RNA-based mechanisms of chromatin and gene regulation

Jordi Planells Rodrigo
RNA-based mechanisms of chromatin and gene regulation

Jordi Planells Rodrigo

Academic dissertation for the Degree of Doctor of Philosophy in Molecular Bioscience at Stockholm University to be publicly defended on Friday 20 October 2023 at 09.30 in Vivi Täckholms (Q-salen), NPQ-huset, Svante Arrhenius väg 20.

Abstract

Cells regulate gene expression in a particular spatio-temporal manner by executing specific transcriptional programs. The different transcriptional programs allow the generation of the different cell types in multicellular organisms and mediate particular cellular responses to external cues such as viral infections. The regulation of how, when and which information is accessed in the DNA is vital for the normal biological function of cells and organisms.

The classic approach for studying gene expression has focused on the proteins involved in chromatin regulation, transcription and post-transcriptional gene expression events. During the years, it has become clear that not only proteins but also RNAs play important roles in the regulation of specific gene expression programs. However, how the RNA component of the chromatin modulates gene expression is to this date not fully understood. The main objective of this thesis has been to expand the knowledge about different mechanisms by which RNA impacts gene expression. In particular, at the transcriptional level, we have studied how RNA influences transcription and chromatin compaction by characterizing the chromatin-associated RNAs in Drosophila melanogaster. At the post-transcriptional level, the thesis aims at understanding the functions of ADAR3, a member of the Adenosine Deaminase Acting on RNA (ADAR) protein family that is catalytically inactive.

In paper I, we have investigated the role of the RNA exosome in the degradation of RNA on chromatin, and we have tried to understand the importance of this phenomenon for the regulation of chromatin accessibility. We hypothesized that deregulation of RNA abundance on chromatin can directly impact chromatin compaction and transcriptional outcome. To address this question, we profiled the chromatin-associated transcriptome of Drosophila melanogaster S2 cells in control cells and in cells depleted of RRP6 and DIS3, the two RNA exosome catalytic subunits. Moreover, to measure chromatin compaction, we performed ATAC-seq and asked whether the deregulation of the chromatin-associated transcriptome had any impact on chromatin compaction. We have identified a subset of genes involved in development and morphogenesis the regulation of which is dependent on the RNA exosome. These genes show alterations of both RNA levels and chromatin compaction in exosome depleted cells, and are characterized by the presence of Polycomb and Trithorax factors, suggesting that RNA degradation by the RNA exosome is necessary to maintain the homeostasis of balanced chromatin states. The RNA exosome has also been linked to the homeostatic control of RNAs derived from repetitive regions of the genome. Thus, we have analyzed the repetitive transcriptome and show that the RNA exosome is required to maintain the structural integrity of the centromeric regions, where we have observed that the RNA exosome is required for the repression of specific transposon families, mainly LTR retrotransposons.

Another finding of paper I is that snoRNAs are enriched on chromatin. In paper II, we have characterized a novel function for a specific snoRNA gene, snoRNA:U3:9B. We hypothesized that the snoRNA:U3:9B can impact gene transcriptional regulation by modulating the chromatin compaction of target genes, and that this regulation is biologically relevant in specific scenarios. We have used an inducible Sindbis viral replicon as a model of viral infection, and we show that the snoRNA:U3:9B is induced upon induction of the Sindbis replicon. We have observed attenuated transcriptional response of immune response genes in flies depleted of snoRNA:U3:9B, accompanied by decreased chromatin accessibility in these genes, which suggests that the snoRNA:U3:9B modulates the chromatin accessibility of genes implicated in D. melanogaster immune response.

In paper III, we have characterized the function of ADAR3 with focus on brain development. We observed that ADAR3 is broadly expressed in cortical neurons during early development but only expressed in a small subpopulation of in vitro differentiated cortical neurons, which suggests a specialized function in the mature brain. We generated a transgenic cell line by ectopically introducing ADAR3 in a neuroblastoma cell line (N2a) that did not express ADAR3 endogenously. To characterize ADAR3 function, we performed co-immunoprecipitation experiments followed by mass spectrometry to identify the ADAR3 interactome, which suggested links to mRNA stability and translation. Studies of mRNA stability based on high-throughput RNA-sequencing revealed that ADAR3 affects the stability of a large number of mRNAs. However, the corresponding protein levels were in general not affected. Moreover, we observed that ADAR3 inhibits translation and is associated with polysomes. Taken together, our results suggest that ADAR3 binds and stabilizes specific mRNAs in non-productive polysome complexes. Some of the affected transcripts are related to neurogenesis and we propose a model in which ADAR3 inhibits neuronal differentiation during early brain development.

Keywords: Gene regulation, chromatin, non-coding RNAs, genome-wide studies, Drosophila melanogaster, ADAR3.
RNA-BASED MECHANISMS OF CHROMATIN AND GENE REGULATION

Jordi Planells Rodrigo
RNA-based mechanisms of chromatin and gene regulation

Jordi Planells Rodrigo
A tota la gent que m'estima i m'ha ajudat a arribar fins ací
Title
RNA-based mechanisms of chromatin and gene regulation

Respondent
Jordi Planells Rodrigo, Department of Molecular Biosciences, The Wenner-Gren Institute
Stockholm University, Sweden

Opponent
Simon Elsässer, Department of Medical Biochemistry and Biophysics
Karolinska Institutet, SciLifeLab, Sweden

Examination committee
Qi Dai, Department of Molecular Biosciences, The Wenner-Gren Institute
Stockholm University, Sweden

Andrea Hinas, Department of Cell and Molecular Biology, Microbiology and immunology
Uppsala universitet, Sweden

Yuri Schwartz, Department of Molecular Biology
Umeå University, Sweden

Einar Hallberg, Department of Biochemistry and Biophysics
Stockholm University (Reserve opponent)

Chairman of the public defense
Ulrich Theopold, Department of Molecular Biosciences, The Wenner-Gren Institute
Stockholm University, Sweden

Supervisor
Neus Visa, Department of Molecular Biosciences, The Wenner-Gren Institute
Stockholm University, Sweden

Co-supervisor
Antonio Jordán Pla, Institute of Biomedicine in Valencia (IBV-CSIC)
University of Valencia, Spain
Thesis abstract

Cells regulate gene expression in a particular spatio-temporal manner by executing specific transcriptional programs. The different transcriptional programs allow the generation of the different cell types in multicellular organisms and mediate particular cellular responses to external cues such as viral infections. The regulation of how, when and which information is accessed in the DNA is vital for the normal biological function of cells and organisms.

The classic approach for studying gene expression has focused on the proteins involved in chromatin regulation, transcription and post-transcriptional gene expression events. During the years, it has become clear that not only proteins but also RNAs play important roles in the regulation of specific gene expression programs. However, how the RNA component of the chromatin modulates gene expression is to this date not fully understood. The main objective of this thesis has been to expand the knowledge about different mechanisms by which RNA impacts gene expression. In particular, at the transcriptional level, we have studied how RNA influences transcription and chromatin compaction by characterizing the chromatin-associated RNAs in *Drosophila melanogaster*. At the post-transcriptional level, the thesis aims at understanding the functions of ADAR3, a member of the Adenosine Deaminase Acting on RNA (ADAR) protein family that is catalytically inactive.

In paper I, we have investigated the role of the RNA exosome in the degradation of RNA on chromatin, and we have tried to understand the importance of this phenomenon for the regulation of chromatin accessibility. We hypothesized that deregulation of RNA abundance on chromatin can directly impact chromatin compaction and transcriptional outcome. To address this question, we profiled the chromatin-associated transcriptome of *Drosophila melanogaster* S2 cells in control cells and in cells depleted of RRP6 and DIS3, the two RNA exosome catalytic subunits. Moreover, to measure chromatin compaction, we performed ATAC-seq and asked whether the deregulation of the chromatin-associated transcriptome had any impact on chromatin compaction. We have identified a subset of genes involved in development and morphogenesis the regulation of which is dependent on the RNA exosome. These genes show alterations of both RNA levels and chromatin compaction in exosome depleted cells, and are characterized by the presence of Polycomb and Trithorax factors, suggesting that RNA degradation by the RNA exosome is necessary to maintain the homeostasis of balanced chromatin states. The RNA exosome has also been linked to the homeostatic control of RNAs derived from repetitive regions of the genome. Thus, we have analyzed the repetitive transcriptome and show that the RNA exosome is required to maintain the structural integrity of the centromeric regions, where we have observed that the
RNA exosome is required for the repression of specific transposon families, mainly LTR retrotransposons.

Another finding of paper I is that snoRNAs are enriched on chromatin. In paper II, we have characterized a novel function for a specific snoRNA gene, snoRNA:U3:9B. We hypothesized that the snoRNA:U3:9B can impact gene transcriptional regulation by modulating the chromatin compaction of target genes, and that this regulation is biologically relevant in specific scenarios. We have used an inducible Sindbis viral replicon as a model of viral infection, and we show that the snoRNA:U3:9B is induced upon induction of the Sindbis replicon. We have observed attenuated transcriptional response of immune response genes in flies depleted of snoRNA:U3:9B, accompanied by decreased chromatin accessibility in these genes, which suggests that the snoRNA:U3:9B modulates the chromatin accessibility of genes implicated in D. melanogaster immune response.

In paper III, we have characterized the function of ADAR3 with focus on brain development. We observed that ADAR3 is broadly expressed in cortical neurons during early development but only expressed in a small subpopulation of in vitro differentiated cortical neurons, which suggests a specialized function in the mature brain. We generated a transgenic cell line by ectopically introducing ADAR3 in a neuroblastoma cell line (N2a) that did not express ADAR3 endogenously. To characterize ADAR3 function, we performed co-immunoprecipitation experiments followed by mass spectrometry to identify the ADAR3 interactome, which suggested links to mRNA stability and translation. Studies of mRNA stability based on high-throughput RNA-sequencing revealed that ADAR3 affects the stability of a large number of mRNAs. However, the corresponding protein levels were in general not affected. Moreover, we observed that ADAR3 inhibits translation and is associated with polysomes. Taken together, our results suggest that ADAR3 binds and stabilizes specific mRNAs in non-productive polysome complexes. Some of the affected transcripts are related to neurogenesis and we propose a model in which ADAR3 inhibits neuronal differentiation during early brain development.

Keywords: Gene regulation, chromatin, non-coding RNAs, genome-wide studies, Drosophila melanogaster, ADAR3
Populärvetenskaplig sammanfattning

Celler reglerar sitt genuttryck både temporärt och rumsligt genom att följa specifika transkriptionsprogram. Olika transkriptionsprogram möjliggör utvecklingen av skilda celltyper i flercelliga organismer samt förmedlar specifika cellulära svar på externa signaler såsom virusinfektioner. Regleringen av hur, när och vilken information i DNA som görs tillgänglig för transkription är avgörande för cellers och organismers normala biologiska funktion.


RNA-exosomen krävs för att upprätthålla strukturen av kromosomernas centromerer genom att hämma uttrycket av specifika familjer av transposoner, främst LTR-retrotransposoner.


Manuscripts and publications

List of manuscripts included in this thesis

I. The exosome degrades chromatin-associated RNAs genome-wide and maintains chromatin homeostasis
   Jordi Planells, Antonio Jordán-Pla, Shruti Jain, Juan José Guadalupe, Estelle Proux-Wera, Anne von Euler, Vicent Pelechano, Neus Visa
   bioRxiv 2023.04.10.536209; doi: https://doi.org/10.1101/2023.04.10.536209
   In revision

II. SnoRNA:U3:9B in required for the activation of immune response genes in Drosophila melanogaster
    Shruti Jain, Jordi Planells, Isabel Regadas, Donal Barrett, Anne von Euler, Bo Lindberg, Ylva Engström, Vicent Pelechano, Mattias Mannervik, Neus Visa
    Manuscript in preparation

III. ADAR3 modulates neuronal differentiation and regulates mRNA stability and translation
     Victor Karlström, Eduardo A. Sagredo, Jordi Planells, Charlotte Welinder, Jennifer Jungfleisch, Andrea Barrera Conde, Chammiran Daniel, Fátima Gebauer, Neus Visa and Marie Öhman
     In revision

Publications not included in the thesis

I. Genetic inactivation of essential HSF1 reveals an isolated transcriptional stress response selectively induced by protein misfolding.
   Ciccarelli M, Masser AE, Kaimal JM, Planells J, Andréasson C.
Table of contents

Thesis abstract.................................................................................................................................................. 1
Populärvetenskaplig sammanfattning............................................................................................................. 3
Manuscripts and publications........................................................................................................................... 5
Abbreviations.................................................................................................................................................... 7

Chapter 1: Introduction...................................................................................................................................... 8
  Brief introduction to DNA and chromatin................................................................................................. 8
    DNA as information vault.......................................................................................................................... 8
  DNA has to be compacted in the cell nucleus ............................................................................................. 8
  The transcription process and transcriptional regulation of gene expression........................................... 10
    Eukaryotic gene expression.................................................................................................................... 10
  An overview of eukaryotic transcription.................................................................................................. 11
  An overview of RNA pol II transcriptional stages................................................................................... 11
  Transcription initiation, where and how? - Promoter and mechanistic insights of transcription initiation.12
  Regulation of transcription by TFs........................................................................................................... 14
  Regulation of transcription by chromatin environment........................................................................... 15

RNA in the nucleus: RNA regulation of chromatin processes and RNA degradation............................... 18
  Chromatin-associated RNAs................................................................................................................... 18
  Regulation of chromatin by caRNAs.......................................................................................................... 19
  The RNA exosome and its functions......................................................................................................... 21

Heterochromatin in eukaryotic genomes: Formation and regulation........................................................... 24
  Constitutive heterochromatin.................................................................................................................... 24
  Facultative heterochromatin - An overview of Polycomb repression................................................... 25
  Polycomb repressive complexes - Who?..................................................................................................... 27
  Polycomb repressive complexes - Where?................................................................................................. 28
  Polycomb repressive complexes - How?..................................................................................................... 30

Post-transcriptional regulation of gene expression..................................................................................... 31
  Post-transcriptional processes and RNA stability.................................................................................. 31
  ADAR family of enzymes and Adenosine-to-Inosine RNA modification.............................................. 33
    A-to-I RNA editing.................................................................................................................................. 33
    ADAR proteins catalyze A-to-I RNA editing......................................................................................... 34
    ADAR3 and its unknown molecular function....................................................................................... 36
  SnoRNAs and their functions: Post-transcriptional RNA modifiers?.................................................... 37
    snoRNAs biochemical properties and classification.............................................................................. 37

Chapter 2: Aims, Results, Conclusions and Future Perspectives.................................................................... 41

  Paper 1....................................................................................................................................................... 42
  Paper 2....................................................................................................................................................... 47
  Paper 3....................................................................................................................................................... 51

Main conclusions of this thesis..................................................................................................................... 55
References....................................................................................................................................................... 56
Acknowledgements......................................................................................................................................... 73
Abbreviations

5/3'UTR = Five/three prime UnTranslated Region
A-to-I = Adenosine to Inosine
A,T,C,G = Adenosine, Thymidine, Cytidine, Guanosine
ADAR = Adenosine Deaminase Acting on RNA
bp = base pair
CGI = CpG Island
caRNAs = Chromatin-associated RNAs
CTD = Carboxy-Terminal Domain
CUT = Cryptic Unstable Transcript
dDNA Degradation Rate
dsRNA = double-stranded RNA
E(z)/EZH1/2 = Enhancer of Zeste
eRNA = enhancer RNA
GAF/TrxG = GAGA-Associated Factor/Trithorax-Group
GTF = General Transcription Factor
KMT = histone Lysine Methyl-Transferase
IncRNA = long non-coding RNA
miRNA = microRNA
mRNA = Messenger RNA
ncRNA = non-coding RNA
NDR = Nucleosome Depleted Region
NEXT = Nuclear EXosome-Targeting complex
NGD = No-Go Decay
NGS = Next Generation Sequencing
NMD = Nonsense Mediated Decay
NSD = Non-Stop Decay
PAS = Poly-Adenylation Signal
PAXT = PolyA eXosome-Targeting complex
Pc = Polycomb
Ph = Polyhomeotic
Pho = Pleiohomeotic
PIC = Pre-Initiation Complex
piRNA = Piwi-interacting RNA
PcG = Polycomb Group
PRE = Polycomb Response Elements
pre-(m)RNA = precursor (m)RNA
PROMPT = PROMoter uPstream Transcript
Psc = Posterior sex combs
PTC = Premature Termination Codon
PTM = Post-Translational Modification
RA = mRNA Abundance
RBBP4/7 = RetinoBlastoma Binding Protein
RDRC = RNA-Dependent RNA polymerase Complex
repRNA = repeat-derived RNA
RNA-seq = high-throughput RNA sequencing
RNApol = RNA polymerase
RNB = ribonuclease domain
RNP = RiboNucleoProtein complex
rRNA = ribosomal RNA
Sce = Sex comb extra
SKI = SuperKiller complex
snoRNA = small nucleolar RNA
snRNA = small nuclear RNA
SR = mRNA Synthesis Rate
ssRNA = single-stranded RNA
Su(z)12/SUZ12 = Suppressor of Zeste
TBP = TATA-Binding protein
TF = Transcription Factor
TFIID = Transcription Factor IID
TRAMP = TRf4/Air2/Mtr4p Polyadenylation complex
tRNA = Transference RNA
TSS = Transcription Start Si
Chapter 1: Introduction

Brief introduction to DNA and chromatin

DNA as information vault
The DNA molecule stores the necessary information for living organisms to exist. What today we consider an axiom, was uncertain 70 years ago. This idea started to bloom back in 1944, with the experiments performed by Avery, McCarthey and Macleod who suggested that the DNA was the driver of information transmission between different bacterial strains [1]. Watson and Crick, shortly after, elucidated the DNA structure in 1953 [2] and highlighted two key features. First, that the DNA consists of two antiparallel helices and second, the complementarity of both strands. These features allowed them to propose the inheritance mechanism of the DNA. The fact that DNA was composed of two complementary helices allowed the possibility of using each strand as a template to copy the other strand.

