UMEÅ UNIVERSITY

Regulation of gene expression in fruit flies;

How does it start, and will it be remembered?

Aman Zare

Department of Molecular Biology
Umeå 2018

Dissertation for PhD
ISBN: 978-91-7601-973-3
Cover designed by Aman Zare
Electronic version available at: http://umu.diva-portal.org/
Printed by: UmU print service, Umeå University
Umeå, Sweden 2018

The only certainty is that nothing is certain.

Pliny the Elder

Table of Contents

Abstract

One of the most distinctive features of eukaryotic chromosomes is the bundling of DNA together with functionally associated RNA and proteins in chromatin. This allows huge amounts of DNA to be packed inside the very tiny space of the nucleus, and alterations in the structure of chromatin enable access to the DNA for transcription ("reading" genes by production of RNA copies). Much of the current knowledge of chromatin structure and regulation comes from studies of *Drosophila melanogaster*. When the chromatin structure is open the transcription of a gene can start after recruitment of the necessary factors. The main enzyme for gene transcription is Polymerase II (Pol II). For successful gene transcription, Pol II must not only be recruited to the gene's promoter, but also escape from a pausing state which occurs soon after transcription initiation. CBP/P300 is one of the co-activators involved in transcriptional activation. In the studies this thesis is based upon, my colleagues and I (hereafter we) discovered a new function for CBP in transcription activation. Using high throughput sequencing techniques, we found that CBP directly stimulates recruitment of Pol II to promoters, and facilitates its release from the paused state, enabling progression to the elongation stage of transcription.

For cells to remember their identity following division during development, the transcriptional state of genes must be transmitted. Intensively studied players involved in this memory are the Polycomb group (PcG) proteins, responsible for maintaining the repressed state of important developmental genes. The core members are Polycomb repressive complex 1 and 2 (PRC1 and PRC2), which are recruited in flies through poorly known mechanisms to target genes by so-called Polycomb response elements (PREs). Using *Drosophila* mutant cell lines, we showed that (in contrast to previous models) some PREs can recruit PRC1 even when PRC2 is absent. We also observed that at many PREs, PRC1 is needed for recruitment of PRC2 and concluded that targeting PRC complexes to PREs is a much more flexible and variable process than previously thought.

Some phenotypic effects of environmental changes can be transferred to subsequent generations. Previous efforts to identify the mechanisms involved have focused on material (mainly, but not only, DNA) transferred through germ cells. However, organisms' microbiomes are also transferred to the next generation. Thus, to investigate possible contributions of microbiomes to such transfer, we used fruit flies as the microbiomes they inherit can be easily controlled. We altered some parents' environmental conditions by lowering the temperature, then grew offspring that received microbiomes from cold-treated and control parents in control conditions and compared their transcriptional patterns. Our results suggest that most of the crosstalk between the microbiome and the fly happens in the gut, and that further investigation of this previously unsuspected mode of inheritance is warranted.

Papers included in this thesis

I. Boija A, Mahat DB, **Zare A**, Holmqvist PH, Philip P, Meyers DJ, Cole PA, Lis JT, Stenberg P, Mannervik M. (2017). CBP **regulates recruitment** and **release** of **promoter-proximal** RNA Polymerase II. *Mol Cell. 2017 Nov 2;68(3):491-503.e5.*

II. Kahn TG, Dorafshan E, Schultheis D, **Zare A**, Stenberg P, Reim I, Pirrotta V, Schwartz YB. (2016). Interdependence of PRC1 and PRC2 for recruitment to Polycomb Response Elements. *Nucleic Acids Res. 2016 Dec 1;44(21):10132-10149.*

III. **Zare A**, Johansson AM, Karlsson E, Delhomme N, Stenberg P. (2018). The gut microbiome participates in transgenerational inheritance of low temperature responses in *Drosophila melanogaster*. *FEBS Lett. 2018 Oct 29. doi: 10.1002/1873-3468.13278. PMID: 30372516*

Abbreviations

ANT-C	Antennapedia complex
Ash1	Absent, small or homeotic 1
Br	bromodomain
BRE	TFIIB-recognition element
BX-C	Bithorax complex
CAF-1	Chromatin assembly factor 1
CAGE	Cap Analysis of Gene Expression
CBP	CREB-Binding Protein
CBX	Chromobox
CGIs	CpG islands
CH1	Cys/His-rich region 1
CH3	Cys/His-rich region 3
ChIP-chip	Chromatin immunoprecipitation followed by microarray analysis
ChIP-Seq	Chromatin immunoprecipitation followed by sequencing
CPF	Cleavage and polyadenylation factor
cPRC1	canonical PRC1
CREs	*cis*-regulatory elements
cryo-EM	cryo-electron microscopy
CStF	Cleavage stimulating factor
CTD	C-terminal domain
DEGs	differentially expressed genes
DSIF	DRB sensitivity-inducing factor
E(Z)	Enhancer of zeste
EED	Embryonic ectoderm development
GAF	GAGA factor
Gro-Seq	Global run-out sequencing
GTFs	General transcription factors
H2AK118ub	ubiquitylation of lysine 118 of histone H2A
H2AK119ub	ubiquitylation of lysine 119 of histone H2A
H3K4me	methylation of histone H3 at lysine 4
H3k9ac	acetylation of histone H3 at lysine 9
H3K9me	methylation of histone H3 at lysine 9
H3K20me	methylation of histone H3 at lysine 20
H3K27ac	acetylation of histone H3 at lysine 27
H3K27me	methylation of histone H3 at lysine 27
H3K36me	methylation of histone H3 at lysine 36

H4K8ac	acetylation of histone H4 at lysine 8
H4K20me	methylation of histone H4 at lysine 20
HATs	Histone acetyltransferases
HDACs	Histone deacetylases
HP1	Heterochromatin protein 1
Inr	Initiator
KATs	Lysine acetyltransferases
KIX	CREB-binding domain
M1BP	Motif 1 binding protein
MBDs	methyl-CpG binding proteins
MYC	Myc proto-oncogene protein
ncPRC1	non-canonical PRC1
ncRNA	noncoding RNA
NELF	Negative elongation factor
NFR	nucleosome-free region
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NPCs	Nuclear pore complexes
OPLS-DA	Orthogonal projections to latent structures Discriminant Analysis
ORFs	Open reading frames
PAS	poly(A) signal
PC	Polycomb
PCA	Principal Component Analysis
PcG	Polycomb group
PCGF	Polycomb group ring finger
PCL	Polycomb-like protein
Ph	Polyhomeotic
PHF1	PHD finger protein 1
Pho-RC	pho repressive complex
PIC	preinitiation complex
Pol I	RNA polymerase I
Pol II	RNA polymerase II
Pol III	RNA polymerase I
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
PREs	Polycomb Response Elements
PSC	Posterior sex combs
P-TEFb	transcription elongation factor b
PTMs	post-translational modifications

RING	Really interesting new gene
RNAP II	RNA polymerase II
rRNAs	ribosomal RNAs
SCE	Sex comb extra
Scr	*Sex comb reduced*
snRNAs	small noncoding RNAs
SU(VAR)3–9	Suppressor of variegation 3-9
Su(Z)2	Suppressor of zeste 2
SUZ12	Suppressor of zeste 12
TADs	Topologically associated domains
TBP	TATA-box binding protein
TEC	Transcription elongation complex
TF	Transcription Factor
Trl	Trithorax-like
TrxG	Trithorax group
TSS	Transcription start site

Introduction

1 Genome architecture

1.1 Nucleosomes

A fundamental characteristic of eukaryotic cells is that the strand of DNA in each chromosome, which may consist of several giga pairs of nucleobases (base pairs, bp), is bundled in chromatin together with structurally and functionally associated proteins and RNAs. Essentially, the DNA is folded around octameric proteins called histones, in structures first observed by electron microscopy and later called nucleosomes (Olins and Olins, 1973; Kornberg, 1974). Each nucleosome contains 145-147 bp of DNA wrapped almost twice around an octameric core consisting of pairs of H2A, H2B, H3 and H4 histones. Another histone subunit, histone H1, acts as a linker enabling packaging into higher-order structures. The packaging of DNA into chromatin is a key evolutionary difference between eukaryotes and prokaryotes, as it enables massive diversity in gene expression patterns and hence inherent biological complexity (Struhl, 1999). It also provides strong safeguards against spontaneous and inappropriate transcription (Lomvardas and Thanos, 2002). However, the significant advantages of the chromatin structure are accompanied by challenging needs for complex regulatory machinery to open the chromatin structure at appropriate times and places, as outlined below.

