nous avons découvert plusieurs hélicases d'ARNs en interaction avec les complexes insulateurs dans des cellules S2 de *Drosophila melanogaster*. Nous nous sommes alors concentrés sur 3 de ces hélicases d'ARNs, à savoir Mtr4, Rm62 (DDX5 chez l'Humain) et Abstrakt (DDX41 chez l'Humain), pour caractériser leur potentiel rôle de co-facteurs d'insulateurs. Dans cette étude, nous utilisons principalement des méthodes de séquençage du génome entier analysés en bio-informatique, pour mesurer la régulation des ARNs codants et non codants (RNAseq), le recrutement à la chromatine de plusieurs facteurs insulateurs (ChIPseq) ainsi que l'influence de ces hélicases d'ARN sur la conformation des chromosomes (HiC). Nous complétons ces approches génomiques par des analyses en microscopie portant sur l'impact de ces hélicases d'ARNs et des ARN non-codants sur la formation des condensats insulateurs et des condensats de transcription. Enfin, nous avons produit une preuve de concept pour cibler l'accumulation de eRNAs spécifiques par CRISPR-dCas9, et donc potentiellement modifier l'expression des gènes et le recrutement des insulateurs. Malgré notre manque de compréhension des mécanismes sous-jacents, nous avons pu déterminer que les 3 hélicases d'ARN d'intérêt ont des effets contextuels dépendants du type de complexe insulateur considéré. Nous proposons que ces hélicases aient des rôles différents, à des sites distincts, potentiellement via plusieurs mécanismes. Nous émettons l'hypothèse que les hélicases d'ARN soient un groupe important de potentiels régulateurs des complexes insulateurs. Ils pourraient alors expliquer une partie des changements de comportement des insulateurs en fonction des sites. Au cours de cette étude, l'accumulation globale d'ARNs non-codants nous a permis d'évaluer leur effet indépendamment de leur transcription, ce qui fait encore débat dans ce domaine. Cette étude apporte de nouveaux éléments quant à la diversité des mécanismes impactés par les ARN non-codants et leur régulation et pose les bases de futures investigations relatives aux régulations locales de la chromatine par les ARN.

Title: Roles of RNA Helicases and ncRNAs in genome spatial and functional organisation

Key words: Chromosome Organization in 3D, Insulators, Chromatin, ncRNAs, Gene regulation

Abstract: Genes' transcription is one of the foremost nuclear processes, influencing cells and organisms at virtually all stages of life and development. It is therefore tightly controlled to ensure proper gene expression programs and avoid deleterious mis-regulations. In the last decades, the blooming fields of Chromatin and Epigenetic have become prevalent in the study of gene expression regulations. Truly, from nucleosomes and their modifications, to the large-scale 3D structures formed to mediate and constraint spatial interactions of genomic elements, all the components of chromatin have been demonstrated as influential in controlling transcription at all steps. Insulator elements, for example, were found to limit spatial contacts between neighbouring regions to avoid unwanted regulatory interactions, thus forming Topologically Associating Domains (TADs). More recently, transcription was shown to happen not only at defined genes, but more pervasively in nearly all contexts. Notably, Enhancers and Promoters, the two main genomic elements enabling gene transcription, are transcribed into relatively unstable non-coding RNAs, named eRNAs and PROMPTs, which are thought to participate in gene regulation. The search for the various factors that mediate and influence genome organisation, both spatially and functionally, is still ongoing, with new roles and factors discovered regularly. Insulators in particular have been shown in the last two decades to be composed of around 15 proteins in *Drosophila*, and at least 4 in Human. Recently, RNAs, and in particular non-coding RNAs, have been proposed as regulators of insulator proteins, influencing their binding and oligomerization at specific loci. In this context, we found several RNA helicases in interactions with insulator complexes in *Drosophila melanogaster* S2 cells. Hence, we focused on three of these RNA helicases, namely Mtr4, Rm62 (DDX5 in Human) and Abstrakt (DDX41 in Human), to decipher their potential roles as insulator co-factors. In this study, we use predominantly genome-wide sequencing and bioinformatic analyses to investigate coding and non-coding RNAs regulations (RNAseq), as well as the recruitment and binding to chromatin of several insulator proteins (ChIPseq), and the influences of these RNA helicases on 3D chromosome conformation (HiC). We complement these genomic approaches with microscopy experiments to examine the impact of RNA helicases and ncRNAs on the formation of transcriptional and insulator condensates. Finally, we also make a proof-of-concept experiment using CRISPR-dCas9 to target the accumulation of specific eRNAs and thus potentially enact changes in insulator recruitment and gene expression. Despite our relative lack of understanding of the underlying mechanisms, we have identified that all three RNA helicases have contextualized effects that depend on the type of insulator complexes that are considered. We propose that these helicases exert different effects and at different sites, possibly through separate mechanisms. Thus, we support that RNA helicases may be an important branch of insulator complexes regulators that account for some of the previously described loci-dependent behaviour of insulators. Along this course, the global accumulation of non-coding RNAs upon depletion of Mtr4 has also permitted us to study the effects of non-coding RNAs accumulation when uncoupled from their transcription, a current subject of intense debate in the field. This study brings new evidence of the diversity of mechanisms impacted by non-coding RNAs and their regulation, laying the ground for future investigations of local chromatin regulations by RNAs.