The DNA is composed of four different deoxynucleotides, adenosine (A), thymidine (T), guanosine (G) and cytidine (C). We now have more insights into the DNA properties. Structurally, the DNA molecule is approximately 2 nm wide and the latest estimate of DNA length in humans is 206.62 ± 1.99 cm on average for the male and female diploid genome (205.00 ± 1.97 and 208.23 ± 2.00 respectively) [3]. Extrapolating this number, if taken an average of 3*10^{12} nucleated cells in a single human organism [4], the total DNA length in the human body results in about 6.2 billion km, sufficient to cover the Earth-Sun distance over 40 times. Piovesan and colleagues [3] also provide an approximate DNA weight per cell of 6.46 pg and using the latest published release of the human genome [5], they calculated an approximate GC content of 40.89%.

In recent years, synthetic DNA has been proposed as an efficient way of data storage, and self-assembling DNA molecules have been used for molecule transport and delivery [6], widening the DNA research outside the molecular biology field.

DNA has to be compacted in the cell nucleus
With the aforementioned biophysical properties, the need of compacting the DNA molecule inside the eukaryotic cell nucleus stands out. The diameter of a human cell spans from 7-8 µm (erythrocytes [7]) to 100-120 µm (megakaryocytes [8] and mature female eggs, respectively) and the average nucleus diameter of human cells is around 10 µm [9]. Packaging the DNA into such a small compartment presents a difficult challenge for the
cells, both physically and operationally. In order to overcome the physical problem of fitting the DNA molecule inside the cell nucleus, DNA is wrapped around proteins called histones, constituting a DNA-protein complex called nucleosome. The canonical nucleosome includes two copies of each of the histone core proteins (H2A, H2B, H3 and H4). Packaging is achieved thanks to the histones’ chemical properties. Histones are small basic proteins that use their positive charge to interact with the DNA [10]. The DNA winds around the histone octamer approximately 1.7 turns (146 bp) to constitute the nucleosome complex [11]. Nucleosomes are approximately 206 KDa and are separated by ~20–90 bp of DNA known as linker DNA [10]. The length of the linker DNA can vary among different species and tissues. Besides core histones, the histone H1 also plays an important role in the nucleosomal organization. The histone H1 sits at the entry/exit of the nucleosomes, protecting additional 20 bp of DNA in the nucleosome [12]. The histone H1 contains numerous basic amino acid residues that neutralize the negative charges of the linker DNA, reducing its flexibility and bringing the nucleosomes closer [13]. All three components form the nucleosome repeat monomer, the basic chromatin unit. In humans, it has an approximate length of 200 bp (197 bp).

The nucleosomes and linker DNA form the 10 nm fiber that was first described in the late seventies, and it was called ‘beads-on-a-string’ structure [14,15]. Wrapping around the DNA onto histones reduces considerably the length of the DNA molecule. However, further compaction is needed in order to accommodate the DNA inside the nucleus. Early in vitro experiments showed the ability of purified chromatin to form more compact structures in the presence of cations (Mg²⁺, Na⁺) and histone H1 [16]. This more compact structure was termed 30 nm fiber. However, several studies in vivo could not prove the existence of the 30 nm fiber, contradicting the in vitro model [17]. In fact, the current accepted model is that 10 nm fiber is heterogeneous in size and can form lumps or clutches ranging from 6 to 24 nm [18]. Why the 30 nm fiber is not formed in living cells depends on several factors that include the different concentration of cations used in vitro and in vivo; that other proteins binding to chromatin also alter the electrostatic interactions required for the 30 nm fiber formation; and the irregular spacing (positioning) of nucleosomes in vivo (extensively reviewed in [19]). For these reasons, the 30 nm fiber is not considered a basic unit of chromatin conformation, but rather a conformation that is transient or confined to heavily heterochromatic regions [20,21]. This latest model is also consistent with the highly dynamic nature described for the chromatin organization in eukaryotic cells [22], which would not be observed if the 30 nm fiber was the predominant chromatin conformation. Moreover, it can also explain the 3D conformation of the genome into dynamic topologically associated domains or chromosome territories in the eukaryotic nucleus [23].
The transcription process and transcriptional regulation of gene expression

Eukaryotic gene expression

The information stored in the DNA is organized in operational units called genes. The process by which a gene generates a functional product is known as gene expression, and includes the synthesis of functional RNAs and proteins. Gene expression is an extremely complex process, and its regulation dictates the cellular identity and fate. Gene expression starts with the transcription of the gene, resulting in a precursor RNA (pre-RNA). The pre-RNA is then processed to form a mature RNA. The modifications that pre-RNAs undergo are specific for each type of RNA. For instance, maturation of pre-messenger RNA (pre-mRNAs) includes capping, splicing of introns, poly-adenylation of the 3’ end, and other chemical modifications such as editing. When referring to pre-mRNAs, another important event is the packaging of the pre-mRNA into a ribonucleoprotein particle (mRNP). Mature mRNAs are exported to the cytoplasm and translated by the ribosomes. With so many different processes involved in gene expression, the cell can regulate protein production through a plethora of different mechanisms.

RNA expression is often measured as RNA abundance (RA) and can be broken down into two separate processes, RNA synthesis and RNA degradation. Both processes have specific kinetics, the synthesis rate (SR) and the turnover or degradation rate (DR). In steady-state cells, the abundance of a specific transcript X can be modeled simplistically as: \[ RA(x) = SR(x) - DR(x) \] [24]. Although synthesis and degradation appear as two different processes, it has been shown that they can be intertwined. Haimovich and collaborators showed that Xrn1p, the protein responsible for the main cytoplasmic RNA degradation pathway, can shuttle to the nucleus in specific scenarios to act as transcription factor and boost the mRNA production of a subset of genes [25].

The complexity of gene expression and its regulation have been extensively studied, but to this day, many questions remain unsolved. In this thesis, the focus has been to study the impact of RNA on the packaging of the chromatin (projects 1 and 2) and how RNA translation and stability impact the process of gene expression (project 3).
An overview of eukaryotic transcription

The protein that transcribes the DNA into RNA is the RNA polymerase (RNAPol). Eukaryotic organisms count with up to five different RNAPols. Three of them are central to all eukaryotes. RNAPol I transcribes ribosomal RNA (rRNA), RNAPol II is devoted to the transcription of a wide catalog of genes, including protein-coding genes and non-coding RNAs (ncRNAs) such as long non-coding RNAs (lncRNAs), small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs) and other ncRNAs that include repetitive transcripts and cryptic transcripts. Lastly RNAPol III transcribes short RNAs, for instance tRNAs and 5.8S rRNA (reviewed in [26]). The transcripts produced by the different RNAPols widely differ in structure, diversity and amount. RNAPol I produces a single transcript, the pre-rRNA, that accounts for between 80-90% of the total RNA in eukaryotic cells [27]. RNAPol III transcripts (tRNAs, 5.8S and U6 snRNA) also account for an important fraction of the total RNA in the cells (3–7%) [27]. In fact, RNAPol III transcripts constitute the majority of the soluble RNA (RNA pool) in eukaryotic cells. Lastly, the RNAPol II produces the most heterogeneous RNA pool. While RNAPol I and III are highly specialized enzymes, RNAPol II is required to be extremely versatile, due to the wide repertoire of genes this enzyme is transcribing in eukaryotic cells. More than 200 thousand transcripts can be produced in a typical human cell [28], and deciphering the regulatory networks and mechanisms that dictate which, how, where and when genes are transcribed by RNAPol II, allows us to understand the biological plan of eukaryotic organisms. Despite the different RNA species each RNAPol is transcribing, the three RNAPol complexes operate mechanistically similarly. Transcription can be subdivided in three different stages for all three different RNA polymerases: transcription initiation, elongation and termination. These stages will be addressed in the following sections with specific focus on RNAPol II transcription.

An overview of RNA pol II transcriptional stages

The major determinant of RNAPol II transcription is the initiation. Transcription initiation includes promoter recognition and pre-initiation complex (PIC) formation. The currently accepted model for initiation involves the formation of the PIC and recruitment of RNAPol II in a stepwise sequential manner. The PIC is assembled in the promoter of the genes, and includes both the RNAPol II and general transcription factors (GTFs) that incentivize the RNAPol II activity [29]. Through a number of conformational changes induced by the binding of specific proteins and the change of carboxy-terminal domain (CTD) phosphorylation status, the RNAPol II escapes the promoter and starts transcription. Transcription is then
stopped in most eukaryotic genes 20-120 bp downstream of the transcription start site (TSS) in a process called promoter proximal pausing [30,31]. This pause constitutes a checkpoint in the transition from initiation to productive elongation. The activity of positive transcription elongation factor b (P-TEFb) is required to release RNApol II into the gene body [32]. Once the proximal-pausing has been resolved, the RNApol II starts the productive elongation phase. Some of the pausing factors, for example DSIF (DRB Sensitivity Inducing Factor), turn into elongation factors that modulate the RNApol II speed and/or processivity. Others, for example PAF1C, modify the chromatin landscape across the gene bodies by recruiting chromosome remodelers [31].

Transcription terminates when both RNApol II and the nascent RNA are released from the DNA. This process differs between protein-coding genes and ncRNA genes [33]. The difference resides in the intrinsic characteristics of the newly transcribed RNA. Protein-coding mRNAs have poly-A tails, while some ncRNAs do not. For most mRNAs with poly-A signals, the cleavage and polyadenylation factor (CPSF) contained in the CPF-cleavage factor complex is recruited by Serine 2 phosphorylated CTD of RNApol II and recognizes the A-rich poly-A signal in the 3’ untranslated region (3’UTR) of the mRNA [33]. Subsequently, this complex will perform an endonucleolytic cleavage, the RNA will be released, and the polyA tail will be added to the transcript. In ncRNAs, because of the absence of a polyA tail, the termination mechanism differs from the protein-coding genes. In yeast, the complex Nrd1-Nab3-Sen1 is required for ncRNA termination. It is thought that Sen1 translocates along the transcript associated with the RNApol II. Sen1 will ultimately reach the RNApol II causing a translocation and releasing the transcript [34].

Understanding the underlying mechanisms and regulation of all transcriptional stages is important. However, only the initiation mechanisms and their regulation will be relevant for this thesis and will be described in the following chapters.

Transcription initiation, where and how? - Promoter and mechanistic insights of transcription initiation

Gene transcription starts at a specific nucleotide at the 5’ end of the genes called TSS. The TSS is found in the core promoter, a DNA sequence that extends away from the TSS approximately 50 bp upstream and downstream. The core promoter of eukaryotic genes is a major regulation hub that serves as a scaffold for PIC and associated GTFs binding (Figure 1, reviewed in [36]).
**Figure 1: Transcription initiation in eukaryotic cells.** A) Transcriptional repressed state. DNA is compacted and nucleosomes are positioned on top of the promoter region, blocking the access to TSS. Nucleosomes are decorated with transcriptional repressing post-transcriptional modifications (PTMs), for instance H3K27me3 (red star). B) Active transcriptional state. Positioned nucleosomes on the promoter have been evicted/repositioned, generating a nucleosome free region that allows the recruitment of GTFs and RNApol II, that ultimately form the PIC. Nucleosomal PTMs differ from the repressed state. In active transcription, nucleosomes are decorated with H3K27ac (green square) and H3K4me3 in the promoter (dark green star).

Genome-wide studies of transcription initiation have identified different features that characterize and constitute the core promoter, among which we find sequence features, initiation pattern and chromatin configuration (reviewed in [37]). Sequence features refer to conserved sequences of nucleotides (DNA motifs) that are recognized by the transcription machinery. One of the most widely studied promoter motifs and one of the first motifs identified is the TATA-box motif [38]. Many other DNA motifs exist
in the promoter region, and extensive research has been done to identify and characterize them, including the development of databases containing sequences, targets and genomic locations [39-41]. Promoters are also characterized by how RNApol II starts transcribing. Two different initiation patterns have been described, including transcription initiated in one specific nucleotide (sharp initiation) or in several positions of the promoter (broad transcription [42]). Promoter chromatin conformation is also an important promoter defining feature. Some specific promoters show a well conserved nucleosome positioning pattern around the TSS, specific histone variants or specific histone tail modifications [43,44]. Taking into consideration the sequence features, the initiation pattern and the chromatin configuration, Haberle and Stark proposed that eukaryotic promoters can be subdivided in three different promoter classes [37].

Type III promoters are of special relevance for this thesis. These promoters are found in developmentally regulated genes, involved in patterning and morphogenesis. They present broad transcription initiation and a particular chromatin configuration. Type III core promoters are enriched both in chromatin modifications that enhance transcription (H3K4me3) and transcriptional repressive marks (H3K27me3), constituting what has been called a bivalent chromatin configuration. This chromatin organization is presumably used to prime transcription of developmental genes in specific cell types, while silencing them in others. In humans, type III promoters are overlapped by one long CGI or multiple CGIs and often produce divergent transcripts [44]. In contrast, type III promoters in flies are associated to a consensus sequence referred to as downstream promoter element and present sharp initiation [45].

Promoters in eukaryotic cells are by default not accessible to the RNApol II [46]. RNApol II access to the core promoter is impaired by the chromatin, more specifically by the nucleosomes. In order to start transcription, the nucleosomes need to be rearranged or evicted from the chromatin for transcription initiation to occur [46,47]. This means that the chromatin accessibility of the promoter for RNApol II is a major determinant in transcription initiation and can be controlled or modified by two different mechanisms: the binding of specific TFs and the chromatin environment.

Regulation of transcription by TFs

Unlike GTFs that are required for transcription of the vast majority of genes, TFs execute specific transcriptional programs. This means that TFs are highly specialized in driving the expression of concrete gene subsets related to a biological process or function. For instance, development is one of the most regulated processes in multicellular organisms. In metazoans, sequential growth and cell fate determination are governed by a family of
transcription factors called Hox (reviewed in [48]). TFs also participate in other processes such as responses to specific stresses or environmental cues. For example, the TF FOXO is involved in response to oxidative stress [49] and the HSP proteins in the heat shock response [50].

Although TFs are associated with specific transcriptional programs and are cell-type specific, some studies show that the same TF can regulate transcription of different genes in different cell types [51]. TFs have been widely implicated in different diseases. For instance, TFs are overrepresented among oncogenes [52], and one third of all human developmental diseases have been linked to a dysfunctional TF [53]. TFs are useful for research and can be used to induce differentiation [54] and even de-differentiation [55]. Evolutionarily, global gene regulatory networks are conserved across organisms [56]. However, when TF genes are duplicated, they rapidly evolve and diverge, giving as a result a specific transcriptional program that drives speciation [57].