1.2 Chromatin types

The multistep packaging of chromatin allows the human DNA, which is nearly 2 meters long, to condense incredibly tightly into a volume about 5 micrometers wide. In mitosis (during prophase) the chromatin adopts its most compact form, now known as the chromosome. However, the DNA is not accessible in this form, so the chromatin must be decondensed to enable essential proteins to access it during processes such as transcription, replication and DNA damage repair. To grasp the decondensation process of chromatin, which needs to happen in the right places along chromosomes at the right times, we need to understand different types of chromatin. Classically, two types of chromatin were recognized, based on their compaction. The first type is heterochromatin; a tight and condensed form of chromatin formed by stacking of multiple nucleosomes. Heterochromatin is associated with repression of gene activities and the DNA within it is not easily accessible. However, heterochromatin contains many proteins, and both the proteins and DNA may have diverse chemical modifications, with various functions (not all known). Some of these

modifications are discussed further in this thesis. The other type of classically recognized chromatin is called euchromatin, which unlike heterochromatin is loosely packed and thus contains more accessible DNA and most active parts of the genome. The euchromatin structure has been likened to "beads on a string", as sets of nucleosomal particles, on average every 200 ±40 bp (Kornberg, 1974), are connected via linker DNA (Olins and Olins, 1973; Zlatanova and Leuba, 2003). However, more types of chromatin and intermediate states are now recognized. For instance, Filion *et al.* (2010) identified five types of chromatin based on genome-wide protein mapping in *Drosophila* cells, and Kharchenko *et al.* (2011) identified 9 or 30 states of chromatin in fruit flies (using different approaches) in higher-resolution genome-wide analyses.

Moreover, nucleosomes and chromatin structure generally are highly dynamic and constantly changing, in accordance with cells' needs to regulate various processes in time and space (Zlatanova and Leuba, 2003). These dynamic changes, sometimes referred to as nucleosome breathing, together with other reorganizations of chromatin, enable access of the most highly compacted DNA (Eslami-Mossallam, Schiessel and van Noort, 2016).

1.3 3D structure of chromatin

Apart from nucleosomes, which are the main physical factors involved in DNA compaction, chromatin also has structures at other orders. For example, chromatin can form loops connecting regions up to one megabase pairs apart and even regions located on different chromosomes. This non-random looping is essential for regulation of transcription in various cell types and organisms. During the past decade, many studies have examined the three-dimensional organization of chromosomes using methods such as chromosome conformation capture or, for short, 3C (Dekker *et al.*, 2002). A similar optimized method used for studying this 3D conformation is 'Hi-C', which combines chromatin crosslinking with fragmentations and proximity ligation followed by high-throughput sequencing (Lieberman-Aiden *et al.*, 2009). With experimental and computational advances, especially in sequencing techniques, new methodologies are being rapidly developed for this purpose, as reviewed by Schmitt, Hu and Ren (2016). Two types of structures generally observed (using different methods) in studies of the 3D conformation of chromosomes are topologically associated domains (TADs) (Nora *et al.*, 2012) and chromosome compartments A and B (Rao *et al.*, 2014). TADs are genomic regions with high levels of contact frequencies within them. TADs have taxa-specific sizes and positions, for instance in *Drosophila* they are smaller than in vertebrates (Pombo and Dillon, 2015). The "compartments" are higher-level structures: "A" compartments are regions of chromosomes that **are** relatively open and

accessible while “B” compartments are less transcriptionally active (Lee *et al.*, 2018) Finally, these compartments shape ‘chromosome territories’, the highest order of chromatin assemblies, dictating chromosomes’ localization in the nucleus (Cremer and Cremer, 2001). These structural features have been utilized in studies of gene expression, development, and diseases (Bonev and Cavalli, 2016). Several factors contribute to the 3D conformation of chromatin and its regulation, including architectural proteins such as Mediator and Cohesin, nucleosome remodelers such as the SWI/SNF family, reviewed by Tang, Nogales and Ciferri (2010), transcription factors (TFs) and non-coding RNAs (ncRNAs).

Although the 3D structure of the chromatin is generally stable, it has high dynamic capabilities allowing it to “jiggle”, thereby enabling tight regulation of the accessibility of specific DNA sequences during developmental stages and cell proliferation. Conversely, expression of some genes may change the 3D structure of the chromatin (Sevier and Levine, 2018).

1.4 Histone modifications

Up to 90% of the genomic DNA of eukaryotic cells is enfolded in nucleosome structures. Nucleosomes are physical barriers that bend the DNA and hinder gene transcription. In addition, a pioneering study by Allfrey and co-workers in the 1960s (Allfrey, Faulkner and Mirsky, 1964), and many other groups since then, have shown that several post-translational modifications (PTMs) of the histone residues modulate the chromatin’s accessibility. PTMs add to the complexity of gene regulation through reversible mechanisms providing the flexibility needed for cells to respond appropriately to environmental signals.
PTMs of histone tails, referred to as histone “marks”, orchestrate which genes should be turned on or off, active or repressed. Histones are marked and unmarked by a system of tightly regulated enzymes that remodel nucleosomes. Some of these enzymes act as “writers”, marking histones with chemical groups that are recognized by another set of “reader” enzymes, often as complexes of effector proteins that recognize and bind to the modified histones. Interestingly, writers can also often read, and reinforce, the modification marks (Zhang, Cooper and Brockdorff, 2015). Finally, “eraser” enzymes remove the marking when necessary (Gates, Foulds and O’Malley, 2017).

Most histone modifications occur at histones’ N-terminal residues, and thus are relatively accessible from the surface. The most common and widely studied histone modifications include acetylation, methylation, and phosphorylation (Luger *et al.*, 1997; Kouzarides, 2007; Lawrence, Daujat and Schneider, 2016). However, there are also several other types, including ubiquitination (addition of a ubiquitin group; a relatively large modification) and sumoylation (addition of

small ubiquitin-like modifiers). More modifications are being discovered and identified with increasing numbers of studies and use of modification-specific antibodies (Kebede, Schneider and Daujat, 2015). The development of chromatin immunoprecipitation techniques following by microarray analysis (ChIP-chip) and sequencing (ChIP-Seq) has enabled acquisition of extensive knowledge of histone modifications promoting or hindering gene expression (Heintzman *et al.*, 2007; Xhemalce, Dawson and Bannister, 2011). However, these techniques cannot detect different modifications of histones within a single nucleosome, for which a combination of mass spectrometry and top-down proteomic analysis is helpful (Macek *et al.*, 2006; Kouzarides, 2007). Over 100 types of histone modification have been detected to date, which are conserved from yeast to humans, and many more may be discovered (Gates, Foulds and O'Malley, 2017). Many histone modifications act indirectly by mediating recruitment of other proteins, including TFs and chromatin modifiers. However, some modifications can act directly and change the chromatin structure themselves (Akhtar and Becker, 2000; Shogren-Knaak *et al.*, 2006; Lu *et al.*, 2008; Lawrence, Daujat and Schneider, 2016). Some of the main types of modification are briefly discussed in the following sections.

1.4.1 Acetylation

One of the most common types of histone PTMs is acetylation, a process governed by histone acetyltransferases (HATs, also known as KATs) and histone deacetylases (HDACs), which respectively catalyze the addition and removal of acetyl groups. The acetyl group is negatively charged, thus its addition neutralizes positively charged nucleosomes, leading to decondensation of chromatin (Carrozza *et al.*, 2003; Chrun, Modolo and Daniel, 2017). Most histone modifications, including acetylation, occur on lysines (Azevedo and Saiardi, 2016). Acetylations on residues of histones H3 and H4 are usually linked to the active state of associated genes, partly due to the nucleosomes' neutralization reducing their instability and thus increasing accessibility for transcriptional machinery (Pradeepa, 2017).

Moreover, acetylation acts as a linker for binding of bromodomain, a common domain of many types of proteins. Upon recognition of an acetyl group, a bromodomain-containing protein will either act directly through other domains (e.g., chromatin remodeling domains), or recruit other proteins to the site. Acetylation and deacetylation are very rapid processes that can even occur within a single cell cycle. Two of the most well-known acetylation marks occur on the 9th and 27th lysines of histone H3 (H3k9ac and H3K27ac, respectively) and are often associated with sites in active promoters, especially H3K27ac.

1.4.2 Methylation

Histone methylation is slightly more complex than acetylation (and phosphorylation), but does not change nucleosomes' charge. It may be associated with gene activity or inactivity. There are three forms: mono-, di-, and tri-methylation, i.e. addition of one, two, and three methyl groups, respectively. The methylated amino acid residues are lysines (by 1-3 methyl groups) or arginines (by one or two methyl groups). The best characterized histone methylation marks occur on lysines of histone H3 at positions 4, 9 and 27 (H3K4me, H3K9me, and H3K27me). H3K4 tri-methylation is associated with gene activity and occurs around promoters of active genes while H3K36 di-methylation seems to occur within gene bodies (Suzuki, Murakami and Takahata, 2017). The other two mentioned lysine methylation marks on histone H3 (H3K9me2/3 and H3K27me1/2/3), as well as H4K20 tri-methylation, are associated with gene inactivity and are not restricted solely to promoter sites. They are often spread along genes' bodies and even into untranslated regions (Schotta *et al.*, 2004; Barski *et al.*, 2007; Wang *et al.*, 2008). However, while H3K20 trimethylation is linked with heterochromatin, H3K20me1 is often associated with active regions, including promoters or coding regions of active genes (Talasz *et al.*, 2005). Similarly, H3K9me3 has suggested enrichment in many active promoters despite its supposedly repressive nature (Squazzo *et al.*, 2006; Barski *et al.*, 2007).

While chromatin modifications such as acetylation and phosphorylation appear to be readily reversible, and removing the marks may generally result in an opposite outcome, histone methylation appears to be more stable (Bannister, Schneider and Kouzarides, 2002). Moreover, methyl groups in modified histone tails may promote binding to domains of various proteins. For example, Heterochromatin protein 1 (HP1) and Suppressor of variegation 3-9, SU(VAR)3–9, form a complex that binds to H3K9me (Schotta *et al.*, 2002; Zeng, Ball and Yokomori, 2010).