TFs act by recognizing and binding to specific sequences that are located in the promoters of their target genes. Motifs recognized by TFs are usually 4-8 base pairs (bp) in length, and therefore, occur with high recurrence along the DNA. Generally, eukaryotic genes possess binding sites for several TFs, which allows eukaryotic cells to have exponential possibilities of combinatorial TF binding events, increasing the plasticity of the eukaryotic gene expression spectrum. In fact, it is commonly observed that several TFs need to collaborate in order to achieve specificity and activate transcription [58]. Mechanistically, TFs recognize their motifs and via protein-protein interaction recruit GTFs or the RNApol II itself to the promoter region. A second mechanism of action is the recruitment of proteins that can modify the chromatin conformation in the promoter region. This modification can be either by recruitment of chromatin remodeling factors (such as SWI/SNF) that use ATP to move or evict positioned nucleosomes, or through histone modifying enzymes that chemically modify the histone tails located in the promoter. The outcome of activator TF binding is always to open the chromatin of the core promoter region for PIC assembly. While the transcriptional activation by TF is an extremely important process in the transcriptional control of eukaryotic genes, the topic is not relevant for this thesis. Models and examples of TF transcriptional activation are well documented and reviewed in [59,60].

Regulation of transcription by chromatin environment

The default genomic DNA conformation, wrapped around histones, limits the accessibility of proteins to promoters for gene activation and transcription. For this reason, active promoters require a specific chromatin organization, inherently different from silenced genes. Active promoters are characterized by the lack of nucleosomes (nucleosome depleted region,
NDR), flanked by precisely positioned nucleosomes [61, reviewed by 62]. Some studies have proven that the NDR is not completely depleted of nucleosomes, but occupied with hyper-dynamic nucleosomes containing histone variants H3.3 and H2A.Z [63]. There are three main chromatin regulatory processes in promoter regions that allow the active promoter conformation: ATP-dependent chromatin remodeling, DNA methylation and chromatin (histones) post-translational modifications (histone PTMs).

Chromatin remodeling complexes utilize ATP to modify the nucleosome position on chromatin, eject nucleosomes from the chromatin or interchange histone subunits of the nucleosomes. Four different families of chromatin remodeling complexes are found in eukaryotic organisms, based on the similarities and differences in their ATPase function. These families are imitation switch (ISWI), chromodomain helicase DNA-binding (CHD), switch/sucrose non-fermentable (SWI/SNF) and inositol requiring 80 (INO80) [64]. The ISWI family complexes recognize the tail of the histone H3 and the linker DNA flanking nucleosomes [65]. Functionally, ISWI family complexes are devoted to regularly space nucleosomes to limit chromatin accessibility and transcription. There are exceptions like the NURF complex, whose function is to increase chromatin accessibility and promoter transcription [66]. The CHD complex remodelers share certain structural similarities to the ISWI complexes. However, CHD remodelers have also been implicated in transcription activation, for instance by incorporating the highly dynamic histone H3.3 in the NDR [67,68].

Chromatin access is granted mostly by the SWI/SNF complexes that operate nucleosomes at different levels, including nucleosome sliding, nucleosome eviction and nucleosome ejection from chromatin [69]. Lastly, the INO80 family is mostly devoted to nucleosome editing, by exchanging canonical histones with specific variants. The histone variants present in the nucleosomes are important, because they impact the dynamics of recruitment, eviction or ejection of the nucleosomes from chromatin [70].

DNA methylation is also a major determinant of promoter configuration in vertebrates. In fact, approximately 70% of the core promoters overlap CGIs [71] and, depending on the cellular context, CGIs exhibit high levels of methylation, ranging between 70-80% [72]. Housekeeping genes and developmentally regulated genes (type III promoters) have the highest CGIs prevalence [43,71]. CGIs are regions with high GC content, and more precisely, a specific dinucleotide combination, CpG. DNA methylation occurs in Cytosines (5mC), and earlier studies about DNA methylation showed a clear correlation with transcriptional repression [73], which increases with the density of CpG dinucleotides contained in CGIs [74]. How DNA methylation drives transcriptional repression involves many different mechanisms, and to this date it is still not fully understood. One mechanism by which methylation can induce repression is by changing the pool of chromatin proteins that are recruited to a specific genomic loci. Yin and colleagues compared the binding 16
profiles of TFs in methylated and unmethylated promoters genome-wide. They demonstrated that approximately 20% of the addressed TFs showed decreased affinity by methylated promoters [75]. Another example of how DNA methylation influences transcription, is the polycomb repressive group (PcG) that in mammals is targeted via methylated CpG (reviewed in [76]). CGIs and their methylation is one of the main differences between mammalian and invertebrate promoters. Most of the studied invertebrates do not present CGIs nor methylation, or its presence is residual and its biological relevance is still in debate, for instance in D. melanogaster [77].

Another chromatin determinant of promoter accessibility is histone modifications. The histone modifications have also been extensively studied, especially the PTMs of the N-Terminal domain (or histone tails). The two most studied histone modifications are the acetylation and methylation of residues in the histone tails, and to a lesser extent, in the globular domains. The histone PTMs were, since long ago, proposed to influence gene transcription [78,79]. However, the identification and mechanistic characterization of the vast majority of the enzymes catalyzing PTMs occurred during the last years of the 1990s decade. It was shortly after, together with the resolution of the human DNA sequence, when the histone code hypothesis was proposed [80,81]. Histone modifications are deposited by specialized enzymes, known as ‘writers’. These modifications can be interpreted by a wide catalog of proteins, the ‘readers’, and can be reversed by a set of enzymes, the ‘erasers’, that are able to catalyze the inverse reaction of the writers. Histone PTMs and their regulation allow complex regulatory networks that act sequentially or in a combinatorial manner to enable the cells to regulate gene expression.

Acetylation and methylation are the most widely studied and characterized histone PTMs. Histone acetylation has been linked to many different cellular processes, including development, DNA repair and disease [reviewed in 82,83 and 84 respectively]. Histone acetylation is thought to incentivize transcription activation because it neutralizes the positive charges of the lysine residues, thus loosening the interaction between the histones and the DNA, as demonstrated both in vitro and in vivo [85,86]. Histone acetylation has been detected in different residues, some examples of it are: H3K4, H3K9, H3K27, H3K36, H3K79, H4K5, H4K12, and H4K20. The most studied acetylation is H3K27ac, which usually localizes at promoters and enhancers of actively transcribed genes, where it co-exists with H3K4me3 [87].

Histone methylation has also been linked to many different cellular processes such as development, DNA damage response (DDR) and disease [reviewed in 88,89 and 90 respectively]. The mechanism of action of histone methylation is not as straightforward as the acetylation mechanism. Methylation does not change the charge of the amino acids, but
acts rather by creating binding sites for the recruitment of chromatin factors. Histone methylation has been both linked to activation and repression of transcription. H3K4me3, can be found in active enhancers [91], whereas constitutive heterochromatin (Polycomb complex regulated regions) and structural heterochromatin (centromeric and telomeric regions) are enriched in H3K27me3 and H3K9me3, respectively [92 and reviewed in 93, 94]. Methylation can also be found in other histone residues, such as H3K36, H3K79 and H4K20. Besides acetylation and methylation, other PTMs can be detected on histones. Some examples are histone phosphorylation, involved in several processes which include DDR [reviewed in 95] and the euchromatin/heterochromatin definition [96] or histone sumoylation, that regulates several processes such as transcription or again the DDR [97].

RNA in the nucleus: RNA regulation of chromatin processes and RNA degradation

Chromatin-associated RNAs

At the same time the mRNAs are synthesized, they undergo a number of co-transcriptional modifications on the chromatin. For a mRNA to be mature, a cap is added to the 5' end, introns are excised and a poly-A tail is added to the 3' end of the transcript. Mature mRNAs are also bound by a plethora of mRNA-binding proteins (RBP) forming mRNPs, including mRNA export factors that will mediate the transport of the mRNP to the cytoplasm for translation. [98]. However, mRNAs represent only 1-2% of the genomic material present in eukaryotic cells and yet, up to 70% of the genome is detected as transcribed RNA [99]. What are the rest of the RNAs being produced? As envisioned by [100], with the development of microarrays and high-throughput RNA sequencing (RNA-seq), many different ncRNA species have been identified. Some examples of these transcripts have been already mentioned in previous sections (see section An overview of eukaryotic transcription). Other ncRNAs produced in eukaryotic genomes are enhancer RNAs (eRNAs), Piwi-interacting RNAs (piRNAs), microRNAs (miRNAs) [101], promoter upstream transcripts (PROMPTS) [102], cryptic transcripts derived from repetitive regions (repRNAs) [103] and damaged-induced long non-coding RNAs (diIncRNAs) that are synthesized at DNA double-strand breaks as part of the DDR [104]. All these types of RNA transcripts have been linked to many different biological processes, most of which occur in the chromatin. There is increasing evidence that many ncRNAs once transcribed are retained and stably associated with the chromatin (reviewed in [105]). Researchers have developed a wide catalog of tools to study the RNAs retained in the chromatin, most of which are based on the use of next
generation sequencing (NGS). The first strategy used was based on standard fractionation procedures. Nuclei were isolated and chromatin was extracted from them. RNA could then be isolated from chromatin and sequenced [106]. Later, different methods have been developed that can be classified into one-to-all and all-to-all strategies. Methods here presented are precise and widely reviewed by Li and Fu in [107]. One-to-all strategies (ChiRP, CHART, RAP) use biotinylated probes against a specific target to purify the chromatin the target binds to. The main limitation of the one-to-all methods is that only one RNA can be interrogated at a time, and the main advantage is that all the interactions are detected irrespective of its abundance in the cell. All-to-all strategies (MARGI, GRID-seq, ChAR-seq) are based on the use of proximity ligation. These methods ligate an adapter to the RNA and to the digested chromatin (digested via either restriction digestion or sonication). The main disadvantage of the all-to-all approaches is their low sensitivity, together with the fact that the methods rely on proximity ligation, which does not ensure a direct RNA-DNA interaction. These methodologies have allowed the identification and characterization of many RNA-chromatin interactions. Most RNA-chromatin interactions are found in the nascent mRNAs [108]. However, many other RNA-chromatin interactions involve ncRNAs transcribed and retained in the chromatin, and increased evidence shows that caRNAs are important for several biological processes [105].

Regulation of chromatin by caRNAs

It has been known for a long time that ncRNAs can modulate the chromatin structure, as shown for instance 20 years ago for the establishment of pericentromeric heterochromatin in Schizosaccharomyces pombe [109] or Xist X chromosome inactivation in humans. In recent years, more and more studies point towards a functional link between caRNAs and biological processes and disease, for instance histone acetylation and transcriptional activation [110] or cancer [reviewed in 111]. Not only caRNAs have been linked to chromatin processes, but they have also been shown to act as structural facilitators, for instance in 3D genome architecture [112,113]. Moreover, RNA and caRNAs have been shown to be a crucial element in the constitution of many phase-separated organelles in the nucleus, for instance paraspeckles [reviewed in 107].

How caRNAs regulate processes on chromatin has been and continues being an extremely active research field. Since most of the caRNAs are nascent transcripts, the most straightforward mechanism by which these transcripts can modulate chromatin organization is via R-loop formation (Figure 2A) or acting via RNA-protein interactions (Figure 2B and Figure 2C) [113,114]. R-loops are formed when RNA invades a double-stranded DNA locus.
and base pairs with one of the strands. This gives as a result a RNA:DNA hybrid that displaces the other DNA strand (not hybridized) as single-stranded DNA. R-loops can be formed in cis at the site of transcription or in trans if the RNA is transcribed elsewhere in the genome [115].

Mechanistically, R-loops can act on the chromatin in many different ways. In humans, R-loops in promoter regions have been associated with transcription. DNA methylases bind poorly to R-loops, thus R-loops prevent CpG methylation usually associated with gene silencing [116]. In mouse embryonic stem cells, it has been shown that R-loops inhibit repressive chromatin modifying enzymes and recruit activating chromatin-remodeling complexes to resolve the proximal pausing of RNApol II, facilitating the transition to productive elongation [117].

R-loops are also formed in genes with a poised/balanced chromatin state. This chromatin configuration is common in the promoter of developmentally regulated genes, and it is characterized by the coexistence of both transcription repression marks (H3K27me3) deposited by the Polycomb repressive complex 2 (PRC2) and activation marks (H3K4me3), deposited by the Trithorax complex [reviewed in 118]. Genes with poised chromatin are repressed, but ready for transcription, and R-loops are important regulators in fine-tuning the balance between the repressive and active state of the chromatin [119]. Excess of unresolved R-loops has been shown to be detrimental for cells, since unresolved R-loops might lead to genomic instability [120].

**Figure 2. Different examples of caRNA regulation of chromatin processes.** A) Nascent mRNA R-loop formation collaborates in RNApol II proximal pause. B) caRNAs can act as scaffolds or recruit
transcriptional activators in cis (nascent mRNA, forming R-loop) or in trans (lncRNA) to help solve the RNAPol II proximal pause. C) LncRNAs have been shown to recruit transcriptional repressors (PcG) to silence actively transcribed genes. D) RNA accumulation in chromatin can lead to eviction of DNA binding proteins, for instance HP1.

Besides base-pairing mechanisms, the mere presence of RNA can alter the chromatin compaction or composition. Dueva and colleagues [121] showed that RNA can compete with DNA for histones’ basic residues (Lysine, Arginine), counteracting the positive electric charge. A similar mechanism was proposed by Schubert et al. [122] in which Df31, a histone binding protein interacts with snoRNAs to form snoRNPs that maintain the chromatin in an open/permissive transcriptional state. RNA not only affects the neutralization of histone tail basic residues. Many proteins found in chromatin can also bind RNA with relatively high affinity, for instance heterochromatin protein 1 (HP1) [123]. Therefore, the presence of RNA in the chromatin causes HP1 binding competition. In cells depleted of the RNA exosome, it has been shown that the compaction of specific repetitive loci decreases. A possible explanation is that HP1 is evicted from the chromatin upon accumulation of RNA transcripts, degraded by the RNA exosome in steady-state conditions (Figure 2D) [124]. Thus, knowing the dynamics of the RNA in chromatin, either through regulation of R-loop formation or resolution, or just by the mere presence of RNA on chromatin proximities, is crucial for our understanding of how the chromatin is regulated in the eukaryotic cells.

The RNA exosome and its functions

From the previous section, it is clear that there is a need to regulate RNA homeostasis on chromatin in order to ensure the correct functioning of eukaryotic cells. This is precisely the RNA exosome endeavor. The RNA exosome is a multiprotein complex present both in the nucleus and cytoplasm of all eukaryotic cells. The RNA exosome consists of a catalytically inactive core of 9 subunits and two different catalytic subunits called Dis3 in fruit flies and mammals (Rrp44 in yeast), and Rrp6 in yeast and fruit fly (EXOSC10 in humans) [125]. The cytoplasmic RNA exosome is composed by the catalytically inactive core and Dis3/Rrp44 and the nuclear RNA exosome is composed by the catalytically inactive core plus both catalytic subunits, Dis3/Rrp44 and Rrp6/EXOSC10 [126,127]. Besides these two main complexes, there has been detected a minor RNA exosome complex in the nucleolus of eukaryotic cells, consisting of the core and Rrp6/EXOSC10 [128]. The proteins constituting the RNA exosome core are organized forming an empty barrel, through which the RNA is channeled, and whose function is to accommodate and incentivize the catalytic subunits.
The subunits with the enzymatic activity are located on top and bottom of the barrel, constituting the 'lids/gates' of the barrel [129].

Dis3/Rrp44 contains 5 different domains, two of which show ribonuclease activity. The N-terminal Pilus-forming domain (PIN) has endoribonuclease activity, and the ribonuclease domain (RNB) has exoribonuclease activity. The RNB domain coordinates two magnesium ions that are required for the $3' \rightarrow 5'$ exoribonuclease activity [130].

Rrp6 has four different domains. The EXO domain catalyzes distributive $3' \rightarrow 5'$ exoribonuclease activity and the HRDC domain (helicase and RNase D carboxy terminal domain) unwinds the possible RNA secondary structures and channels the RNA into the central channel of the RNA exosome complex (structural details reviewed in [131,132]).

When it comes to evolutionary conservation, the RNA exosome is present in all eukaryotic cells, and also in archaea. Protein homology analyses have also found similar complexes devoted to analogous activities in bacteria, the bacterial RNase PH and PNPase. The presence of homologous proteins and the conservation across the entire eukaryotic domain of life suggests that the function of the RNA exosome is ancient and evolutionary conserved [133].

The RNA exosome by itself shows low specificity for its RNA targets. Therefore, it requires a number of associated cofactors that confer specificity to the complex. There are three main known protein complexes that perform this function, the nuclear exosome-targeting (NEXT) complex [134], the polyA exosome-targeting (PAXT) [135] and the Trf4/Air2/Mtr4p Polyadenylation (TRAMP) complex [136]. These complexes guide the RNA exosome to a number of specific RNA species, such as pre-rRNAs, aberrant tRNAs, snRNAs, snoRNAs, and/or cryptic unstable transcripts (CUTs). The NEXT complex targets heterogeneous RNAs lacking polyA tail whereas PAXT complex targets well defined polyA+ RNA transcripts. Irrespective of the specificity, these two complexes show redundancy in their targets. The reason being, most of the RNA species can be processed so that they can be targeted by both complexes. For instance, RNA transcripts lacking a polyA tail can be polyadenylated and therefore, become PAXT targets [137]. Thus, both PAXT and NEXT complexes enhance the RNA exosome activity by recruiting it to its targets. The RNA exosome degradation can only be counteracted by competing for the binding sites of the aforementioned complexes. This is the case of proteins with RNA-stabilizing capacity, for instance of Nab2 in S. cerevisiae [138].

The nuclear RNA exosome has been linked to many different processes in the nucleus, ranging from mRNA quality control, regulation of gene expression or other RNA-degradation related functions such as pre-rRNA processing, transcription termination and DDR (reviewed in [139], further explained below). It has been shown that the RNA stability and susceptibility to degradation is tightly linked to the RNA exosome [140]. As presented above, the mRNA...
production requires a number of co-transcriptional processes and modifications that produce a mature mRNA from the pre-mRNA. The involvement of the RNA exosome in mRNA quality control is supported by several reasons. For instance, the RNA exosome is recruited at genes that fail to terminate transcription at PAS [141]. The RNA exosome is also involved in splicing quality control and it is recruited to transcripts that fail to excise the introns from pre-mRNA. This is shown by the increased levels of pre-mRNA detected in the cytoplasm of cells lacking components of the nuclear RNA decay machinery (either the RNA exosome or the adaptor complexes) [142]. The RNA exosome has also been shown to be enriched in promoter regions of actively transcribed genes, as demonstrated in [143], supporting the idea that the RNA exosome works in mRNA quality control. Another function related to the RNA exosome, is the retention of aberrantly processed transcripts on the chromatin. This function prevents the nuclear export of incorrectly processed pre-mRNAs and has been specifically linked to Rrp6 [144]. Moreover, the nuclear RNA exosome plays a critical role in the degradation of cryptic transcripts tethered to the chromatin, both in yeast and mammalian cells [145].

Several studies have linked the RNA exosome to regulation of gene expression. For instance, in S. cerevisiae, Rrp6 is responsible for the rapid degradation of mRNAs of replication-dependent histones when cells exit the S-phase [146]. Another study by Wu and colleagues [147] showed that depletion of one or several RNA exosome components leads to the production of exosome-sensitive transcripts, out of which at least a thousand were found to overlap annotated mRNA, fine-tuning the expression of protein-coding genes. The RNA exosome has been involved in DNA damage repair and genome integrity. In DNA double-strand breaks, transcription takes place in the neighborhood of the lesions. The RNA exosome is required to remove the RNA produced to allow DNA end resection and the binding of the replication protein A (RPA), later resulting in homologous recombination repair [148]. As mentioned above, unresolved R-loops lead to increased genomic instability. The link of the RNA exosome to genome integrity was first described by Richard and colleagues [149] who proposed that the RNA exosome cooperating with Senataxin, was in charge of R-loop resolution at DNA-damaged loci.

Another function attributed to the nuclear RNA exosome is the maturation process of several types of ncRNAs. In fact, the RNA exosome was first identified as a protein required for the 3’ end processing of rRNA. 3’ ends of multiple snRNAs and snoRNAs are also trimmed by the RNA exosome (reviewed in [150]).

The cytoplasmic RNA exosome has been mainly implicated in mRNA surveillance (quality control), turnover of mRNA transcripts and immunity. The RNA exosome and Xrn1, a 5’ → 3’ exoribonuclease, are the two main RNA degradation pathways in the cytoplasm. The RNA
exosome has been shown to be recruited to non-functional mRNAs lacking termination codons (non-stop decay, NSD) or to mRNAs with stalled ribosomes (no-go decay, NGD, reviewed in [151]), whereas mRNAs that contain premature stop codons trigger the nonsense mediated decay (NMD) pathway that is mainly handled by Xrn1, the 5' → 3' degradation machinery [152].

Lastly, the cytoplasmic RNA exosome complex has been linked to antiviral response in flies, demonstrated by the fact that the depletion of several subunits of the TRAMP complex resulted in increased levels of viral particles and toxin production [153].

Heterochromatin in eukaryotic genomes: Formation and regulation

Constitutive heterochromatin

Not only activation of gene transcription is important for living organisms. Accurate transcription silencing of genes is also an essential component of transcriptional programs. The silenced regions of the genome that are not expressed are referred to as heterochromatin. There are two different types of heterochromatin in eukaryotic cells. The constitutive heterochromatin located in the centromeric and telomeric regions of the chromosomes, more related to a structural role; and the facultative heterochromatin, that corresponds to genes that are kept silent during development (reviewed in [154]) or in specific cell types. The two types of heterochromatin have specific signatures and are established and maintained by the involvement of different cellular processes and proteins.

The constitutive heterochromatin is characterized by being heavily decorated with H3K9me2/3 and bound by HP1. The H3K9me2/3 modification is catalyzed by the SET domains of the histone-lysine N-methyltransferases (KMT) [155]. In D. melanogaster, it is catalyzed by the protein suppression of variegation 3-9 or Su(var)3-9 [156]. SUV39H1/2 and Clr4 are the human and yeast orthologs, respectively. However, the mechanisms of establishment and maintenance of this chromatin mark are not fully understood. In fission yeast, the establishment of H3K9me3 is dependent on transcription and the RNAi mechanism. Low levels of transcription arising from pericentromeric loci generate nascent RNAs that can directly be used to recruit silencing factors [109]. Transcription of heterochromatin repeats gives rise to rep-RNAs that will be bound by Argonaute protein (Ago1) and used to target the same repetitive loci these transcripts have been transcribed from. Ago1 is part of the RITS complex, which associates with the RNA-dependent RNA
polymerase complex (RDRC) and the Clr4 protein [157]. Clr4 KMT is the responsible of depositing the histone PTM and RDRC transcribes the second strand of the pericentromeric transcripts, allowing the amplification of the siRNAs, thus establishing a positive loop in order to spread the H3K9me3 marks to the neighboring histones. This positive forward loop ultimately results in the establishment of the centromeric heterochromatin [109]. Once H3K9me3 is established, it is recognized and bound by HP1, which contacts two adjacent nucleosomes and induces a change in the chromatin conformation. This conformational change translates into increased chromatin compaction (less accessibility for TFs) which constitutes the basic principle to minimize transcription at heterochromatin loci (reviewed in [158]).

Pericentromeric heterochromatin is characterized by low gene density, high density of repetitive elements and low levels of transcription. It was thought to be a transcriptionally inactive region of low functional significance. Today, we know that this simplistic vision was wrong. The centromeric and pericentromeric heterochromatin have been shown to be extremely important for chromosome segregation [159]. It has also been shown that the centromeric chromatin contacts other heterochromatic regions embedded in euchromatic regions, which is important for the 3D structure organization of the chromatin and genome integrity [160]. As illustrated by the mechanism of heterochromatin formation, transcription and the presence of RNA is crucial for heterochromatin dynamics and thus, the regulation of the caRNAs in heterochromatic loci is key to ensure proper chromatin organization.

Facultative heterochromatin - An overview of Polycomb repression

Despite having the same ultimate goal as constitutive heterochromatin, transcriptional repression, the facultative heterochromatin is established through different mechanisms and exhibits specific characteristics. This type of heterochromatin is embedded in euchromatic regions and is characterized by high gene density. These regions usually harbor the genes conferring cell-type characteristics and developmentally regulated genes. While the constitutive heterochromatin is decorated with H3K9me2/3, the facultative heterochromatin is instead enriched in H3K27me3. This chromatin mark is deposited by the Polycomb repressive system (PcG), which comprises two different protein complexes, the Polycomb repressive complexes 1 and 2 (PRC1 and PRC2) [161].

PRC1 monoubiquitylates histone H2A at Lys119 (H2AK119ub) and PRC2 methylates H3K27, from mono to trimethylation. The classical molecular mechanism described for PRC1 and PRC2 is a collaboration between these two complexes in order to repress their target genes [162]. This classic view comes from the Polycomb regulation of the homeotic
(Hox) genes during *D. melanogaster* development. The Hox genes constitute a family of TFs that are spatiotemporally expressed to control the development and patterning of *D. melanogaster* morphology (reviewed in [48]). Despite the early research in PcG regulation was carried out in *Drosophila*, it has been demonstrated that the PcG and its molecular functions are evolutionarily conserved in mammals [163]. Interestingly, the PcG is present in some unicellular eukaryotes, but has been lost in *S. cerevisiae* or *S. pombe* [164]. Besides H3K27me3 and the H2AK119ub modifications, many of the PcG regulated genes are also enriched in a specific active transcription modification, H3K4me3. This seemingly counterintuitive chromatin conformation was discovered in 2006 in mouse embryonic stem cells [165]. This study described the co-occurrence of H3K4me3 and H3K27me3 in the promoters of developmentally regulated genes and termed this chromatin conformation as bivalent chromatin (Figure 3A, next page). Later studies further developed this idea, and soon the bivalent chromatin conformation was observed also in human embryonic stem cells, embryonic epiblast and human T-cells [166,167]. Importantly, this co-occurrence happens in the same genomic loci [168]. Besides mice and humans, the presence of bivalent chromatin domains has been found in several other organisms, for instance, avians, planarians, zebrafish or *Drosophila* [169]. H3K4me3 is deposited by the Trithorax-group (TrxG), and is thought to be counteracting the repressive action of PcG. The following sections will describe several aspects of the bivalent chromatin states, PcG and TrxG gene regulation, including who are the members of the different Polycomb complexes, how these complexes are recruited, the chromatin signature of the target genes and the regulation of the PcGs.
Figure 3. Bivalent chromatin conformation. A) Bivalent chromatin conformation in developmentally regulated genes. Promoters are decorated with activating histone modifications (H3K4me, green star) deposited by TrxG (green blobs) and repressing histone modifications (H3K27me, red star) deposited by PcG (blue blobs). The transcriptional steady-state in this equilibrium is repression (left part of the figure). The equilibrium can be broken by several mechanisms that recruit or evict one of the two complexes leading to heterochromatin formation and repression (top) or to activation and transcription (bottom) B) Possible role hypothesized for the RNA exosome in the regulation in bivalent states. The absence of the RNA exosome catalytic subunits can generate changes in caRNA levels. Because of PcG to bind RNA, it can potentially be evicted from chromatin by accumulated caRNAs, leading to derepression of target genes (see paper I).

Polycomb repressive complexes - Who?

The human PRC2 complex is formed by a core of four proteins: enhancer of Zeste 2 (EZH2) and its paralog EZH1, embryonic ectoderm development, Suppressor of Zeste 12 (SUZ12), and retinoblastoma-binding protein 4 (RBBP4) or RBBP7. Structural analyses show that PRC2 is organized into a catalytic domain and a regulatory domain, bridged by SUZ12 [170]. EZH1/2 possesses the catalytic KMT activity of the complex. Upon SUZ12 binding, a conformational change is induced in EZH1/2, transitioning from an autoinhibitory conformation into a productive KMT conformation [171]. In Drosophila, the PRC2 core
components are E(z), Extra-sex combs (Esc), Su(z)12 and Caf1-55. The core proteins are complemented by a plethora of accessory proteins, forming different subcomplexes that render different specificity and targets [172].

In Drosophila, the canonical PRC1 core consists of Polycomb (Pc), the protein that recognizes and binds the H3K27me3 (methylated by PRC2); Polyhomeotic (Ph), consisting of two paralogues (proximal and distal, Ph-p and Ph-d respectively), RING1/Sex comb extra (Sce) and Posterior sex combs (Psc). Sce and Psc form a heterodimer that enhances Sce E3 ubiquitin-ligase activity on histone H2A [173,174]. PRC1 also has a wide catalog of accessory proteins and forms different subcomplexes. For instance, the dRING-associated factors (dRAF) complex has been proposed as the responsible for most of the ubiquitination in Drosophila [175]. Homologous proteins to the aforementioned ones have been found in mammals, many of which are present as different paralogues (for an extensive review, see [176]).

Moreover, besides accessory proteins that regulate PcG activity, there are a number of accessory complexes that modulate PcG target recognition. In Drosophila, PcG target recognition is achieved with a specific conserved DNA sequence, known as Polycomb Response Elements (PREs, see next section). The PhoRC complex is composed of two different proteins; Pleiohomeotic (Pho) and Sfmbt. Pho is able to recognize PREs, and is thought to help with the recruitment of Polycomb complexes to their targets [177].

Polycomb repressive complexes - Where?

Target recognition is perhaps the most divergent aspect of the PcG repression between Drosophila and mammals. In Drosophila, as previously mentioned, the recognition of target genes by PcG and auxiliary complexes is achieved through the PRE elements [178]. PRE elements are usually located close to the TSS of the gene to be repressed. However, PREs can also be found up/downstream of the promoter, in introns of genes and even several thousand base pairs away from the promoters [179]. PREs consist of a set of consensus sequences among which it is possible to find binding motifs for Pho, Trithorax-like (Trl; also known as GAF), Dorsal switch protein 1 (Dsp1) and other DNA-binding factors [180]. Machine learning has also been utilized to predict and further characterize the different PRE motifs found in Drosophila [181]. The fact that Trl recognizes PREs suggests a possible explanation on how bivalent chromatin domains are formed.

PREs recruit all the different PcGs. However, it still remains elusive which protein or proteins might recruit them to these DNA motifs. Only Pho, contained in the Pho-RC auxiliary complex, contains a DNA binding domain that recognizes PREs, and it is proposed as one of
the proteins devoted to PcG recruitment. Supporting this idea, in vitro experiments confirmed that Pho could directly interact with E(z) and Esc [182,183]. However, it cannot efficiently target the different complexes on its own. Despite not knowing the precise recruitment mechanisms, PREs are believed to be essential for PcG repression. It has been shown that depletion of PREs in fly larvae resulted in the dilution of H3K27me3 modification through divisions, leading to de-repression of Hox genes in wing imaginal discs [184]. This study supports the view that PRE elements are required for H3K27me3 transcriptional repression, and that H3K27me3 chromatin modification alone is not able to self-maintain the transcriptional repression.

In humans however, no PRE has been identified so far and the recruitment of the different PcGs is suggested to occur via DNA sequence and composition. PcGs are thought to be recruited by CGIs that normally overlap with the promoter of the developmentally regulated genes [185,186]. CGIs are located closely to the TSS of the genes, but span a relatively broad region (1-2 Kb). PcG related CGIs are characterized by low levels of CpG methylation in opposition to CpG dinucleotides in the rest of the promoters, where CGIs are heavily methylated to restrict transcription outside the genes bodies [186]. Supporting the idea that recruitment is not sequence-dependent, Lynch and collaborators [187] demonstrated that bacterial-derived GC-rich sequences substituting CGIs could recruit PcGs in mouse embryonic stem cells.

It has been demonstrated that PcGs can also be targeted by RNAs. Davidovich and colleagues [188] demonstrated the ability of the PRC2 complex to promiscuously bind RNA in vitro, opening the possibility for a different recruitment mechanism for PRC2. This recruitment mechanism can be exemplified with several lncRNA, including HOTAIR, ANRIL or NEAT1 (reviewed in [189]). In fact, the interactome of PRC2 has been studied in vivo with RIP-seq and it has been demonstrated that PRC2 associates to thousands of RNA transcripts [190], supporting the in vitro results obtained by Davidovich et al. Moreover, in humans, transcription of short ncRNAs has been detected in CGIs and it has been proposed that these short non-coding transcripts could regulate the function of PRC2 [191]. In the last few years, several reports have described the ability of R-loops to regulate Polycomb activity. However, there is not a consensus yet on whether R-loops promote [192,193] or inhibit [117] Polycomb activity.