1.4.3 Phosphorylation

Like acetylation, histone phosphorylation is a more transient mechanism than methylation. It may have various outcomes depending on the residue that is modified and the cell cycle stage when it occurs (Barth and Imhof, 2010). Phosphorylation rates can change dramatically, e.g. following DNA damage or when the cell enters mitosis (Sawicka and Seiser, 2014). Phosphorylation has proven association with DNA condensation during mitotic events, but it is also linked with chromatin opening and thus promotion of transcriptional activation. Several phosphatases and kinases regulate this modification by respectively adding and removing phosphate groups.

1.5 CpG methylation

DNA methylation may occur in flies, according to some controversial indications (Dunwell and Pfeifer, 2014), but at most rarely. However, it often occurs in mammalian genomes, where methyl groups frequently bind to the fifth carbons of cytosines in CG dinucleotides in DNA strands (Bird, 1980). These nucleotide pairs are usually designated CpG, to distinguish them from CG nucleotide pairs in double-stranded DNA. A group of enzymes called DNA methyltransferases, Dnmts, catalyze this marking (Jabbari and Bernardi, 2004). CpG sites encompass about 1% of the genome in vertebrates; substantially less than the random probability (ca. 4%). Some studies suggest that the depletion of CpG dinucleotides in vertebrate genomes is due to side effects of cytosine methylation (Alexandrov *et al.*, 2013; Tomkova and Schuster-Böckler, 2018). DNA methylation is the most intensively studied type of DNA modification, which has vital functions in chromatin structure and nucleosome stability (at least in genomes where it definitely occurs). Its action mechanism has not been clearly elucidated, but there are two main hypotheses. One suggests that DNA methylation directly affects DNA-histone interactions by overwrapping DNA around histones or changing DNA's affinity for the histone proteins. The other suggests that DNA methylation serves as a mark for recruitment of Methyl-CpG binding proteins (MBDs), which then alter the chromatin structure (Collings, Waddell and Anderson, 2013). Either way, addition of a methyl group to CpG dinucleotides usually results in repression of the neighboring genes, which is essential in many processes of vertebrate cells such as X chromosome inactivation, gene imprinting and tissue-specific gene regulation during development (Jones, 2012). Unsurprisingly, any unwanted alteration in DNA methylation can cause gene deregulation, leading to a variety of diseases in humans (Robertson, 2005).

CpG dinucleotides are particularly abundant at some genomic sites (usually 200 to 1,000 bp long) known as CpG islands or CGIs. Since their discovery, many studies have shown that CGIs are associated with transcription initiation, especially for housekeeping genes. In mammals, CGIs are associated with promoters of about 70% of the genes in total, generally including those of housekeeping genes (Saxonov, Berg and Brutlag, 2006; Long *et al.*, 2013). It is estimated that the human genome has 28 million CGIs (Stevens *et al.*, 2013). Mapping of all human and mouse CpG islands has shown that more than half of them, so-called "orphan" CGIs, are located remotely from known promoters. Orphan CGIs have characteristics of functional promoters but are often methylated during development (Illingworth *et al.*, 2010). Nevertheless, Cap analysis of gene expression (CAGE) (Shiraki *et al.*, 2003) and Global run-out sequencing (Gro-Seq) (Core, Waterfall and Lis, 2008) have provided evidence of transcriptional initiation in nearly 40% of orphan CGIs. Moreover, trimethylation of histone H3 at lysine 4 (H3K4me3), a hallmark of active genes, is frequently

found at these sites. Collectively the evidence suggests that orphan CGIs are associated with transcription of genes that are activated tissue-specifically and potentially only under certain conditions (Illingworth *et al.*, 2010; Deaton and Bird, 2011).

2 Transcription

Transcription of DNA is deployed and regulated in four dimensions of space and time. For example, locations of active genes in the three spatial dimensions of both chromatin fibers and nucleic space may change during development or in response to environmental stimuli (Papantonis and Cook, 2010). In addition, the timing of transcription is controlled through continuous adjustable processes influenced by internal and external stimuli. Briefly, transcription is the process whereby complementary RNA (ribonucleic acid) molecules are generated from DNA (deoxyribonucleic acid) sequences. The main enzymes involved in this process are RNA polymerases. Transcription in both eukaryotes and prokaryotes involves three main steps: initiation, elongation, and termination. However, this thesis focuses solely on transcription in eukaryotes, particularly *D. melanogaster*.

2.1 Initiation of transcription

The first step in transcription is start site selection, in which key proteins bind to DNA at the site where transcription will begin (Smale and Kadonaga, 2003), but before that the chromatin must be opened, as described in the following section.

2.1.1 Opening of chromatin

Chromatin is opened partly through neutralization of DNA by an enzyme called DNA acetyltransferase (HAT) (Brownell et al., 1996; Carrozza et al., 2003). This neutralization reduces histones' affinity for DNA strings and may result in nucleosome remodeling. However, the remodeling alters the relative positions of histone molecules along the DNA string. Nucleosomes have dynamic and mobile characteristics that are dependent on the DNA sequence and promoted by histone modifications (Li et al., 2004). The DNA sequence, to some extent, encodes the nucleosome positioning (Segal et al., 2006). Moreover, the sequence's basal properties determine the flexibility of the double helix and thus its capacity for bending around histone octamers, which can cause up to 1000-fold variations in DNA-histone affinities (Thåström et al., 1999).

Important sets of enzymes, mentioned above, are chromatin remodelers. These are all the enzymes involved in the dynamic process of nucleosomal arrangements, including placement, displacement and removal of nucleosomes. Blossey and Schiessel (2018) have recently presented a nice analogy of chromatin remodelers as "clip-on" motors of eBikes, which can engage or disengage the bikes' chains to regulate their velocity. Similarly, chromatin remodelers can bind to nucleosomes and pull on the DNA string, thereby triggering nucleosomal reconfiguration or rearrangement. Chromatin remodelers or other proteins

involved in transcription initiation must identify and interact at proper genomic sites to trigger appropriate changes in nucleosomal positioning, initially dictated by the DNA sequence characteristics. An important question is whether the DNA sequence governs which proteins bind and where they bind, or the remodelers override the genetic code and govern nucleosomes' re-positioning. An attractive hypothesis is that the system's state is governed by equilibrial effects of nucleosomes and chromatin remodelers competing for certain binding sites along DNA strings (Segal *et al.*, 2006).

2.1.2 Recruitment of Pol II

Unlike bacteria, in which one type of polymerase transcribes all active genes into RNA, three classes of polymerases are involved in transcription in eukaryotic cells: RNA polymerase (Pol) I, II and III. Pol I and Pol III are associated with transcription of ribosomal RNAs (rRNAs) and small noncoding RNAs (snRNAs), respectively (Khatter, Vorländer and Müller, 2017). **Pol** II, which has 12 subunits, catalyzes formation of mRNA from DNA, and thus transcribes all protein-coding genes, but also some ncRNA genes.

In eukaryotic cells, a multicomponent system known as transcriptional machinery drives the complex process of transcription. This large consortium includes several multiunit proteins that work in concert with Pol II and several TFs as well as many other complexes that serve as coregulators. In humans, to date, we have identified more than 100 proteins that participate in transcription, from initiation to elongation and termination. Transcription starts with assembly of a preinitiation complex (PIC), consisting of Pol II and General Transcription Factors (GTFs TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH), which are coupled together at promoters of active genes. TFIID has a subunit known as TATA-box binding protein (TBP) that recognizes the core promoter and creates an anchor for PIC formation. After that, Pol II and other GTFs are recruited to the site. DNA is unwound by the helicase subunit of the TFIIH transcription factor, resulting in an open structure of single-stranded DNA close to the transcription start site (TSS) (Fishburn *et al.*, 2015). Another essential player in transcription initiation is a multiunit protein known as Mediator (Kelleher, Flanagan and Kornberg, 1990; Flanagan *et al.*, 1991), which co-activates recruitment of the initiation complex and stabilizes it by binding to Pol II, TFIID and TFIIH. It then participates in further stages of transcription, from elongation to production of mature mRNAs (Conaway and Conaway, 2011; Plaschka *et al.*, 2015).

Transcription initiation was classically regarded as a highly ordered process, involving a sequence of steps that had been elucidated through many years of painstaking studies. However, advances in single particle cryo-electron microscopy (cryo-EM) leading to discoveries of new subunits and new models

suggest that we should not regard the players as discrete, tightly bound complexes with clearly defined action mechanisms (Kandiah *et al.*, 2014; Gupta *et al.*, 2016; Hantsche and Cramer, 2017).

Many proteins and complexes contribute to the sophisticated process of transcription. However, the first actors assemble at tiny genomic regions of DNA that have unique (but diverse) binding sites, called core promoters, for recruitment of all the proteins and cofactors involved in transcription. Each of these sites is centered at a TSS and extends about 40 bp in both directions, providing foundations for recruitment of many members of the transcriptional machinery. Several signs in the genome reveal the presence of promoters, including openness of chromatin, certain histone marks (e.g. H3K4me3 or H3K9ac) and presence of PIC (Haberle and Lenhard, 2016). With advances in single nucleotide resolution sequencing techniques, such as CAGE (Kodzius *et al.*, 2006) and GRO-Seq (Gardini, 2017), we can now map TSSs and associated promoters precisely (Core *et al.*, 2014). Various motifs are found in core promoters, such as TATA-boxes (AT-rich DNA sequences) and Initiator (Inr). TFIID binds nearly all identified core promoter elements during transcription, but not the TFIIB-recognition element (BRE) (Lagrange *et al.*, 1998). It has been shown that some of these elements are associated with specific biological functions. For instance, the TCT motif reportedly plays a key role in transcription of ribosomal protein genes in *Drosophila* (Parry *et al.*, 2010). However, no such links have been clearly defined for many others. Moreover, although these elements are very diverse and abundant in the genome, many promoters do not have any known motifs. Unsurprisingly, nucleosome positioning is related to the presence of promoters. Two nucleosomes, commonly known as +1 and -1 nucleosomes, surround the nucleosome-free region (NFR) at the beginning of a gene. Nucleosome formation seems to be driven by specific sequence patterns, implying that promoter characteristics dictate nucleosome positioning. However, evidence from studies of mammalian cells, based on different types of promoter, suggests that more than one mechanism may be involved in nucleosome positioning at promoters (Valouev *et al.*, 2011).

2.1.3 General transcription factors

As already stated, GTFs are proteins that participate in regulation of gene transcription. To date, six GTFs have been identified and extensively studied. TFIID recognizes the TATA-box motifs of promoters, binds via its TBP subunit, and thus is the main DNA-binding protein. However, TFIIB also participates in promoter selection and DNA binding. TFIIH has a helicase subunit that unwinds the DNA and exposes it to Pol II. TFIIE and TFIIF are responsible for recruiting other players to the PIC (TFIID and Pol II, respectively). Lastly, TFIIA does not seem to be associated with PIC formation but rather seems to stabilize the binding

of TFIID to DNA through disassociation of TFIID/TBP dimers. This facilitates TFIID loading onto the promoter (Coleman *et al.*, 1999). While these are general TFs, hundreds of others, some of which are cell-type-specific, are also involved in transcription initiation and promotion of PIC formation. For further information about TFs (especially in humans), readers are referred to a recent review by Lambert *et al.* (2018).

2.1.4 Enhancers and negative regulators

Vital components of gene regulatory networks include short ncDNA sequences known as *cis*-regulatory elements (CREs) that are usually found in the vicinity of genes and regulate their transcription. Promoters and enhancers are examples of CREs that are located on the same chromosome as the genes they regulate. Enhancers are short sequences (from 100 to a few hundred bp) that may control both neighboring genes and (unlike promoters) genes located up to 1 Mbp away. This enables regulation of genes by more than one distal enhancer, and hence greater variability in gene regulation. Distal enhancers can interact with promoters by looping to the target genes (Levine, Cattoglio and Tjian, 2014).

Like promoters, enhancers have DNA motifs for binding activators including TFs. Since enhancers have high affinity for histone octamers, clusters of TFs often compete with these transcription barriers and thus facilitate chromatin remodeling via activation of functional enhancers (Spitz and Furlong, 2012). However, this is often accompanied by recruitment of cofactors (to enhancers or promoters), which facilitate transcription, but unlike TFs do not have specific binding sites (Weake and Workman, 2010). Coactivators often have chromatin remodeling capacities, sometimes through histone modifications (e.g., p300/CBP), or facilitate the action of transcriptional machinery at promoters (e.g., Mediator) (Long, Prescott and Wysocka, 2016). Silencers, another class of CREs, act antagonistically to enhancers, suppressing transcription by recruiting proteins known as repressors. Like enhancers, silencers are usually located upstream of TSSs, but may be located downstream of them in 3' UTRs (untranslated regions) or introns of gene bodies (Maston, Evans and Green, 2006).

In a metazoan cell, activators and repressors acting together in the gene regulatory network dictate the transcriptional status of thousands of genes. In addition to thousands of TFs, hundreds of cofactors (coactivators and corepressors) that also assist the transcriptional machinery have been identified. These transcriptional coregulators include CREB-binding protein (CBP), which is encoded by a multigene family in *Drosophila*, and its mammalian paralog P300. The main characterized function of CBP/p300 is acetylation of several lysine residues of histones H3 and H4, as well as specific lysines of other

regulators, via HAT (mainly lysine acetyltransferase; KAT) activity (Dancy and Cole, 2015). However, as well as a HAT domain, CBP/p300 proteins have several other functional protein-binding domains, such as a CREB-binding domain (KIX), Cys/His-rich regions 1 and 3 (CH1 and CH3), and a bromodomain (Br). In addition to histone PTMs, CBP also acts as a hub for recruitment of many TFs and other cofactors to the promoter and enhancer regions. To date, over 400 proteins have been identified that network with CBP via protein-protein interactions or interact functionally (Bedford *et al.*, 2010). Hence, CBP is a key player in gene regulation and cell development, so CBP deficiency is lethal in the embryonic stage in flies (Goodman and Smolik, 2000). Unsurprisingly, CBP and p300 are only found in higher organisms, and their presence or absence seems to be indicative of an organism's complexity. Relatively simple eukaryotes, such as *D. melanogaster*, only have one CBP ortholog (known as *nejire* in flies), while yeasts have no CBP/p300 family members. In mice, more than 90% of histone H3 acetylations at lysines 18 and 27 are associated with CBP/p300. Together with histone acetylation at H3K18 and H3K27, CBP has been used for mapping active promoters and enhancers (Jin *et al.*, 2010; Kasper *et al.*, 2014). Although CBP binds to active promoters and enhancers, in flies it can also bind to inactive enhancers as well as insulators, suggesting it has even broader functions in chromosomal interactions (Philip *et al.*, 2015).

2.2 Promoter-proximal pausing

For many years it was believed that gene regulation occurs mainly in Pol II initiation stages when Pol II is recruited, and a PIC is formed. Thanks to recent technological advances, it is now clear that there is another critical checkpoint for gene regulation shortly after transcription initiation and before productive elongation (Mayer, Landry and Churchman, 2017). After Pol II has catalyzed synthesis of 10-12 nucleotides of a nascent RNA, it escapes from the promoter region and initiates transcription elongation. However, a pause occurs about 30-50 bp downstream of the TSS, known as promoter-proximal pausing. The main contributors to promoter-proximal pausing are negative regulators such as Negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF). These factors induce and stabilize the pausing state via interaction with Pol II. However, negative elongation factors are not the only players promoting Pol II pausing. Certain modifications of Pol II, some features of the underlying DNA sequence, and even the nascent RNA can contribute to it. A high GC content may induce pausing and in mammals CpG island promoters are associated with high Pol II pausing (Kellner, Bell and Vertino, 2015). Moreover, some DNA sequences may cause Pol II to "backtrack", i.e. move backwards (Cheung and Cramer, 2011). More interestingly, nascent RNA can also induce pausing, partly via RNA hairpins limiting the movements of Pol II on DNA (Core, Waterfall and Lis,

2008). Additionally, an increasing number of discovered TFs are associated with promotion of Pol II pausing, e.g. GAGA factor (GAF) and Motif 1 binding protein (M1BP) in *D. melanogaster* (Li and Gilmour, 2013).

Pol II pausing does not only occur at promoter-proximal regions, and variations in Pol II distribution have been detected along gene bodies, showing accumulation of Pol II (peaks) due to its paused state. In many metazoans, polymerase pausing seems to be elevated at core promoters, exon-intron junctions and 3' polyadenylation sites, and the rate of Pol II pausing is much higher in promoter regions than in gene bodies (Kwak *et al.*, 2013; Day *et al.*, 2016). The ratio between these rates is called the pausing index, which varies among genes and cell types, and is often used for distinguishing "paused" genes from the others. Nevertheless, pausing should be regarded as a general feature of the transcription cycle and not restricted to some genes or specific regions in gene bodies. The primary outcome of Pol II pausing is provision of opportunities for gene regulation in response to internal and external stimuli, as well as fine-tuning of gene transcription in cell differentiation or different stages of development.

2.3 Release from pausing

After promoter-proximal pausing, Pol II must be released from the paused state. The main protein mediating this release is Positive transcription elongation factor b (P-TEFb), which in addition to Pol II also phosphorylates NELF and DSIF. Phosphorylation of NELF results in dissociation from Pol II, whereas phosphorylation of DSIF transforms it into an elongation-promoting state (Peterlin and Price, 2006). Hence, P-TEFb is a key regulator in early elongation. Transcription pausing of highly expressed genes has been detected, so Pol II pausing and release from the paused state are key regulatory steps in the transcription of bodies of active genes (Jonkers and Lis, 2015). Like many processes in the nucleus, release from pausing does not always proceed linearly. DNA looping can bring enhancers close to promoters and thus promote P-TEFb recruitment by binding other cofactors such as Mediator (Allen and Taatjes, 2015). Interestingly, this looping, mediated by the Cohesin protein complex, has been detected in genes showing high levels of Pol II pausing (Schaaf *et al.*, 2013; Ghavi-Helm *et al.*, 2014). Some DNA-binding TFs are also associated with release from pausing, of which Myc proto-oncogene protein (MYC) and Nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) are the best known examples that interact with P-TEFb (Barboric *et al.*, 2001; Rahl *et al.*, 2010).