Thus, because of the ability of PcGs to bind RNA, the RNA exosome could play a potential regulatory function of the Polycomb repressive complexes (Figure 3B). Deregluation of polycomb repression has been long observed in model organisms such as Drosophila and there is an increasing interest in Polycomb repression deregulation, not only from a developmental perspective, but also in a number of human diseases, including cancer [194].
Polycomb repressive complexes - How?

It is widely accepted that Polycomb repression is the default state (steady-state) of developmental genes, and TrxG activity is required to counteract the Polycomb repressive action [195]. As introduced before, both activities co-exist on chromatin in a determined chromatin conformation denominated bivalent chromatin state. How the fine balance of repression/activation is established and/or resolved is a biological question that has been widely studied and yet, not fully understood.

How can Polycomb function repress transcription in developmental genes? The most widely studied mechanisms of Polycomb repression involve H3K27me3, linked to the PRC2 complex, and the H2AK119ub, associated with the PRC1 complex. However, additional activities are required for optimal gene repression. These activities are provided by the auxiliary complexes, and a clear example of this is the recruitment of H3K4 demethylase activity (KDM1A/LSD1), that counteracts the Trithorax H3K4me3 deposition [196]. The ultimate goal is to impede the access of the transcription machinery to the repressed genes, hence one of the main mechanisms proposed for Polycomb regulation is the modification of chromatin compaction. The increase of chromatin compaction allows exclusion of chromatin remodeling factors such as SWI/SNF [174] and extrudes TFs from promoter regions. There are several mechanisms by which this is accomplished. For instance, it has been proposed, in both Drosophila and mice, that the PRC1 proteins Psc and CBX2 ([197,198] respectively) increase the chromatin compaction thanks to the great number of basic residues in their sequence, directly interacting with linker DNA to bring nucleosome closer to each other.

Another mechanism is that H3K27me3 and H3K27ac, considered a hallmark of active transcription, are both deposited in Lysine 27 and cannot coexist. On top of substrate competition (H3K27), PRC2 also inhibits the histone acetyl-transferase catalytic activity of CBP that deposits the H3K27ac [199]. Moreover, a beautiful study by Pengelly and collaborators showed that H3K27me3 is essential for repression of developmentally regulated genes by mutating the Lysine 27 to Arginine of the H3 histone and reproducing the same phenotypes generated by PRC2 mutations [200].

Another interesting question about PcG repression is how the heterochromatin from PRE/CGIs spreads into the gene bodies. The answer to this question seems to rely on the ability of the different complexes to act both as ‘readers’ and ‘writers’, and the cooperativity between complexes. For instance, PRC2 complex contains both activities. As mentioned before, E(z) (Ezh1/2) is a histone-methyltransferase and Esc can bind H3K27me3. Thus, upon deposition of first H3K27me3 in the PREs/CGIs, a positive feedforward loop is established, where PRC2 docks H3K27me3 and spreads the modification further to the neighbor nucleosome [201]. A clear example of the cooperativity between PRC1 and PRC2...
is shown in [202]. This study demonstrated that tethering of PRC1 was sufficient to drive de novo enrichment of PRC2, accumulation of H3K27me3 and the formation of a Polycomb chromatin domain that spreads several kilobases from the original deposition site. All in all, PRC1 and PRC2 regulate the formation of heterochromatin through many different mechanisms, which act in a cooperative manner to ensure the PcG gene repression.

Post-transcriptional regulation of gene expression

Post-transcriptional processes and RNA stability

Transcription and degradation are two processes that dictate RNA abundance. Once mRNA has been transcribed, many biological processes are required to ultimately generate a functional protein. The regulation of these processes will dictate the life, degradation or translational output of the mRNAs. We refer to these processes as post-transcriptional mRNA steps, and include: mRNA processing, nuclear export, subcellular transport or localization, translation, and mRNA turnover.

The very first post-transcriptional process is pre-mRNA processing. The pre-mRNA has to undergo a number of chemical modifications in order to produce a functional mRNA. The main modifications are addition of 7-methylguanosine cap (capping), intron removal (splicing), and cleavage and addition of a poly-A tail to the 3’ end of the transcript (polyadenylation) [203]. To a certain extent, these modifications occur cotranscriptionally, as shown for capping and splicing [203]. At the same time, the mRNAs being transcribed are quickly bound by a number of RBPs which impacts the mRNA maturation. The process is concluded with the packaging of the mRNA into a mRNP that contains the required proteins for mRNA export to the cytoplasm (reviewed in [204]). The mRNP export is achieved by passing the mRNPs through the nuclear pore complex, and the proteins that form the mRNP can influence the export dynamics.

Once the mRNA has reached the cytoplasm, there are four possible scenarios: 1) it can be directly translated by the ribosomes, 2) it can be stored in cytoplasmic bodies, 3) it can be transported to a specific subcellular localization or, 4) it can be flagged for degradation. The way the fate of the mRNA is decided is encoded by the proteins that form the mRNP, which in turn is dictated by the mRNA sequence and the relative abundance of RNA-binding proteins in each particular cell type. For instance, in neurons it is extremely common that specific mRNPs containing mRNAs required for synapse and axonal homeostasis are actively transported to the axons where they are locally translated (reviewed in [205]). Another possibility is that the mRNP is localized to a cytoplasmic membrane-less organelle.
This is the case of P-bodies or stress granules, that are liquid-liquid phase separated compartments in which the mRNPs are targeted for 'long-term storage' and degradation, respectively [206].

The importance of mRNA degradation in RNA abundance comes from the observation that the transcription rate does not directly correlate with the abundance of mRNA, suggesting that mRNA degradation is regulated [207]. Interestingly, the range of mRNA half lives is very broad. In fact, mRNAs of housekeeping genes are generally more stable (longer half life) than mRNAs encoding regulatory proteins [208]. Moreover, mRNA half lives are organism-specific, as described by Chen and collaborators [209] and can differ across tissues.

Different post-transcriptional processes influence cytoplasmic mRNA degradation, translation itself being one of them. In fact, there are several lines of evidence suggesting a tight regulation and cross-talk between the translation and RNA degradation machineries (reviewed [210]). The mRNA is looped while being translated by the ribosome, so that the 5' (cap) and the 3' (poly-A) ends are in close proximity. This process has been called 'the closed loop model', and it is required for effective translation [211]. When the mRNA adopts such conformation, it is protected from exoribonucleases, since these proteins cannot access either of the mRNAs ends [212]. Translation can also negatively influence mRNA stability by targeting the mRNA for degradation. This mechanism is part of the mRNA quality control, and occurs when ribosomes are stalled on non-productive mRNAs [213]. Faulty mRNAs trigger NMD, NSD or NGD (see section The RNA exosome and its functions). In these degradation pathways, the ribosome acts as a sensor, along with a plethora of specific factors, triggers the decay of the problematic mRNA either via 5'→3' (Xrn1) or 3'→5' (RNA exosome, extensively reviewed in [214]).

Another type of post-transcriptional process is chemical RNA modification. There are over 170 detected RNA modifications in the eukaryotic transcriptome [215]. RNA modifications were discovered over 50 years ago with the sequencing of the first biological RNA, the yeast Alanine tRNA. RNA modifications have been widely studied for tRNAs and rRNAs [216]. However in recent years, attention has shifted towards mRNA modifications that constitute the so-called epitranscriptome. The development of specific methods to study RNA modifications and third generation sequencing (Oxford nanopore), has been crucial to study mRNA modifications [217]. The new data suggest that mRNA modifications are relevant for many biological processes and when deregulated, can contribute to diseases such as cancer [218]. Some of the most studied RNA modifications are N^6-Methyladenosine (m^6A), N'-Methyladenosine, 2'-O-ribose methylation, 5-Methylcytosine, 5-Hydroxymethylcytosine, pseudouridylation or RNA editing (see more below). RNA modifications can impact mRNAs by altering their charge, secondary structure, and/or modifying the protein-RNA interactome.
(reviewed in [219]). An RNA modification that has been extensively studied is m^6A, which interferes with the secondary structure favoring unstructured transcripts and in this way predisposing transcripts for protein recognition (HNRNPC and HNRNPG) [220]. Another RNA modification, and relevant for this thesis, is adenosine-to-inosine (A-to-I) RNA editing. Adenosine is deaminated giving as a result a different base, called Inosine. Inosine chemical structure is more similar to Guanosine, which can impact the RNA secondary structure, the RNA-protein recognition pattern and its coding during translation (further discussed in next sections).

ADAR family of enzymes and Adenosine-to-Inosine RNA modification

A-to-I RNA editing

RNA editing was discovered in the cytochrome oxidase coxII transcript of Trypanosoma mitochondria, which contained four uridines not encoded by the genomic DNA [221]. Later, this type of post-transcriptional modification was also found in mammalian mRNAs [222]. RNA editing includes the following different RNA modifications: Cytosine-to-Uracil substitution editing; Uracil insertion/deletion editing; Cytosine insertion editing; Guanosine insertion editing and A-to-I substitution (reviewed in [223]). A-to-I substitution occurs in double-strand RNA (dsRNA) structures and is widely spread across metazoans [224]. This enzymatic reaction is catalyzed by the adenosine deaminases acting on RNA (ADARs, see next section) [225].

Inosine is chemically similar to Guanosine. This implies that Inosines in the mammalian transcriptome are recognized by the cellular machinery as Guanosines and not as Adenosines. This opens many possibilities in RNA-protein and RNA-RNA mediated recognition. An important consequence of A-to-I editing is amino acid substitution, known as recoding editing. The classic example of recoding is the change of Glutamine by Arginine (Q/R) of the AMPA receptor subunit GluR-B [222]. This protein is a subunit of the AMPA receptor, whose mRNA (Glu2A) in physiological conditions is edited virtually in 100% of the mRNA molecules. This A-to-I editing event implies a Q/R amino acid change in the synthesized protein, and the recoding event makes the AMPA channel impermeable to calcium [226]. Very rarely transcripts show 100% of RNA A-to-I editing, and the extent of editing widely varies across transcripts. In vitro and in vivo editing efficiencies are on average 50% and 16% respectively [224]. In the human transcriptome, the vast majority of the editing events are detected in the transcripts arising from Alu retrotransposons and in
introns [227]. Besides recoding, A-to-I editing affects RNA metabolism in different ways, by regulating alternative splicing and promoting different transcript isoforms [228], altering RNA secondary structure changing the protein interactors of specific mRNPs [229], and changing RNA-RNA interactions, for instance through miRNA retargeting [230].

In summary, A-to-I RNA editing can impact the transcriptome at several levels, not only by recoding the protein encoded by the edited mRNA, but also altering the mRNA interactions with other molecules present in the cells.

ADAR proteins catalyze A-to-I RNA editing

ADAR1, ADAR2 (also known as ADARB1) and ADAR3 (ADARB2) form the ADAR protein family. All three proteins are structurally very similar to each other [225], but it is ADAR1 and ADAR2 the proteins that catalyze A-to-I RNA editing while ADAR3 is catalytically inactive. All three proteins possess one deamination domain. However, the ADAR3 deamination domain carries five amino acid substitutions that result in ADAR3 inability to stabilize the substrate in the catalytic site, leading to its catalytic inactivation (Figure 4A) [231]. The three proteins also differ in the RNA binding domains. All three proteins contain dsRNA binding domains, but in different number. ADAR1 contains three dsRNA binding domains and ADAR2 and ADAR3 have two. ADAR1 is present in two different isoforms in mammalian cells that contain one or two Z-DNA binding domains, respectively. The isoforms are known as ADAR1-p110 and ADAR1-p150 [232]. ADAR3 binds both dsRNA (DRBM domain) and single stranded RNA (ssRNA), via the Arginine-rich domain (R-domain) [233]. The differences in the RNA-binding domain dictate the target specificity (reviewed in [234]), which are dsRNA molecules with a length ranging from 20 bp to 100 bp [235].

Besides structural particularities, the ADAR enzymes also show different expression profiles in mammals. In humans, ADAR1 and ADAR2 are expressed ubiquitously. However, the expression levels widely differ across cell types, and ADAR1 and ADAR2 typically show elevated levels of expression in the brain and vascular system. In contrast, ADAR3 is expressed primarily in the brain and germ line, at a relatively low abundance [236]. The ADAR enzymes are thought to have evolved from gene duplication events [237] and are conserved across different organisms, with slight differences. In D. melanogaster, there is only one member of the ADAR family, and two in C. elegans. Interestingly, in C. elegans one of the ADAR members (ceADR1) is inactive and is thought to regulate the active member of the family (csADR2, see discussion below) [238].
Figure 4. ADAR protein family, structure and function. A) Different members of the ADAR family of RNA modifying enzymes. ADAR1 is present in two different isoforms. Red squares symbolize the deaminase domain. ADAR3 red cross denotes the amino acid substitutions in the deaminase domain that have inactivated the enzyme. B) ADAR proteins function. ADAR1,2 catalyze deamination of adenines (A) into inosines (I, pink). ADAR3 has been proposed to inhibit A-to-I editing by substrate binding competition with ADAR1 and ADAR2.

In humans, ADAR1 is responsible for most of the A-to-I editing detected. However, the biological relevance of ADAR1 is not completely understood. ADAR1 has been associated with untargeted hyper-editing, probably because it is a highly and ubiquitously expressed enzyme [239]. In fact, in the brain where ADAR2 is more expressed than in other cellular types, it appears to outperform A-to-I editing catalyzed by ADAR1 [240]. ADAR1, and in particular the ADAR1-p150 isoform, seems to be pivotal for immune system regulation. Several studies suggest the ability of ADAR1-p150 to downregulate interferon gamma by preventing melanoma differentiation-associated protein 5 (MDA5) and mitochondrial antiviral signaling proteins (MAVS) from triggering an immune response [241]. MDA5 is a dsRNA sensor activated by the presence of dsRNA, and it has been proposed that ADAR1 could play an important role in differentiating host dsRNA from exogenous (viral) dsRNA [242].

ADAR2 instead has been shown to be critical in the nervous system (reviewed in [243]). In concordance with this statement, it has been observed that ADAR2 deficient mice die three weeks after birth due to overstimulated neuronal activity [244]. Thus, ADAR2 is essential for mice viability. The molecular mechanism underlying this lethal phenotype is the lack of editing in the Glu2A mRNA. While this phenotype is drastic, there are several other pieces of
evidence of a more subtle A-to-I editing regulation. For instance, the voltage-dependent potassium channel Kv1.1 or the calcium channel Cav1.3. Mutations of RNA editors lead to abnormal brain development and a plethora of neurological diseases (reviewed in [245]).

ADAR3 and its unknown molecular function
ADAR3 was identified over twenty years ago in a screening with probes against the deamination domain to identify more proteins of the ADAR family [233]. This protein however, has historically attracted less attention than ADAR1 and ADAR2. A possible explanation for this fact is that ADAR3 lacks catalytic activity. Also, ADAR3-KO mice exhibit milder phenotypes such as stress-related and impaired memory formation [246]. Contrarily, ADAR1 and ADAR2 present lethal phenotypes. In recent years, several studies have tried to address the biological function of ADAR3. By homology with C. elegans (see above), it has been hypothesized that ADAR3 could regulate the A-to-I RNA editing exerted by ADAR1 and ADAR2 by target competition or by formation of inactive heterodimers [247]. There are some studies supporting this idea. For instance, Oakes and collaborators demonstrated the existence of RNA editing competition for Q/R site of Glu2A mRNA in overexpressed ADAR3 cells [248] and a global study of RNA editing regulators showed a negative correlation between ADAR3 levels and RNA editing [249]. Moreover, a recent study showed that ADAR3 binding to the 3'UTR of the MAVS transcript inhibits the RNA editing reaction [250].