2.4 Elongation

Elongation is essentially the extension of transcripts produced by RNA Pol II. However, this is not a straightforward process, and many factors may affect gene bodies' transcription rates. Generally, elongation is relatively slow during transcription of the first part of a gene body, and accelerates up to 4- to 10-fold, depending on the type and condition of the cell, after the first ca. 15 kbp (Danko *et al.*, 2013). The slow start ensures engagement of all the factors required for productive elongation and is crucial for some events such as splicing. We now know that certain gene features, as well as the presence of exons or polyadenylation sites, can slow elongation rates while other factors (such as H3K79me2, H4K20me1, and ubiquitylation of histone H2B) seem to accelerate it (Jonkers, Kwak and Lis, 2014; Veloso *et al.*, 2014). In contrast to these histone methylation marks, tri-methylation of H3K36 and H3K9 is associated with reduction in elongation rates (Jonkers and Lis, 2015).

2.5 Termination and nuclei export

Ending gene transcription at the appropriate point is as important as its correct initiation. *Inter alia*, termination determines the fate of the produced transcript, e.g. whether a mRNA is stabilized and transported into the cytoplasm, or degraded. Since transcription by polymerases is widely spread throughout the genome, and not limited solely to open reading frames (ORFs), tight control of termination is essential for proper expression of neighboring genes. Termination processes differ between eukaryotes and prokaryotes, and the processes catalyzed by the three eukaryotic polymerases also vary. Here, the termination of Pol II transcription in mRNA production is briefly discussed, but there are many similarities in the processes and subunits involved in the termination of transcription catalyzed by all three types of eukaryotic polymerases.

Pol II has a C-terminal domain (CTD), which is heavily modified during transcription. One well-studied CTD modification is phosphorylation, which mediates recruitment of factors needed for processing the pre-mRNA: Cleavage and polyadenylation factor (CPF) and Cleavage stimulating factor (CStF). Phosphorylation of Ser5 during transcription initiation, which occurs at promoters with the association of TFIIH, is necessary for recruitment of 5' capping machinery (Komarnitsky, Cho and Buratowski, 2000). Ser2 phosphorylation, in contrast, occurs during elongation and results in recruitment of 3' processing factors (e.g., CPF) to poly(A) sites (Ahn, Kim and Buratowski, 2004). CPF has an endonuclease subunit that is responsible for cleavage of pre-mRNA. Another CPF subunit then adds a poly(A) tail to the new 3' end of the pre-

mRNA. This polyadenylation is essential for mRNA stability and facilitates the mRNA's exit from the nucleus into the cytoplasm (Schreieck *et al.*, 2014) .

There are at least two hypotheses regarding the mechanism that terminates Pol II-mediated transcription at the poly(A) signal (PAS) on the DNA strand. The first, known as the "allosteric" (or anti-terminator) model, suggests that conformational changes in Pol II triggered when it senses the PAS region induce termination. However, Pol II continues to transcribe nascent RNA downstream of the PAS. Therefore, after RNA cleavage by CPF there will be a new 5' end, which is not protected by capping. According to this and the rapid degradation of the newly synthesized RNA, a second hypothesis, known as "torpedo", proposes that the region of RNA transcribed downstream of the PAS is recognized and digested by an exonuclease enzyme that somehow causes dissociation of Pol II from DNA, and hence termination. However, the two presented models are not necessarily exclusive, and a combination of both may better explain the complex process of transcription termination (Loya and Reines, 2016).

The process discussed above is known as PAS-dependent termination. However, there are also other mechanisms, such as transcriptional arrest (where Pol II is degraded rather than recycled), and a process involving backtracking of Pol II, enabling its use as a proofreading tool (Proudfoot, 2016). Finally, after the pre-mRNA has been released from the transcription elongation complex (TEC), which consists of Pol II, DNA and nascent RNA, it undergoes further modifications, including splicing, that complete its maturation. Then, the mature mRNA is ready for release from the nucleus into the cytoplasm, where it will be translated and eventually degraded. Its translocation is mediated by multiunit proteins forming nuclear pore complexes (NPCs), and is considered to be another major regulatory step in gene expression (Kubitscheck and Siebrasse, 2017).

3 Transcription Maintenance

Gene regulation is vital for both maintenance of cells' homeostasis and appropriate responses to various conditions. Moreover, their gene expression status must often be remembered and passed on during proliferation and developmental stages. All somatic cells in a given eukaryotic organism have copies of the same DNA, raising questions about how cells become different from each other and remember their identities. This brings us to the fast-growing and captivating subject known as epigenetics. In Latin, the prefix epi- means above, so the term epigenetics implies phenomena "on top" of genetics, or additional to the traditional genetic basis of inheritance. Today we know this involves "epigenetic marks", which are indirect modifications of chromatin that dictate cells' fates, determine which genes are switched on or off, and whether specific genes will continue to be expressed or silenced. Collectively, these states are referred to as the "transcriptional status", "transcriptional pattern" or "transcriptional profile" of a given gene, tissue, cell line or organism.

Probably the best-known pioneer of epigenetics, who originally proposed the concept of epigenetics, is Waddington (1942). He used this term to explain the occurrence of phenotypic variability from a single genome. Today, definitions of epigenetics are widely controversial, but overall it refers to all mitotically or meiotically heritable changes related to gene expression that are independent of direct changes in underlying DNA sequences (Deans and Maggert, 2015). Mechanisms involved in the inheritance of epigenetic information, such as modifications of DNA and histones, are debated, but DNA methylation, histone modifications, and processes involving ncRNAs are among the most intensively studied suggested mechanisms.

Almost all cells in the body have the same DNA, but not all cells are the same, because cells use the genomic codes in different ways. Genes may be changed in various ways during developmental stages and following exposure to various environmental factors (often simply called "exposures") and switched on or off at various times depending on their type and function. A cell's combined set of all these changes is called its epigenome. If we think of a cell as a computer, while DNA is its hardware, the epigenome is the software telling the hardware what to do. The epigenome is the secondary structure of chromatin consisting of DNA, proteins (including histones), chemical tags (i.e. DNA methylation and histone PTMs), and ncRNAs. In contrast to DNA, which remains fixed for life (unless mutations or DNA damage occur), the epigenome is a flexible construction that adjusts states of specific genes in response to rapid changes in the cellular environment. Moreover, life style conditions including nutrition, and some

activities such as smoking (in humans) can also influence the epigenome (Kanherkar, Bhatia-Dey and Csoka, 2014).

3.1 Maintenance through cell division

When cells divide, the daughter cells must remember their parental cells' transcriptional patterns and preserve their identity. While this mechanism has not been fully elucidated, some studies suggest that the DNA's organization in the nucleus is involved. Therefore, it is suggested that the 3D packaging of the DNA partly determines genes' expression status (Poleshko *et al.*, 2017). Accordingly, important characteristics of a genomic region and important elements of epigenetic information include the pattern of histones' PTMs. These histone modifications are involved in regulation of each genomic region's activities, including their transcriptional status. Classically, it was suggested that the histones are split between daughter cells (each receiving H3-H4 dimers from H3-H4 parental tetramers) or randomly distributed during cell division. We now know that during DNA replication almost all the histone modifications are erased when chromatin is disassembled (Budhavarapu, Chavez and Tyler, 2013), raising questions about how daughter cells regain their epigenetic information after cell division. In other words, how are parental histone modifications remembered during replication? While the exact mechanism whereby cells remember local-specific histone modifications is still a mystery, some evidence suggests involvement of histone chaperones in recycling the modified histones (Hammond *et al.*, 2017). One of these histone modifiers, Chromatin assembly factor 1 (CAF-1), is putatively a key player in inheritance of H3K9me3 marks (Liu *et al.*, 2012). Another suggested mechanism for inheritance of histone modification involves a "pre-replicative boost". For instance, Polycomb group (PcG) proteins and their associated repressive mark (H3K27me3) as well as H3K4me3 reportedly accumulate at Polycomb response elements (PREs) in early mitosis during S-phase. This accumulation declines later in S-phase upon PRE replication, suggesting an inheritance mechanism for H3K27me3 involving diluted retention of this mark upon cell division in TSSs (Lanzuolo *et al.*, 2011; Lanzuolo, Lo Sardo and Orlando, 2012; Budhavarapu, Chavez and Tyler, 2013). However, mitotic inheritance of epigenetic information does not always involve histones. For example, in *D. melanogaster* embryos PcG proteins reportedly remain bound to the DNA during DNA replication and can carry epigenetic memory in the absence of histones (Francis *et al.*, 2009; Lo *et al.*, 2012).

3.1.1 Polycomb group proteins

Elements of one the most crucial epigenetic silencing systems, which play a key role in maintenance of cells' identity during differentiation and development, are the Polycomb group (PcG) proteins. Core units of these multi-unit complexes are conserved in organisms ranging from yeasts through flies to humans, and they are responsible for maintaining the repressed status of key developmental genes (*Hox* gene clusters). PcG complexes have two major parts known as Polycomb repressive complexes 1 and 2 (PRC1 and PRC2), and a DNA-binding part (known as pho Repressive Complex, PHO-RC, in flies). This group of proteins acts cooperatively but antagonistically with other proteins known as Trithorax group (TrxG) proteins in maintenance of genes' expression status. Many cellular processes, such as X-chromosome inactivation and gene imprinting, require proper functionality of these two sets of complexes, and any deficits may lead to various diseases including cancer (Poynter and Kadoch, 2016). Most subunits of PcG proteins have been discovered, and studied, in *D. melanogaster*. The name "polycomb" comes from observations of phenotypic changes in mutant flies caused by overexpression of the *Sex comb reduced (Scr)* gene, and as the name Polycomb implies, additional, ectopic sex combs develop on the legs of adult male flies (Lewis, 1947, 1978). While PcG proteins in mammals share many similarities to those of flies, they are more diverse, especially in the PRC1 complex (Blackledge, Rose and Klose, 2015). PcG protein complexes and their subunits have received much attention in recent years and our understanding of their action mechanism is rapidly increasing.

3.1.1.1 Polycomb response elements (PREs)

Relatively small sequence-specific DNA elements known as Polycomb response elements (PREs) usually recruit polycomb group proteins. These regions are enriched in many DNA motifs responsible for recruitment of DNA-binding parts of polycomb complexes. PREs were first discovered in studies of transgenic effects of regulatory Bithorax complex genes in *D. melanogaster* (Müller and Bienz, 1991). Flies were also used in the first characterizations of these elements. Hundreds of PREs have also been detected in mammals, but while showing similarities to their counterparts in flies, they have some significant differences. Notably, what we call "true PREs" are only present in *Drosophila* and our understanding of PRE-like elements in mammals is evolving with emerging studies in the field (Bauer, Trupke and Ringrose, 2016; Cameron *et al.*, 2018). Mammalian PREs, also known as PRE-like regions, target many genes such as developmental genes and genes involved in cell proliferation, and are also engaged in various processes including X-chromosome inactivation. While PREs in mammals are much more diverse than those of flies, they have similar target genes such as *Hox* genes (which control segmentation of embryos along their anterior-posterior axis), and signaling pathways including morphogenetic and

developmental pathways (Ringrose, 2007). PcG- and TrxG-proteins are highly conserved in flies and mammals, and there are remarkable similarities in genes that are controlled by PREs in both taxa (Sharif *et al.*, 2013). In mammals, these regions are often located in hypomethylated CGIs, but as yet no clear DNA motif has been associated with them (Bauer, Trupke and Ringrose, 2016). In fruit flies, on the other hand, several short motifs have been characterized in PREs, which vary both among PREs and in the same PREs of different species (Kassis and Brown, 2013; Bauer, Trupke and Ringrose, 2016). In flies, PREs are independent functional elements, meaning that they can be inserted anywhere in the genome and still recruit PcG proteins. These *cis*-regulatory elements act co-operatively with enhancers and promoters and can establish an active or repressed state of the genes. Moreover, PREs may maintain a gene's expression status with the ability to switch upon receipt of new signals.

3.1.1.2 PRC2

The PRC2 complex is so important that it has been called "a guardian of cell fate" (Holoch and Margueron, 2017). It contains three essential subunits in its core: a SET domain, Suppressor of zeste (SUZ12), and Embryonic ectoderm development (EED). In flies, the SET domain has histone methyltransferase activity through Enhancer of zeste, E(Z), which can add one to three methyl groups to histone H3 at lysine 27 (H3K27me1/2/3) (Müller *et al.*, 2002). In mammals, homologous protein subunits EZH1 and EZH2 execute addition of these repressive marks. Binding of EED to the tri-methylated histone may putatively contribute to self-propagation of H3K27me3 in maintenance of a repressive status of the chromatin and transmission of the histone mark to daughter cells (Margueron *et al.*, 2009). Thus, EED promotes positive feedback mediating persistence of specific transcriptional profiles. Several other accessory proteins also contribute to the PRC2 core's enzymatic activity. In *Drosophila,* Polycomb-like protein (PCL) contributes to maintenance of a repressed state by stimulating high levels of H3K27me3 deposition on Polycomb target genes (Nekrasov *et al.*, 2007). Similarly, its homolog in mammals, PHD finger protein 1 (PHF1), is required for efficient trimethylation of H3K27, but (interestingly) not for either its mono- or di-methylation (Sarma *et al.*, 2008). Apart from the mentioned core subunits, there may be additional subunits of PRC2, enabling creation of alternative forms of the complex. For a review of these "facultative" subunits and their action mechanisms see Holoch and Margueron (2017).

3.1.1.3 PRC1

While PRC2 mediates addition of 1-3 methyl groups to H3K27, PRC1 ligates a ubiquityl group to histone H2A at lysine 118, creating the H2AK118ub mark (H2AK119ub in mammals). As already mentioned, mammalian PRC1 complexes are more diverse because there are multiple variants of some subunits, and

corresponding variations in PRC1's functionality, as a transcriptional repressor or even activator (Gil and O'Loghlen, 2014). In *Drosophila*, PRC1 contains Polycomb (PC), Polyhomeotic (PH), Posterior sex combs (PSC) or its paralog Suppressor of zeste 2 [Su(Z)2], and RING (also known as Sex comb extra, Sce). PC recognizes and binds to the H3K27me3 mark, while RING catalyzes ubiquitylation (Chittock *et al.*, 2017) . Mammalian PRC1 variants have several homologs for each subunit and historically were divided into two major categories: canonical PRC1 (cPRC1) and non-canonical PRC1 (ncPRC1). This classification resulted from paucity of knowledge of PRC1 at the time (Vidal and Starowicz, 2017). Today, despite numerous advances in the field, there are still many 'unknowns' about these complexes' biochemical properties and action mechanisms. Briefly, cPRC1 complexes are like PRC1 in flies, with homologs for the main components, while ncPRC1 refers to complexes with variations in Chromobox (CBX) (PC homolog) and PHC subunits (Bajusz *et al.*, 2018). However, both categories contain the Polycomb group ring finger (PCGF) catalytic subunit. For further information on the increasing numbers of PRC1 variants in mammals, see Blackledge, Rose and Klose (2015).

3.1.1.4 Trx – Ash1

Like PcG proteins, TrxG proteins contribute to specification of segmental identity and are crucial for proper development. As already mentioned, these two groups of proteins have opposing effects on expression of *Hox* genes. In *Drosophila*, *Hox* genes comprise two groups of homeotic genes: Bithorax complex (BX-C) and Antennapedia complex (ANT-C) genes, which encode homeodomain transcription factors and thus direct the fate of different body segments through development (Gellon and McGinnis, 1998). Knocking out these genes in flies results in homeotic transformations, where body parts do not develop in their regular segments. The expression of *Hox* genes must be precisely regulated, and any deficits may lead to drastic alterations in cell fate, which emphasizes the importance of PcG and TrxG proteins (Kassis, Kennison and Tamkun, 2017).

Unlike PcG proteins, the TrxG complex promotes maintenance of active gene transcription (counteracting PcG-mediated repression) by a Set domain catalyzing methylation of H3K4 or H3K36 (Geisler and Paro, 2015). While this might initially seem a simple ON-OFF switch, there is much more complex interplay between these two groups of proteins (Schuettengruber *et al.*, 2017). TrxG terminology is based on the first discovered gene in this cluster, *Trithorax*. Many protein-coding genes were previously identified as TrxG members, based on similarities of observed phenotypic effects of their mutation and other criteria such as biochemical activity, sequence homology and similarity of effects on the transcription of *Hox* genes. However, only two proteins, Trithorax and Absent, small or homeotic 1 (Ash1), seem to be true members of TrxG complexes, based

on their PRE-binding profile (Kingston and Tamkun, 2014; Kassis, Kennison and Tamkun, 2017). This view might change later, since our understanding of the interplay between PcG and TrxG complexes is rapidly growing with continuing advances in the field (see Geisler and Paro, 2015; Schuettengruber *et al.*, 2017).

3.2 Maintenance through generations

For many decades it was thought that no changes in somatic cells are transferred to gametes, because of a barrier between somatic and germline cells called the Weisman barrier (Weismann, 1892). Today, we know this is not entirely true and some epigenetic changes may be passed through gametes and sometimes sustained for several generations (Sharma, 2013). We know very little about the mechanisms involved, but the changes may be triggered by endogenous events (e.g. processes during cell differentiation) or responses to exogenous environmental conditions. However, the acquired epigenetics traits must be removed for zygotes to be pluripotent and totipotent. This reprogramming of the germline is not perfect, and when it fails epigenetic marks may be transmitted to the next generation, but generally at much lower rates than genetic information.

3.2.1 Epigenetic reprogramming

Epigenetic changes do not have high chances of transfer to the next generation, as a simple mechanism called epigenetic reprogramming erases most epigenetic marks in early embryogenesis. One of the mechanisms involved in this reprogramming in mammals is DNA demethylation, which occurs during early developmental stages (Morgan *et al.*, 2005; Seisenberger *et al.*, 2013). After fertilization and during gametogenesis and embryogenesis, almost all the parental DNA methylation is erased. Then, during embryogenesis a wave of DNA methylation occurs that re-establishes the repressive marks, resulting in global gene silencing. Notably, CpG islands are protected from remethylation, thereby safeguarding expression of housekeeping genes in all cell types (Cedar and Bergman, 2012). The mechanism involved in protection of CGIs is not very clear. However, some sequence motifs in mice (e.g. Sp1) play a role in protecting CpG islands from remethylation (Macleod *et al.*, 1994), and some epigenetic information is transmissible and may persist for several generations.