In this study, the researchers managed to generate variable degrees of RNA-binding deficient ADAR3 proteins, and they observed that the RNA-binding deficient ADAR3 resulted in elevated RNA-editing levels in the MAVS transcript. Despite the elevated levels of RNA-editing, the MAVS protein expression was also found upregulated in the RNA-binding deficient ADAR3, suggesting a RNA-independent ADAR3 regulation. These examples presented here are the exception to the rule, since ADAR3-KO mice do not show global alterations in the editome levels. A plausible explanation for it might rely on the fact that the expression of ADAR3 is not ubiquitous, but confined to specific neuronal cell types [236,246,251]. To further characterize ADAR3 biological function and its binding partners, Wang and colleagues [231] genetically engineered a catalytically active ADAR3 enzyme. Utilizing this approach they were able to show that ADAR3 has binding preference for the 3'UTR of the transcripts. Interrogating the newly detected A-to-I editing events, they found a genomic position that overlaps a miRNA binding site in the EGR1 transcript, suggesting that ADAR3 could actively compete with miRNA for the same targets. Indeed, overexpression of ADAR3 led to five-fold upregulation of the EGR1, which could not be replicated by ADAR2 overexpression [231].
Besides the potential as A-to-I RNA editing regulator, ADAR3 has been linked to several other biological processes. For instance, ADAR3 has been proposed to enhance circular RNA-formation from the NT5ET transcript. This transcript functions as a miRNA sponge in Glioblastoma cells, acting as a tumor-suppressor mechanism [252]. ADAR3, through interaction with the C9ORF72 hexanucleotide expansion, has also been linked to frontotemporal dementia and amyotrophic lateral sclerosis [253].

Taken all the observations together, it seems that ADAR3 might play a role in A-to-I RNA editing regulation. However, the low severity phenotypes, the fact that ADAR3 is only expressed in highly-specific cell types and the links between ADAR3 and several biological processes in an editing independent manner suggest that ADAR3 cannot be a general regulator (inhibitor) of A-to-I RNA editing.

**SnoRNAs and their functions: Post-transcriptional RNA modifiers?**

**snoRNAs biochemical properties and classification**

The snoRNAs are ncRNAs that usually range between 50 and 300 nucleotides in length, highly structured, and enriched in the nucleoli of eukaryotic cells (Figure 5A). SnoRNAs were early described to act as guides for the modification of rRNAs and snRNAs, which is considered to be their canonical function [254]. However, only a few snoRNA species are devoted or have been directly linked to rRNAs or snRNAs maturation. Many other annotated snoRNAs remain without an associated function and are known as orphan snoRNAs. SnoRNA genes are mainly harbored in intronic regions of both protein coding and non-protein coding genes [255] or in fewer cases, as independent transcriptional units (Figure 5B, top panel). SnoRNAs are classified into two different groups according to their sequence and structural features. These groups are C/D box snoRNAs and H/ACA box snoRNAs [256] (Figure 5A, left and right panel respectively). Eukaryotic C/D-box snoRNAs usually range from 70 to 120 nucleotides and contain two conserved sequences, the C and D box. The C box consists of the consensus sequence RUGAUGA located near the 5’ of the molecule. The D box instead, is formed by the consensus sequence CUGA, and is located in the 3’ end of the snoRNA. C/D-box snoRNAs acquire a specific secondary structure named kink-turn [257]. This structure is recognized and bound by Snu13p which recruits Nop1p (or Fibrillarin in *Drosophila*), Nop58p and Nop56p. Together, C/D box snoRNAs and aforementioned proteins constitute a complex called C/D box snoRNPs, whose function is to mediate 2’O-ribose methylation of specific residues [258].
H/ACA snoRNAs are slightly shorter in size, ranging from 60-75 nucleotides in length. These snoRNAs form a specific secondary structure called pseudouridylation pockets, where the pseudouridylation reaction occurs and RNA uridine residues are isomerized [259]. Similar to C/D box snoRNAs, H/ACA snoRNAs are bound by a specific subset of proteins to form snoRNP complexes. The proteins associated with H/ACA snoRNAs are Cbf5p (Dyskerin in Drosophila), Nop10p, Gar1p, and Nhp2p; among which Dyskerin/Cbf5p is the protein that catalyzes the pseudouridylation reaction (reviewed in [260]). Eukaryotic H/ACA box snoRNAs contain two conserved sequences: the H box and the ACA box, which are located downstream of the first and second hairpins, respectively [261]. There is a third group of snoRNAs referred to as small Cajal body RNAs (scaRNAs) based on their very specific localization. ScaRNAs are a specialized type of snoRNA located at Cajal bodies, and are thought to modify snRNAs [262].

The snoRNA functions, more than post-transcriptional RNA modifiers

As presented in the previous section, canonical snoRNAs contain sequences that are complementary to specific RNA targets and act as guides for RNA modification. The sequence complementarity is crucial for guiding the snoRNP to the correct sequence in the target RNA, mostly rRNAs [263] (Figure 5C). Besides 2'O-methylation and pseudouridylation, there is an immense variety of RNA modifications detected in eukaryotic RNAs [215]. In recent years, there has been increased evidence that snoRNAs can guide RNA modifications other than the two canonical RNA modifications aforementioned. In particular, N4-acetylcytidine (ac4C) modification, which can be found on rRNAs, tRNAs and mRNAs, can be guided by snoRNAs [264]. Bioinformatic tools have been developed to predict snoRNA targets [265] but yet, many snoRNAs do not seem to be complementary to any other transcript and have no associated function. A general assumption is that orphan snoRNAs do not exert canonical functions and their functions have yet to be interrogated individually. Recent data shows that snoRNAs are involved in a wide variety of biological processes. Among these processes, snoRNAs have been linked to cancer, disease in which snoRNAs levels have been observed to be dysregulated. SnoRNAs have been considered both as tumor-suppressing [266] and tumor-inducing [267,268] (for extended review of snoRNAs implications with cancer, see [269]). SnoRNAs expression has also been shown to be affected in viral infection [270] and latest evidence suggests that snoRNAs are highly enriched in chromatin [271,272].
SnoRNA regulation of biological processes can be found in particular studies. For instance, SNORA73 forms a complex with PARP1 and DKC1/NHP2 to inhibit PARP1 auto-PARylation to affect cancer genome stability [273]. SnoRNAs can also regulate alternative splicing. This is the case of HBII-52/SNORD115, which has been found to bind the exon Vb of the 5-HT2C receptor and to regulate alternative splicing [274]. Reduced levels of SNORD27, present in...
nuclear snoRNPs lacking Fibrillarin, are associated with reduced levels of alternative exon skipping [275].

SnoRNAs have recently been associated with miRNA-like functions. MiRNAs can silence gene expression by binding to specific sequences in the 3'UTR regions of the mRNAs, triggering the transcript decay [276]. A particular example of this behavior is the C/D box snoRNA HBII-180C. This snoRNA forms a snoRNP with fibrillarin. It contains an M-box region, used to recognize its targets and induces the transcript knockdown through a silencing mechanism that is similar to that triggered by miRNAs [277]. An interesting piece of evidence suggesting that snoRNAs could have non-canonical functions on chromatin comes from experiments that map RNA-DNA interactions. Bell and colleagues [271], developed ChAR-seq to analyze RNA-DNA interactions genome-wide in Drosophila melanogaster. They observed that several snoRNAs were among the top ranked RNAs detected on chromatin. Interestingly, an early study by Schubert et al. [122] proposed that snoRNAs can interact with histone tails and contribute to the establishment of snoRNA-chromatin network that promote open chromatin domains. Placing all available pieces of evidence together, it is highly probable that snoRNAs can regulate chromatin organization or mRNA abundance through many different molecular mechanisms that are still to be fully understood.
Chapter 2: Aims, Results, Conclusions and Future Perspectives

The main objective of this thesis has been to expand the knowledge about different mechanisms that impact gene expression. More specifically, I have addressed this broad question at two different levels: 1) at transcriptional level by investigating the role of chromatin-associated RNAs (caRNAs) in Drosophila melanogaster, their regulation by the RNA exosome and the biological relevance of RNA degradation on chromatin, and 2) at post-transcriptional level by studying the editing-independent functions of ADAR3. The main focus has been the characterization of caRNAs on Drosophila melanogaster S2 cells chromatin, and how the transcriptional landscape is shaped on chromatin by the RNA exosome.

To tackle the questions presented above, this thesis has been subdivided in three main aims:

- Characterize chromatin-associated RNAs in Drosophila melanogaster. Understand the turnover control that the RNA exosome exerts on caRNAs, and the possible biological consequences of such control (Paper 1).

- Disentangle the role of snoRNAs in controlling chromatin organization and gene expression using a viral infection model (Paper 2).

- Establish the developmental spatio-temporal expression pattern of ADAR3 in developing mice brain and elucidate the RNA-editing dependent and independent functions of ADAR3 (Paper 3).
Paper 1

Results

Single-copy gene analysis
Several studies have tried to characterize the chromatin-associated transcriptome in different organisms using a variety of methods, including GRID-seq in mouse [108], RADICL-seq in mouse embryonic stem cells [278] or ChAR-seq in Drosophila [271]. These studies however, have not investigated the dynamics and the regulation the RNA exosome might exert on caRNAs turnover. To this end, we performed a caRNA enrichment followed by sequencing in Drosophila melanogaster S2 cells. We performed this experiment in control cells (GFP-kd) and in cells depleted of the RNA exosome catalytic subunits, individually (Rrp6-kd, Dis3-kd) or both catalytic subunits simultaneously (Double-kd). We also isolated total RNA from the same cells to be used as reference. Taking into consideration that caRNAs have been implicated in many cellular processes, including chromatin homeostasis (reviewed in [107]), we parallelly analyzed the chromatin accessibility with ATAC-seq in the same four conditions. By jointly profiling RNA-seq and ATAC-seq, we have characterized how the RNA exosome regulates caRNAs levels and, most importantly, we have obtained new insights into the mechanism by which the RNA exosome maintains chromatin.

Our first objective in this study was to characterize the chromatin-associated transcriptome in control cells. In concordance with previous published data, we found a number of non-coding RNA species to be enriched in chromatin, including snRNAs or snoRNAs [271,272]. Moreover, we found most of the protein-coding genes to be more abundant in the total RNA preparation, which also agrees with previous observations indicating that the vast majority of RNA-DNA interactions at protein-coding loci are nascent transcripts [108,278].

We then characterized the changes in the ca-transcriptome upon depletion of RNA exosome catalytic subunits depletion. We observed approximately 500 differentially expressed transcripts when depleting Rrp6, and over 900 transcripts when depleting both catalytic subunits (Double-kd). Interestingly, almost no caRNAs abundance change was observed in Dis3 depleted cells. This result indicates that both catalytic subunits have common targets. The fact that the Double-kd shows the most striking phenotype supports the mechanistic redundancy of the catalytic subunits, and the lack of deregulation in Dis3 depleted cells indicates that Rrp6 can take over Dis3 function when this protein is depleted from cells. The target redundancy can potentially be explained by RNA exosome structural and functional data. Rrp6 contains helicase activity in addition to the exoribonuclease activity. With this helicase activity, Rrp6 unwinds RNA and channels it through the catalytically inactive core, to
present the RNAs to Dis3 [129]. Altogether, these results indicate that the RNA exosome is necessary for caRNAs turnover and homeostasis. We then addressed the question of whether depleting the RNA exosome resulted in chromatin organization alterations, and if so, whether we could link the alteration of RNA levels to the chromatin changes. In the ATAC-seq experiments, we observed that single depletion of catalytic activity of the RNA exosome (Rrp6, Dis3) did not have any significant impact on chromatin compaction. Contrarily, when depleting both Rrp6 and Dis3 simultaneously, we observed over 5000 chromatin loci with altered chromatin compaction, out of which over 80% showed increased accessibility, which suggests that the exosome is required to maintain the packaging of the chromatin.

Our ultimate goal was to correlate changes in caRNA levels to accessibility levels, and to do so we intersected both RNA-seq and ATAC-seq data sets. We defined a group of 346 genes that presented both RNA and accessibility changes, and we named them differentially expressed genes with differential accessibility regions (DEG_DARs). We wondered whether these genes possessed particular chromatin and genic features. Characterization of this subset of genes showed higher RNA exosome (Rrp6) occupancy with respect to a random set of genes, longer genes ORFs, and lower GC content. The fact that DEG_DARs show higher Rrp6 occupancy suggests that these regions are direct targets of the RNA exosome. Gene ontology (GO) enrichment analysis showed that DEG_DARs are associated with morphogenesis and developmental processes. We then analyzed chromatin features for these genes. To do so, we performed a meta-analysis of modENCODE [279] published data. This data includes the chromatin binding profiles of over 40 proteins, including histone modifications. We found a striking enrichment of Polycomb and Trithorax-group proteins in the DEG_DAR gene set.

A number of tools have been recently developed to infer the binding profiles of DNA-binding proteins from chromatin accessibility data. Taking advantage of one of these tools (TOBIAS, [280]), we asked whether the occupancy in chromatin of Pho and Trl/GAF were changed in Double-kd condition. We found indeed that Trithorax-group protein Trl/GAF binding was increased in cells depleted of both Rpr6 and Dis3 proteins; while Pho binding was reduced. These results suggest that the RNA exosome could be an important regulator of a developmental gene regulation network that involves Polycomb and Trithorax in the balanced state of the chromatin. This particular chromatin conformation in which both Polycomb and Trithorax complexes co-exist has been associated to developmental expression control (reviewed in [281]). The RNA exosome could play an important role in the regulation of this chromatin conformation at two different levels. First, increased RNA levels on chromatin together with the ability of PRC2 to promiscuously bind RNA, could result in
PRC2 eviction from the chromatin. Second, Polycomb binding has been shown to be regulated via R-loop [117,192,193], and the RNA exosome has been implicated in R-loop resolution in eukaryotic cells [149]. These observations could explain the embryonic lethality phenotypes observed in RNA exosome knockout flies [282].

Taking all together, we propose that the RNA exosome controls the turnover of caRNAs, and that the RNA degradation impacts on the chromatin organization of balanced chromatin states of developmental genes.

Repetitive chromatin-associated RNA analysis

Similarly to caRNAs, it has been proposed that rep-caRNAs can play key roles in important biological processes. One example of it, is the establishment of pericentromeric heterochromatin in fission yeast by RNAi mechanism [109]. However, RNA-seq experiments usually disregard the repetitive genome, mainly because of technical difficulties in assigning reads to specific loci. To overcome the technical difficulties of mapping reads to repetitive regions, we utilized a three fold strategy to quantify the RNA abundance in repetitive regions: 1) counting only the uniquely mapped reads (unique mappers), 2) counting the reads mapping to multiple locations (multi-mappers), and 3) extending the repetitive regions annotations 50 bp upstream and downstream, in order to escape from the repetitive regions of the genome, thus increasing the amount of uniquely mapped reads.

We quantified the extent of rep-caRNA transcription and detected rep-caRNAs transcribed from over 10 000 repetitive loci in control cells, accounting for more than 10% of the RNA amount in the ca-transcriptome. When considering multi-mapping reads, this number increased substantially, up to almost 20 000 detected repetitive elements (19 458).

We then wanted to assess whether the RNA exosome was also controlling the rep-caRNAs turnover. We observed that when depleting the RNA exosome from S2 cells, the number of detected repetitive elements increased by approximately 50% in Rrp6-kd and Double-kd cells. Differential expression analysis identified 544 increased and 112 decreased rep-caRNAs in Double-kd cells compared to GFP control. On average, all repetitive families analyzed showed increased RNA abundance. In order to characterize the chromatin signatures of this group of rep-caRNAs, we intersected them with the chromatin states model published by Karchenko and collaborators [179], and observed a strong enrichment in pericentromeric chromatin. Indeed, when plotting the differentially expressed rep-caRNAs in a karyotype plot, we could observe clear clusters of repetitive elements in the pericentromeric regions of chromosomes 2R, 2L and 3L. Moreover, re-analyses of publicly available ChIP-seq datasets showed that the rep-caRNAs located in the pericentromeric regions showed higher levels of H3K9me3 and HP1 occupancy than both a random set of
repetitive regions and a set of repetitive regions located in the chromosomal arms. Mirroring the analysis realized for single-copy transcripts presented above, we also asked whether the differentially expressed rep-caRNAs presented chromatin alterations. Indeed, we could demonstrate that the differentially expressed rep-caRNAs were associated with higher chromatin accessibility changes in Double-kd cells than a random set of repetitive transcripts, and that the pericentromeric region presented a general decondensation. To further confirm the pericentromeric chromatin decompaction, we performed immunofluorescence analysis against CID. CID is a specific histone variant that is only located in centromeric nucleosomes of Drosophila chromosomes. We compared control cells versus cells depleted of both Rrp6 and Dis3, and found a significantly more intense fluorescence signal in Double-kd S2 cells. Whether this increased signal was due to chromatin decondensation and increased accessibility of CID or whether more CID could be incorporated to the pericentromeric chromatin in the double-knockdown condition could not be determined with this experiment.