3.2.2 Genomic imprinting

Epigenetic marks, such as DNA methylation, are not entirely reset in all genes, and some genomic regions may escape this reprogramming partly or entirely. Imprinted genes and retrotransposable elements are classical examples of sequences that are partially protected from the second round of demethylation in epigenetic reprogramming. This protection allows offspring to maintain some of

the epigenetic modifications of their parents. In diploid organisms, parental imprints are maintained in one allele at a given locus, and erased in the other allele during gamete formation. This monoallelic silencing through erasure of epigenetic marks (e.g., DNA methylation) occurs in an allele-specific fashion, and either the maternal or paternal imprints will be re-established, depending on the gene (Barlow and Bartolomei, 2014). Although genomic imprinting involves monoallelic silencing at certain loci, it is not necessarily a silencing mechanism. Rather, it is considered a *cis*-regulatory tool that may introduce variation in parental-specific gene expression at various gene regulatory steps. It is also a valuable phenomenon for studying epigenetic regulation because both parental alleles are within the same nucleus and thus exposed to identical environmental conditions (Henckel and Arnaud, 2010).

3.2.3 Clarifications

Epigenetic marks can be transmitted within one generation (i.e., from one cell to its daughter cells), or to subsequent generations. The former is defined as intergenerational epigenetic inheritance, and the latter as transgenerational epigenetic inheritance. However, when studying epigenetic memories induced by environmental factors, direct and parental epigenetic effects must be carefully distinguished. In other words, not all changes in further generations' epigenomes induced by initial environmental exposures should be categorized as epigenetic inheritance. For example, any exposure to environmental factors during pregnancy, known as "intrauterine experiences", may potentially alter the epigenome and germ cells of a fetus, independently of parental inheritance, and thus should not be regarded as epigenetic inheritance (van Otterdijk and Michels, 2016). Similarly, a male's exposure to environmental cues may affect his germ cells and induce epigenetic alterations that should not be regarded as transgenerational epigenetic inheritance. Like intrauterine exposures, maternal behaviors such as nursing, licking and grooming in mice, which may putatively affect offspring via gene expression changes in the brain, fall outside of this category (Weaver *et al.*, 2004).

3.2.4 Inheritance of environmental responses

Recent observations have indicated that responses to environmental exposures can be transferred to offspring (Heard and Martienssen, 2014; Soubry *et al.*, 2014; Aiken, Tarry-Adkins and Ozanne, 2016). Therefore, such environmental exposures not only affect the epigenome of one individual but may also affect the subsequent generation or generations. These environmental factors include lifestyle behaviors such as nutrition, physical activities, traumatic stress, and smoking. The non-genetic transgenerational inheritance of obesity or type II diabetes are examples of the observed traits (Stegemann and Buchner, 2015). The

underlying mechanisms of non-genetic inheritance are unclear, but some studies suggest involvement of alterations of histone modifications and/or DNA methylation in germ cells (Greer *et al.*, 2014; Duan *et al.*, 2016).

Traditionally, studies of transgenerational epigenetic inheritance focused on transmission through the maternal line, and any significant paternal effects were considered to be mediated by maternal responses, except possibly in paternally caring species (Wolf and Wade, 2009). The lack of attention to paternal effects was mainly due to spermatozoa having much smaller sizes and cytoplasmic contents than the eggs they fertilize, which contain many proteins and RNAs (Immler, 2018). Moreover, during spermatogenesis most histones are removed and replaced by proteins called protamines in both mice and humans (Wykes and Krawetz, 2003). However, more recent studies have started to address paternal epigenetic inheritance as a distinct route from the maternal effects (Crean and Bonduriansky, 2014). Some mechanisms such as DNA methylation, histone modifications and processes mediated by small non-coding RNAs (sncRNAs) have suggested involvement as non-genetic factors transferred through sperm. For instance, microRNAs (miRNAs) are reportedly involved in transfer of effects of traumatic stress to offspring in mice (Gapp *et al.*, 2014). Today, we have incipient understanding of transgenerational inheritance through germ cells, especially sperm, and more studies are needed to understand the underlying mechanisms and potentially discover further routes for transmission of epigenetic information.

Results and Discussion

This thesis is based upon results of studies reported in the three included papers, designated Papers I-III, outlined in the following sections.

Paper 1

CBP Regulates Recruitment and Release of Promoter-Proximal RNA Polymerase II

As discussed throughout this thesis, numerous factors and cofactors are involved in gene transcription from initiation to termination. These factors include many proteins, chemical modifications, DNA sequence features and even elements located far from the regulated genes. One of the major players in gene expression is CBP/P300. The study reported in Paper I focused on roles of this protein and GAGA binding factor (GAF), also known as Trithorax-like (Trl) in transcription initiation and formation of PIC, as well as in releasing Pol II from its paused state. These "checkpoints" are recognized as two major regulatory steps in gene transcription. CBP is found at enhancers and previous studies have shown that it plays an important role at this position. It has a KAT domain that catalyzes acetylation of H3K18, H3K27 and H4K8, all associated with active chromatin states. In addition to enhancers, CBP can also be found at other regulatory elements, including promoters.

To study CBP's functions at promoters, we used *D. melanogaster* cell line S2. Initially, by examining promoters of expressed genes (using RNA-seq techniques), we observed that CBP occupation is higher in almost all expressed genes than in non-expressed genes. We then examined enrichment of DNA-motifs at strongly CBP-binding promoters, and found a GAF-motif (GAGA-motif). GAF is another abundant protein with known functions in gene expression. It can be found at a subset of expressed genes and its ability to remove local nucleosomes clarifies its positioning at these sites. GAF also has known functions in PIC formation and Pol II pausing, through interaction with TFIID and recruitment of NELF, respectively.

Next, we analyzed the co-occupation of CBP and GAF at promoters and observed high correlation in their binding patterns. We also found that knocking down GAF significantly reduced CBP occupancy, possibly due to GAF's ability to remove nucleosomes and thus facilitate binding of all factors, including CBP.

GAF is known to have strong associations with Pol II pausing, but this link was unknown for CBP. Therefore, we analyzed correlations between CBP, GAF, and

Pol II pausing at the promoters of expressed genes and observed high correlation of the three factors. Next, we classed the promoters in 20 bins, based on their levels of CBP and GAF enrichment, and found that genes with higher levels of CBP and GAF at their promoters generally have higher pausing indices (ratio of Pol II occupancy between their promoters and gene bodies). We labelled two bins with the highest enrichment of CBP, GAF and Pol II pausing as HCG promoters, four with moderate enrichment as MCG promoters, and the remaining 14 (with the lowest enrichment of the three factors) LCG promoters. These observations indicate that CBP may contribute to promoter-proximal pausing.

One of the factors heavily involved in pausing is NELF, also known as pause-inducing factor. Unsurprisingly, we found higher levels of NELF in HCG promoters than in other promoters. Moreover, we also found higher enrichment of TFIIA at HCG promoters. Further analysis showed that genes with HCG promoters are distinct from other expressed genes, with a distinctive set of factors and co-factors binding at their promoters. Moreover, we found that HCG promoters differ from other promoters in chromatin state and associated histone marks. Briefly, HCG promoters have higher enrichment of histone marks associated with enhancers and open-chromatin states. Considering the distinctive properties of HCG promoters, we compared expression of their genes in early embryos and late embryos that S2 cells are derived from. The results showed that HCG promoters preserve their observed differences in terms of high CBP, GAF and Pol II pausing in different developmental stages as well as different cell types, which emphasizes their unique properties.

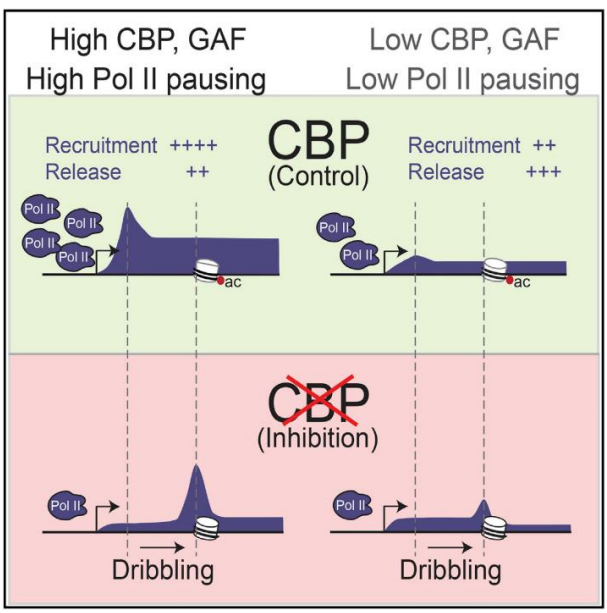
We next used inhibitor C464 to shut down CBP's catalytic activity and monitored results with single base resolution PRO-seq (measuring Pol II occupancy) after 10 minutes of inhibition. The CBP inhibition reduced expression of many genes, most strongly genes with high CBP occupancy. CBP RNAi analysis confirmed these effects. Interestingly, we observed increases in Pol II occupancy after CBP inhibition in the promoter proximal regions of LCG and MCG genes. As the level of expression in associated gene bodies is decreased by CBP inhibition at these promoters, these results suggest that CBP participates in releasing Pol II from its paused state. In contrast, at HCG promoters Pol II decreased in the promoter proximal region, indicating that CBP recruits Pol II to these promoters. We also found that pausing indices of all genes increased following CBP inhibition, indicating an association of CBP with release from pausing in all genes.