With all the presented results we concluded that the RNA exosome plays an important role in degrading the rep-caRNAs, and that the lack of RNA exosome catalytic activity led to a deregulation of pericentromeric chromatin organization.

Conclusions and future perspectives

In this study, we have characterized several important aspects of the RNA exosome function. However, to our knowledge this is the first study that has addressed in a genome-wide manner the RNA degradation function of the RNA exosome directly in the chromatin of eukaryotic cells. The main findings of the study are:

The RNA exosome controls the homeostasis of two major sets of transcripts. Firstly, we have proven that the caRNA levels of almost a thousand single-copy genes are affected upon depletion of both catalytic subunits of the RNA exosome. Secondly, we have proven that the RNA exosome is essential for controlling the spurious transcription of rep-caRNAs, specifically in the pericentromeric region of the chromosomes. caRNA abundance deregulation is linked to chromatin compaction alterations. The first set of transcripts consists of almost 350 transcripts produced from genomic regions that show chromatin alterations in exosome-depleted cells (DEG_DARs). We have proven that these genes are closely related to developmental functions, and very likely to be located in bivalent domains of the chromatin. We propose that the RNA exosome is an important regulator of this chromatin conformation, by fine-tuning the RNA abundance on the chromatin.
Repetitive transcriptome accounts for more than 10% of the caRNA detected. Moreover, we have determined that eliminating the RNA exosome catalytic activity from cells results in a 50% increase of detectable rep-caRNAs, highlighting the importance of the RNA exosome in turning over the spurious transcripts arising from repetitive regions. The RNA exosome is required for normal chromatin compaction in pericentromeric heterochromatin. We observed a global decompaction in these chromatin regions, presumably because the increase of RNA levels compete and hijack the protein HP1 from the chromatin, resulting in loosening of the chromatin [124].

Taken together, these results enhance the importance of the RNA exosome function of maintaining the RNA homeostasis, and specifically this study has proven that this function occurs directly on the chromatin.
Paper 2

Results

In recent years, snoRNAs have attracted the interest of the research community, mainly due to three interesting observations. First, with the development of high-throughput techniques able to map global RNA-DNA interactions, it has been shown that snoRNAs are highly enriched in the chromatin [271]. This observation is consistent with the result of the analysis of Drosophila melanogaster S2 cells caRNAs presented in Paper 1. Second, it has been proven that snoRNAs are upregulated in response to viral infection [270]. Lastly, snoRNAs have been shown to be deregulated in pathological processes, and have been associated, for instance, with several types of cancer [267,269]. These observations, together with the fact that the function of many snoRNAs remains elusive, led us to investigate the functions of snoRNAs on chromatin in D. melanogaster.

In paper 2, we have investigated the potential of snoRNA:U3:9B to influence chromatin and gene expression. To this end, we have used a model of Sindbis virus infection to study chromatin and gene expression changes. We chose to focus our analysis on snoRNA:U3:9B for three main reasons: its chromatin association in our RNA-seq analysis (paper 1), CHAR-seq chromatin association, and upregulation in viral response (see below).

We first validated by RT-qPCR a number of snoRNAs that had been shown to bind chromatin both by ChAR-seq [271] and in our own caRNA experiment in S2 cells (Paper I). Besides RNA levels on chromatin quantification, we performed RNA FISH to analyze the distribution of snoRNA:U3:9B in the chromatin, both at polytene chromosomes and S2 cells. This experiment corroborated the chromatin association detected in our Paper 1 and the genome-wide distribution of snoRNAs observed in the ChAR-seq study [271]. Moreover, we observed that the snoRNA:U3:9B overlapped with active transcription marks (H3K27ac) and the protein Df31, which in a previous study has been shown to be associated with snoRNAs to keep chromatin conformation open [122]. Taken together, there is compelling evidence indicating that snoRNA:U3:9B is associated with chromatin, and located in active transcriptional regions.

Our next question was to elucidate whether snoRNA:U3:9B was involved in the regulation of genes linked to any specific biological function or molecular process. To answer this question, we carried out a gene ontology (GO) analysis of protein-coding genes bound by snoRNA:U3:9B in the ChAR-seq study and found links to the immune system and signal transduction. A second piece of evidence supporting a link to immune response came from modENCODE metadata analysis. In this analysis, we extracted gene expression data for all annotated snoRNA genes in the genome of D. melanogaster and asked whether their
expression changed upon Sindbis virus infection. We observed a statistically significant increase that was specific for snoRNAs and not observed in a random set of protein-coding genes.

To validate the results of the meta-analysis and confirm the relation of snoRNA:U3:9B with Sindbis virus infection, we used a previously published model that induces a pseudo Sindbis virus infection in S2 cells [283]. This model induces the controlled expression of non-structural proteins (nSPs) required for Sindbis virus RNA replication (viral replicon) in S2 cells. The same replicon was used by the modENCODE project to model Sindbis virus infection. The expression of nSPs is controlled by the UAS-Gal4 expression system. The viral replicon is contained in a vector consisting of the nSPs of the virus fused to GFP. The GFP protein replaces the structural proteins of the virus impairing the ability of the Sindbis virus to produce virions. This implies that the expression of the nSPs occurring within the cells containing the construct will not derive in a systemic infection. As a control, in the same vector, we partially deleted the nSPs (no replicon). This deletion is unable to induce the innate immune response pathway as shown by Avadhanula and colleagues [283].

We stably transfected the Sindbis vectors (viral replicon and no replicon) into S2 cells. We induced the viral replicon and quantified snoRNA:U3:9B levels in both replicon and no replicon conditions. Not only could we detect induction of the snoRNA, but also of genes related to infection, that we considered as targets of snoRNA:U3:9B, based on ChAR-seq data. This subset of immune response genes include TotM, Gclc, PGRP-SD, vir-1, CG4572, Vago, DptB, Tep2, dnr1 and Drs. The observed response in S2 cells (increased snoRNA:U3:9B and target genes expression) was abolished when performing snoRNA:U3:9B knockdown experiments using small hairpin RNAs (shRNAs). To further investigate the immune response to the viral challenge, we analyzed the chromatin accessibility of the target genes that showed RNA expression changes upon viral replicon induction. We observed a significant increase of chromatin accessibility in these genes, measured with ATAC-qPCR, upon induction of Sindbis infection. The increase of chromatin accessibility was abolished by snoRNA:U3:9B knockdown, which suggests that snoRNA:U3:9B is required for the correct induction of immune response genes through a mechanism that entails chromatin regulation.

In order to move into a more physiological context, we used the same infection model in flies. The Sindbis replicon transgenic flies were procured from Dr. Hardy lab, Indiana University, Bloomington [283]. We could reproduce the same findings observed in S2 cells. In particular, upon Sindbis virus replicon induction, we observed significantly increased expression (both of snoRNA:U3:9B and target genes) and increased accessibility of target genes in third instar larval brains. The increased expression and chromatin accessibility were not observed in the no-replicon D. melanogaster control.
To test whether the expression and chromatin changes of the target genes were dependent on snoRNA:U3:9B, we created a transgenic fly with snoRNA:U3:9B knock-out by CRISPR-cas9 (snoRNA:U3:9B KO). The snoRNA:U3:9B knock-out strain was viable. We then crossed snoRNA:U3:9B knock-out flies with the viral replicon transgenic flies. We performed it in two independent deletion lines to draw robust conclusions. Interestingly, upon induction of the Sindbis replicon, we observed a lethality phenotype in the KO lines. Upon Sindbis virus infection, the snoRNA:U3:9B KO flies could not reach the wandering larvae stage and the percentage of pupation was drastically reduced. To characterize this phenotype, we again quantified the immune response target gene expression and chromatin accessibility upon infection. We observed ablation of target gene expression induction and the chromatin conformation remained unperturbed in viral replication conditions in KO flies. These results support the conclusion that snoRNA:U3:9B is an integral part of the innate immune response, and it is required to modulate the expression and chromatin compaction of a subset of genes implicated in the cellular response to Sindbis virus replication.

To further investigate the molecular mechanism by which snoRNA:U3:9B influences the expression of immune response, we asked whether snoRNA:U3:9B was bound to the target genes loci. To this end, we utilized ChIRP-qPCR, a technique that relies on target pull-down with antisense oligonucleotides. We designed antisense probes for snoRNA:U3:9B, and quantified the binding of this snoRNA to target genes in larval brains with and without replicon. We could detect the binding of snoRNA:U3:9B to the target genes and we observed increased binding upon viral replicon induction. These results indicate that snoRNA:U3:9B is required to stimulate immune response through binding to chromatin and regulation of immune response genes activated upon Sindbis viral replicon induction.

Conclusions and future perspectives

We have demonstrated in vivo that snoRNAs are enriched in the chromatin of Drosophila melanogaster. We have also shown the upregulation of snoRNA:U3:9B in response to the induction of a viral replicon. We have demonstrated, both in S2 cells and flies, that snoRNA:U3:9B modulates the expression and the chromatin accessibility of a subset of target genes implicated in the immune response of D. melanogaster to Sindbis viral infection. Finally, by utilizing CRISPR-Cas9 we genetically engineered flies that lack the snoRNA:U3:9B gene, and showed that these KO flies show a lethality phenotype upon induction of the Sindbis virus replicon. To our knowledge, this is the first study that shows the implication of a snoRNA in the antiviral defense mechanism.
Future perspectives could include investigating the mechanistic aspect of RNA modifications. snoRNA:U3:9B is a C/D-box snoRNA. The canonical function of C/D-box family members is 2’O-methylation RNA modification in rRNA. It would be extremely informative to profile the epigenome of non-infected versus infected flies, in control and KO flies for snoRNAs in order to find differences in methylation profiles, since C/D box are known to drive methylation and such RNA modifications have been shown to impact RNA stability and cellular fate [216].
This study addresses the biological function of ADAR3, the editing-deficient member of the ADAR protein family. This question remains open to date, although several studies have started to shed some light on the involvement of ADAR3 in several cellular processes. Due to its high similarity with ADAR2, it has been hypothesized that ADAR3 could play a role in RNA-editing inhibition [246] by competing and excluding ADAR2 from its targets. Since ADAR2 and RNA editing are developmentally regulated, we asked whether ADAR3 could also be controlled in mice brain during development and if ADAR3, by sharing spatio-temporal expression with ADAR2, could act as RNA-editing modulator.

In this study, we approached this question from a developmental perspective, and profiled ADAR3 at several levels, including expression, subcellular location, and interaction partners in cellulo. To characterize the spatial and expression patterns of ADAR3 in mice developing brains, we quantified ADAR3 at mRNA and protein levels using real-time quantitative PCR (RT-qPCR) and Western blotting, respectively, in whole mice brains during development. We observed a steady increase of ADAR3 at mRNA and protein levels throughout embryonic development, peaking at embryonic stage 18, prior to birth. After birth, mRNA levels remained comparable with the levels observed before birth. However, the protein abundance drastically dropped. To further characterize ADAR3 subcellular distribution in vivo, we performed immunofluorescence studies of this protein together with ADAR2 in primary cultured cortical neurons differentiated in vitro. We observed a widespread cellular distribution of ADAR3 in the first days of in vitro culture. After day 5, the vast majority of ADAR3 had been translocated to the nucleoli. ADAR2 was also redistributed to nucleoli, as previously shown [284], but ADAR3 translocation to the nucleoli was observed prior to that of ADAR2. The early translocation to the nucleus could be attributed to the fact that ADAR3 and ADAR2 utilize different proteins for nuclear import (Importin-a1 and Importin-a4, respectively) [285]. The strong nucleolar ADAR3 localization could be explained by the presence of the R-domain. This domain binds ssRNA, and the nucleolus harbors the most prominent transcriptional patch in the entire cell nucleus, the rRNA transcription. Interestingly, ADAR3 signal decreased during in vitro differentiation from most of the primary cortical neurons, and remained confined to a small subpopulation after day 6 of in vitro differentiation. This population could belong to the VIP/PAX/LAMP5 inhibitory interneurons subtype, based on previous observations [286]. The fact that ADAR2 and ADAR3 do not share spatio-temporal location after day 6 of in vitro differentiation argues against ADAR3 acting as a general inhibitor of ADAR2 RNA-editing.
To further characterize the function of ADAR3, we produced a lentiviral expression vector for ectopic expression of a FLAG-tagged ADAR3 protein and stably transduced the mouse neuroblastoma cell line N2a, which does not express endogenous ADAR3. The mixture of selected ADAR3-expressing cells was called N2a-A3 cells, and the parental control line N2a-C. In order to identify the ADAR3 interactors, we performed ADAR3 Co-IP in N2a-A3 cells followed by mass spectrometry (LC/MS-MS). We were able to identify different enriched biological processes among the ADAR3 interactors, and we focused specifically on RNA stability and translation.

In order to elucidate the relation of ADAR3 with RNA stability, we performed polyA+ deep RNA sequencing, both in N2a-A3 and N2a-C cells in two different conditions; unperturbed (time = 0h), and treated for 8 hours with triptolide (time = 8h), a transcription inhibitor [287]. Therefore, with transcription blocked, we could compare the RNA decay in control cells and N2a-A3 to infer whether ADAR3 has any impact on RNA stability. We identified a total of 1497 transcripts with affected RNA stability, out of which 1024 and 473 were stabilized and destabilized respectively. GO enrichment analysis ascribed the stabilized transcripts to biological processes related to cell proliferation and pluripotency maintenance (ribosomal large subunit biogenesis and RNA processing), while the destabilized transcripts were associated to cell fate commitment and cell fate specification. These results suggest that ADAR3 could play a role in regulating neural differentiation.

To test the possibility of ADAR3 being a RNA editing inhibitor, we utilized a common bioinformatic pipeline to call the editing events in RNA-seq data of N2a-C and N2a-A3 cells. This pipeline simulates a variant calling, taking the advantage of the fact that RNA editing events are detected as mismatches (A/G) when aligning the reads against the reference genome. We observed minor changes in the global profile of N2a-A3 editome compared to N2a-C cells. Only 260 differential editing events were detected out of the total 19000 editing events detected. Despite not being striking, the observed differences were more prominent than those observed by Mladenova et al. [246] as their study could not detect editing alterations in ADAR3 deficient mice. To rule out the possibility that these editing events could be the main drivers of the RNA stability changes observed in N2a-A3 cells, we intersected both RNA stability and RNA editing data sets. No significant overlap was found between changes in RNA stability and new/lost RNA editing events in N2a-A3 cells. We concluded that ADAR3 influences RNA stability genome-wide, and that this function is editing-independent.

Based on the obtained results and previously published data, we asked if ADAR3 could indeed influence neuronal differentiation. To test this possibility, we subjected N2a-C and N2a-A3 cells to differentiation using a well established neuronal differentiation agent, retinoic acid, for 24 hours [288]. After differentiation, we could observe that N2a-A3 cells showed
shorter protrusions, and morphologically were more similar to pluripotent cells (more rounded, less attached) than to differentiated neurons. In order to further characterize this phenotype, we performed whole proteome mass spectrometry and found striking differences between control and ADAR3 cells. GO enrichment analysis of differentially expressed proteins in control N2a-C cells revealed links to pathways such as differentiation and cell trafficking, while these GOs were missing in N2a-A3 cells, in agreement with limited ability of N2a-A3 cells to respond to differentiation cues.

RNA stability has been shown to be tightly linked to translation. This fact, together with the enriched GO terms related to translation obtained in the analysis of ADAR3 interactome, we decided to investigate whether ADAR3 could directly impact translation. To test this idea, we performed a puromycin incorporation experiment, in which we compared the bulk of translational output in N2a-C and N2a-A3 cell lines. We observed that N2a-A3 cells incorporated less puromycin compared to control cells. To further investigate the cause of this translational defect, we carried out a polysome profiling experiment and observed no significant changes in the 40S, 60S and 80S fractions. However, N2a-A3 cells showed increased polysome levels. Polysomes are usually associated with active translation. These apparently contradictory results (less translation, but higher polysome levels) could be explained if ADAR3 reduced the translation elongation rate, stabilizing the ribosomes on the mRNAs in a non-productive translational state. Summarizing, together with the metabolic labeling experiments, we concluded that ADAR3 can act as a modulator of translation.

Conclusions and future perspectives

In this study, we have shown the ADAR3 spatio-temporal expression in the mice developing brain. We have been able to characterize the ADAR3 protein interactome, linking this protein to RNA stability and translation processes. When investigating the relation between ADAR3 and RNA stability, we determined that ADAR3 expression alters the mRNA half-life of over a thousand transcripts. Interestingly, the genes with affected RNA stability are related to differentiation and neurogenesis. ADAR3 has been proposed as a general inhibitor of RNA editing. Instead we observed only a few specific editing sites changing between control cells and ADAR3 cells. Thus, we concluded that in our N2a-A3 cells we could not find a genome-wide relation between ADAR3 and regulation of editing.

Due to the fact that RNA stability and translation are closely related and that we could detect several translation factors among the ADAR3 interactome, we measured the translation competency of N2a-A3 cells and found a significant decrease of translation output compared
to control cells. We also observed higher polysome fractions in N2a-A3 cells in relation to the control cell line. With these results, we propose that ADAR3 regulates translation elongation. Several transcripts destabilized in N2a-A3 cells are related to neuronal differentiation. To test the implication of ADAR3 in neuronal differentiation, we treated N2A-C and ADAR3 expressing cells with retinoic acid to induce differentiation. Consistently, we observed N2a-A3 cells to have attenuated response to the differentiation stimulus, presenting shorter neuronal projections and a morphology more similar to non-treated cells. Taking all presented results, we propose that ADAR3 plays an important role in neuronal cell fate specification, by instructing the cells into a specific inhibitory neuronal type.

Among the future perspectives, it would be interesting to look at whether the diminished translation output in ADAR3 cells is due to decreased elongation rate, less efficient translation termination or initiation, or suboptimal recycling of the translation machinery. To address this question, genome-wide experiments profiling the ribosome are required, for example Ribo-seq or HT-5Pseq. With this last method, HT-5Pseq, we could also investigate the relation of translation and RNA degradation, to link both pieces of evidence, presented in the current study. This experiment could be carried out in steady-state cells and in cells treated with retinoic acid, to simultaneously study the changes that ADAR3 induces in cells that are undergoing neuronal differentiation.
Main conclusions of this thesis

- The RNA exosome is required to maintain the homeostasis of chromatin-associated RNAs. In absence of the RNA exosome, chromatin processes such as the pericentromeric region organization or balanced chromatin states are deregulated.


- ADAR3 is developmentally regulated, both at expression and subcellular location levels. ADAR3 influences the RNA stability of a subset of transcripts involved in neuronal differentiation. ADAR3 also modulates translation, through regulation of the translation efficiency. ADAR3 regulates cell-fate determination and neuronal specification. During early developmental stages, ADAR3 is cytoplasmic and represses the expression of certain mRNAs involved in neuronal differentiation. We propose that at later developmental stages, the reduced expression of ADAR3 and its nucleolar localization allow for the expression of such genes and favor neuronal specification.
References


2. WATSON JD, CRICK FHC. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature. 1953;171(4356):737-738. doi:https://doi.org/10.1038/171737a0


24. Pelechano V, Pérez-Ortín JE. There is a steady-state transcriptome in exponentially growing yeast cells. Yeast. 2010;27(7):413-422. doi:https://doi.org/10.1002/yea.1768
34. Han Z, Libri D, Porrua O. Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. Nucleic Acids Research. 2016;45(3):1355-1370. doi:https://doi.org/10.1093/nar/gkw1230
40. Burke TW, Kadonaga JT. Drosophilia TFID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes & Development. 1996;10(6):711-724. doi:https://doi.org/10.1011/gad.10.6.711


73. Iguchi-Ariga SM, Schaffner W. CpG methylation of the cAMPresponsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. Genes & Development. 1989;3(5):612-619. doi:https://doi.org/10.1101/gad.3.5.612
76. Owen BM, Davidovich C. DNA binding by polycomb-group proteins: searching for the link to CpG islands. Nucleic Acids Research. 2022;50(9):4813-4839. doi:https://doi.org/10.1093/nar/gkac290
78. Allfrey VG, Mirsky AE. Structural Modifications of Histones and their Possible Role in the Regulation of RNA Synthesis. Science. 1964;144(3618):559-559. doi:https://doi.org/10.1126/science.144.3618.559


102. Preker P, Almvig K, Christensen MS, et al. PROMoter uStream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. Nucleic Acids Research. 2011;39(16):7179-7193. doi: https://doi.org/10.1093/nar/gkr370


130. Lorentzen E, Basquin J, Tomecki R, Dziembowski A, Conti E. Structure of the Active Subunit of the Yeast Exosome Core, Rrp44: Diverse Modes of Substrate Recruitment in the


146. Canavan R, Bond U. Deletion of the nuclear exosome component RRP6 leads to continued accumulation of the histone mRNA HTB1 in S-phase of the cell cycle in Saccharomyces cerevisiae. Nucleic Acids Reasearch. 2007;35(18):6268-6279. doi:https://doi.org/10.1093/nar/gkm691

63


149. Richard P, Feng S, Manley JL. A SUMO-dependent interaction between Senataxin and the exosome, disrupted in the neurodegenerative disease AOA2, targets the exosome to sites of transcription-induced DNA damage. Genes & Development. 2013;27(20):2227-2232. doi:https://doi.org/10.1101/gad.224923.113


211. Amrani N, Ghosh S, Mangus DA, Jacobson A. Translation factors promote the formation of two states of the closed-loop mRNA. Nature. 2008;453(7199):1276-1280. doi:https://doi.org/10.1038/nature06974


250. Raghava Kurup R, Oakes EK, Manning AC, Mukherjee P, Vadlamani P, Hundley HA. RNA binding by ADAR3 inhibits adenosine-to-inosine editing and promotes expression of immune
response protein MAVS. Journal of Biological Chemistry. 2022;298(9):102267. doi:https://doi.org/10.1016/j.jbc.2022.102267


Vispé S, DeVries L, Créancier L, et al. Triptolide is an inhibitor of RNA polymerase I and II-dependent transcription leading predominantly to down-regulation of short-lived mRNA. Molecular Cancer Therapeutics. 2009;8(10):2780-2790. doi:https://doi.org/10.1158/1535-7163.MCT-09-0549

Acknowledgements

Probably the hardest part of writing this thesis... I have been staring at the blank page for several minutes and still don't know where to start. Let's try to start by the beginning of my scientific journey shall we?

Before the scientific little Jordi was born, I studied biology at school, and my biology teacher awakened the passion for molecular biology, Javier Subias. Todavía me acuerdo de entrar en clase y ver los glucidos los dibujados en la pizarra, o cuando explicaste la transcripción. Gracias a ti, también entré por primera vez en un laboratorio. La vida es dura... pero un poquito menos si encuentras lo que te gusta!

Little Jordi started his scientific career in Universitat de València, in the Department of Biochemistry and Molecular Biology. And he did it because Jose Enrique Pérez Ortín gave him the opportunity. I cannot express the gratitude I feel for having been given this opportunity. I learned how science works, and started to develop my interest in RNA, transcription and RNA degradation. Not only Jose Enrique, but also the entire GFL group at the moment, who welcomed me warmly and immediately made me part of the group. This includes Paula Alepuz, a smart and amazing researcher also in charge of the GFL group, Vicente Tordera, director of my bachelor thesis, Pepe, always bringing the energy and good vibes to the group, Tian, Daniel, Adriana, always ready to help disconnect from a frustrating day with a beer, María, Alba, Alice, Mangels, Fani, Ester, Marta, and last but not least Toni Jordán. The recently graduated PhD student that emigrated to Sweden, to do the postdoc in a lab working in the RNA exosome, led by professor Neus Visa... Can anyone spot the link? Thanks Toni, for offering me this many opportunities, for fruitful scientific advice and for helping me when I couldn't find what or how to do the analysis. I remember with great affection my early days in science. I spent three amazing years in the GFL group and if I do it is because of you all!

After the master thesis, little Jordi wants to do a PhD, but he can't stay in GFL. Toni suggests that his boss might be looking for someone. And that's how little Jordi contacts professor Neus Visa to meet via Skype call. Good morning, my name is Jordi... Bon dia Jordi, sóc Neus. And the rest you know it. Little Jordi comes to Sweden to do a short internship that then becomes PhD studies. Neus, I have no words to acknowledge that you offered me the possibility to come here and do the PhD in your group. It has been an incredible journey, and you have helped me immensely to grow, not only as a scientist but also as a person. Jordi is
no longer little Jordi, and you have had a big impact on it. Despite ups and downs during the PhD, I have really enjoyed the journey. And the journey has been great because group Visa provides the perfect environment. This happens only because of your enthusiasm, energy and passion for science.

To my co-supervisor, Toni (can you stop being everywhere in acknowledgements?). Thanks for your help and mentorship. Bioinformatics is definitely not easy, especially in the beginning. Uncountable hours in front of the screen, with the prompt only spitting errors. Just before falling into despair, Toni would always offer a life vest to prevent me from sinking.

To the Visa lab, past, present and future. When I arrived, Judit, Franziska, Martin and Anne were the heart and soul of Visa lab. I’d like to thank Judit, you made the transition to Sweden really easy for me. Beers, dinners and even climbing! We have even lived together… for a week? It has always been great to hang out with you. Martin, partner in crime as PhD students, I hope everything goes great! Anne, always ready to click you some nice images, FISH is her speciality! Patri and Shruti, recruited more or less at the same time as I was. Who would have said that you two would become such an important part of my life. Shruti, what can I say about you… Even though your first impression of me wasn’t the best hehe (I know you are still upset because I didn’t order a beer the first night we met). In the end, we shared beers, dinners, scientific achievements, scientific disappointments, congresses, experiments, analysis, lunch in the balcony, trips… Basically, we have shared 6 years of our lives, and more to come! You are a brilliant scientist, and an even more brilliant person. You’ll always find a place to stay in Valencia, and I hope to visit you in Delhi at some point (even though now you have become Swedish). Patri, amiga! I have been terribly lucky to find you in my way. You have been home when I was missing it, you have been essential in many aspects of my life when being in Sweden. I miss you a lot!

To the latest additions of the Visa lab, Elin the more I got to know you, the more and more I appreciate how fun you are, always ready for a big laugh and a little gossip. Research-wise, you are brilliant. You always have the right suggestion and the right criticism to bring the projects closer to the end. You are amazing, I’m still trying to understand how you are capable of doing 17 experiments at a time, boss of multitasking!

Akshit, you were the latest to arrive, but I’m sure the future of the lab is in good hands!

To all the students that have passed by the Visa lab: Akif, Kristofer, Andrea, Fredrik, you always bring freshness to the lab and help keep things working out! Special mention to Núria, I really had so much fun having you around. I tried to teach you the essence of life…Haber estudiao! We’ll have another one, in my defense!

To Victor, Albin and Eduardo, the siamese part of the Visa lab. It has been a great pleasure to share beers with you in the pubs and BBQs, but it has been even more enjoyable to collaborate with you. In the end, after much struggle and discussions (imputation, yes/no?), 74
we are close to providing the world with the true function of the almighty ADAR3 protein. Cheers!
To all collaborators across my projects. Here in MBW Mattias and Claes. Thanks for your fruitful discussions, whereas it has been in the ISPs or with a quick chit-chat in the corridor. In SciLifeLab, Vicent Pelechano, thanks for always welcoming us to your lab for all imaginable kinds of sequencings.
People in the MBW department, the OGs: Franz, Victor, Albin, Lukas, Carlotta, Juta, Fredrik, Mattias. Thanks for all the good times and moments we have shared together. Sergi, company d’aventures. No només científiques, però també d’experiències, inquietuds, escalada i hamburgueses! Em va tocar traduir el discurs de ta mare a la teua defensa, ara et toca a tu fer-ho en la meua.
People from Wojcik lab, thanks for all the fikas, BBQs and trips to Rome (Alpin)! Mila, as of today, I know I still owe you a paella. I promise I’ll cook it before you leave, mark my words! Thanks to the fairy of the sweet tooth, Zuza, for always feeding my brain with glucose when I needed it the most.
To the rest of MBW, you make the perfect environment to do a PhD.

Enough of scientific acknowledgements. Time to acknowledge the people that made me survive during my PhD.
People I met in Sweden can be divided into two major chunks. The Spanish gang and people I met climbing, although the overlap of both groups is surely quite significant. In the Spanish gang: Fermín, Maria, Irene (animitos con el doctorado!), Guille, Ruben, Alex, Alma, Maddi, Joaquín, Carol, Javi, Xabi, Dani, Blanca, Juanjo. Suecia es un poquito más casa con vosotros aquí. Por estar ahí siempre que era necesaria una paella, unos calçots, unas cervezas, ir a por setas o por los saraos que hemos montado! Especial mención a mi compañero de doctorado Fermín. Por esas noches en las que se ha pegado una lloradita y a seguir existiendo. Recuerda, never smile, only compile.
Climbing gang: I started climbing seriously when I came here to Stockholm. And climbing I have met amazing people. Marcel, Gerard, Nico ja no esteu ací, però uno dels millors estius de la meua vida ha sigut ací a Estocolm escalant en vosaltres. Pol, Enric, de les últimes adquisicions! Gràcies per totes les rises entrenant a gasverket. Bella, the Finnish less Finnish that I know. You are amazing, always bringing joy and good energy. We still have to climb together in Albarracín and use the guidebook I got for you!
A la gent de la meua carrera bioquímica i ciències biomèdiques. Pere, Josep, Marina, Miriam, Carlota, Polo, Maria, Nora, Mireia, Nuria, Pablo. Especialment a Marina, que ha sigut l’única que ha vingut a vorem cabrons!
Als meus amics més propers de València: Seller, Elias, Miguel, Jaume, Carles, Maria, Laura. Seller gràcies per vindre a visitar-me ací, tinc moltes ganes de reprendre les nostres històries germà! Elias, Jaume, Miguel, Carles, els de sempre. Junts. 20 anys i ací seguim. Com canvia la vida, d’anar tots els caps de setmana als concerts de Burjassot, a estar a punt d’entrar en la trentena. A les hermanas Franco. Maria, per venir a verme ací, tinc moltes ganes de reprendre les nostres històries germà!

Elias, Jaume, Miguel, Carles, els de sempre. Junts. 20 anys i ací seguim. Com canvia la vida, d’anar tots els caps de setmana als concerts de Burjassot, a estar a punt d’entrar en la trentena. A les hermanas Franco. Maria, per venir a verme ací, tinc moltes ganes de reprendre les nostres històries germà!

Elias, Jaume, Miguel, Carles, els de sempre. Junts. 20 anys i ací seguim. Com canvia la vida, d’anar tots els caps de setmana als concerts de Burjassot, a estar a punt d’entrar en la trentena. A les hermanas Franco. Maria, per venir a verme ací, tinc moltes ganes de reprendre les nostres històries germà!

A mon pare i ma mare. Podria donar-vos les gràcies per recolzar-me durant aquestos sis anys. Però que jo haja pogut acabar aquesta tesi comença fa 29 anys. Gràcies per l’educació, l’estima i les oportunitats que m’heu donat en aquesta vida. Tan malament no ha eixit el xiquet, no? A la meua iaia, aunque no hayas visto el final de la tesis, también va por ti. A la meua família de campanar, Luís, Paqui i Laura. Supose que al principi tot era un poquet extrany, només arribar i me’n vaig a Suècia. Malgrat això, sempre us he sentit molt a prop. Sempre m’heu donat el vostre suport i estima que han sigut molt necessaris per a aconseguir acabar aquesta tesi!

I per últim, la part més important de la meua vida, Silvia. Podria escriure un text més llarg que la propia tesi, però preferisc simplement donar-te les gràcies. Gràcies per aquestos sis anys. Gràcies per la teua valentia de vindre a Suècia per a viure amb mi més d’un any i mig sense dubtar-lo. Gràcies per la teua comprensió, recolzament i per haver aguantat la distància. Gràcies per tot el que hem compartit, el que compartim i que estic segur compartirem al nostre futur. En definitiva, per estimar-me i compartir aquesta vida en mi. T’estime infinit!