Further examination of effects of CBP inhibition on selected genes showed that it did not affect levels of TBP and TFIIA occupancy but significantly reduced TFIIB and TFIIF occupancy in HCG genes. Moreover, TFIB ChIP-seq after CBP inhibition (10 min) detected a strong genome-wide reduction in TFIIB

occupancy. These findings, together with co-immunoprecipitation of CBP-TFIIB, suggest a regulatory function for CBP, possibly via direct interaction with TFIIB.

Interestingly, CBP inhibition not only affects Pol II rates in promoter-proximal paused regions but also repositions Pol II downstream of the sites. This "dribbling" effect of CBP was observed by genome-wide PRO-seq analysis indicating that the repositioning occurs 60 to 180 bp downstream of the paused site. Next, we performed MNase-seq analysis to detect nucleosomes based on their ability to protect DNA from digestion. Results of this screening indicate that nucleosome density and positioning do not significantly change upon CBP inhibition. Moreover, our data suggest that CBP helps Pol II to overcome the +1 nucleosome barrier, possibly due to its histone acetylation capacity.

The Pol II dribbling effect after CBP inhibition could be explained by CBP's ability to recruit Pol II and facilitate its movement past the +1 nucleosome. Consequently, upon CBP inhibition Pol II will either decrease or accumulate just before the +1 nucleosome barrier. The main results of this study are summarized in the figure below.



Paper 2

Interdependence of PRC1 and PRC2 for recruitment to Polycomb Response Elements.

Traditionally, the suggested model for PcG recruitment was as follows. The PHO-RC subunit binds to PRE and promotes recruitment of PRC2 to the site. Then, PRC2 tri-methylates H3K27, which is a mark for recognition by the PRC1 subunit. Subsequently, PRC1 recognizes this mark via its PC part and deploys the ubiquitylation mark. In recent years, conflicting observations have challenged this view. In the study reported in Paper 2, to elucidate the functional interdependence of the two main parts of PcG complexes (PRC1 and PRC2) we made two cell lines of *D. melanogaster*: one with a *Psc/Su(z)2* mutation bestowing PRC1-deficiency and the other with a *Su(z)12* mutation bestowing PRC2-deficiency. The results are summarized in the following sections.

PRC1 deficiency:

PRC1-deficiency reduced the RING subunit level 4-fold, but did not affect transcription of the gene encoding it. Therefore, PRC1 subunits are apparently degraded upon mutation, but no effect on PRC2 integrity was detected in PRC1-deficient cells. This was also validated by ChIP-qPCR analysis of PSC and PC subunits at six selected PREs. Notably, ChIP signals for parts of the PRC2 complex as well as H3K27me3 were dramatically decreased in PRC1-deficient cells. Our results suggest that although PRC2 integrity seems to be independent of PRC1, at many PREs PRC2 binding is impaired without successful recruitment of PRC1. Nevertheless, PREs seem to vary in terms of coordinating PcG complexes and their recruitment.

Next, we tested the PRC2:PRC1 dependency genome-wide by ChIP-seq analysis. For this, we utilized E(Z) ChIP-seq signals to distinguish PRC2 binding to PREs upon PRC1 ablation. All PREs showed lower levels of PRC2, but about two thirds of PREs showed severe reduction of PRC2 complex binding when PRC1 was impaired. Our data suggest that the level of PRC2:PRC1-dependency varies across PREs.

We next examined computationally defined PREs, where there was no PRC1 dependency for recruitment of PRC2 complexes (independent PREs). By using a transgene consisting of a *white* reporter gene we checked whether these PREs are functional units and can recruit both PcG complexes. Our ChIP data suggest that PREs in which PRC2 recruitment is independent of PRC1 can recruit PRC complexes and repress the reporter gene.

In this study, we also investigated the link between H2AK118ub and PRC2:PRC1 dependency as this mark was reportedly responsible for PRC2 recruitment in mice (Blackledge *et al.*, 2014). Our ChIP-qPCR results showed no correlation between H2K118ub and dependency between PRC complexes, and the ubiquitylation mark was lost in both types of PREs (PRC1-dependent and PRC1-independent) upon PRC1 deficiency. Therefore, we concluded that H2AK118ub is not a requirement for recruitment of PRC2.

PRC2 deficiency:

Next, we examined the PRC2-deficient cell line, in which *Su(z)12* mutation results in degradation of E(Z) protein. Hence, there is no H3K27me2,3 marking mediated by E(Z). In PRC2-deficient cells we observed that loss of PRC2 has no effect on total levels of PRC1 or H2AK119ub in the cells. Through ChIP-qPCR with antibodies against different subunits of PRC1 complexes, we checked PRC1:PRC2 dependency in PRC2-deficient cells. Our results suggest that in both types of PREs (PRC1-dependent and PRC1-independent), the PRC1 recruitment is independent of PRC2 and H3K27me3 marks. This finding was further tested by comparing ChIP-seq signals for PSC subunits in mutant cells versus controls. We found no measurable effect on PSC, but this effect did not differ in the two types of PREs and was not as strong as the PRC2:PRC1 dependency. We observed no indication of strict dependency of PRC1 recruitment on PRC2.

Another part of PcG proteins is the PHO-RC complex. Although little is known about its function, one hypothesis is that PHO-RC directly interacts with PRC2, anchoring it to PREs. To investigate the role of PHO-RC complexes in PRC recruitment, we used ChIP-qPCR with antibodies against PHO in both mutant cell lines (PRC1- and PRC2-deficient). The results showed that PHO-RC:PRC1 interaction does not involve PRC2 and the PHO signal did not change in the absence of PRC2 or H3K27me3.

In this study we showed that PREs can differ in terms of recruitment strategy for PRC complexes. Some PREs need PRC1 recruitment for successful PRC2 binding while others can recruit both complexes independently. Our results also challenged the classical view of H3K27me3 and H2AK119ub marks as they were previously believed to be essential requirements for recruitment of PRC complexes. Overall, our results suggest that there is flexibility in recruitment of PRC complexes by different PREs.

Paper 3

The gut microbiome participates in transgenerational inheritance of low-temperature responses in *Drosophila melanogaster*

In recent years epigenetic studies have revealed that transcriptional responses to environmental stimuli can be transferred to the next generation and often persist for several more generations. Several mechanisms have been suggested for this transmittance, including DNA methylation, histone modifications and processes mediated by ncRNAs. All investigations of this phenomenon before the study presented in Paper 3 addressed the mechanism(s) involved in the germline. However, germ cells are not the only materials transferred from mothers to their children. We know that in metazoans the microbiome is also transferred from parents to their offspring. Briefly, the microbiome is the consortium of all bacteria that share an organism's body space, and their relations to several physiological and psychological disorders have been reported and studied in several cases. Interestingly, a number of studies have observed undesirable phenotype transmittance upon fecal transplantation, which is an old remedy for gastrointestinal disorders especially *C. difficile* infections. These observations raised the possibility that the microbiome can participate in transgenerational inheritance of transcriptional responses.

To test this hypothesis, we used *D. melanogaster* as it is an ideal model for this purpose since fruit flies defecate in their food and the offspring consume the parents' feces (and thus their microbiomes) together with food. We designed our experiment to control the set of bacteria transmitted from parents to their offspring while the parents were exposed to a cold (18°C) or control (25°C) temperature. All offspring were reared at the control temperature. We set up the experiment so that the offspring received a controlled set of parental bacteria in their media. We then used 16S rDNA amplicon sequencing to evaluate the microbial transfer and RNA-seq to study the transcriptional responses in both generations.

Next, we compared differentially expressed genes (DEGs) in representatives of the two generations. To identify DEGs in the F1 generation, we compared control and cold-treated F1 flies, and to identify DEGs in the F2 generation we compared flies that had received their germline and microbiome from control flies to flies that obtained both from cold-treated flies. Analysis of the DEGs in parents and their offspring revealed a significant overlap, indicating that some parental transcriptional responses were retained in their offspring although the initial stimuli was absent (as all offspring were grown at the control temperature). Hierarchical clustering of all samples showed that the major route for inheritance

of transcriptional patterns is, unsurprisingly, through the germline. Almost all the offspring samples clustered together with the parents providing their germline. This was also seen in Principal Component Analysis (PCA) of all offspring samples based on their expression profiles obtained from RNA-seq read counts.

The first Principal Component clearly separated offspring samples according to their germline, while the second separated offspring samples according to their acquired microbiome. To further validate this observation, we used OPLS-DA models, which confirmed that the offspring samples can be separated based on their acquired microbiomes.

In agreement with previous studies, we found that two families of bacteria (Acetobacteriaceae and Lactobacillaceae) are the main components of the gut microbiome in *Drosophila* flies. Our observations also revealed that relative frequencies of these bacterial families differ between control and cold-treated flies and the difference is maintained in offspring that receive the respective microbiomes. Moreover, analysis of DEGs' expression in different *Drosophila* tissues showed that most of the selected genes (with trans-generationally inherited expression patterns, linked to the microbiome exchange), are highly and mainly expressed in the gut. Thus, some of the interplay between microbiome and fly host cells apparently occurs in the gut. Identification of the specific types of bacteria involved and their action mechanisms would require further investigation. Nevertheless, we believe that this study provided the first evidence that the microbiome participates in transgenerational inheritance of acquired transcriptional responses and should be considered together with previously studied mechanisms.

Acknowledgements

First and foremost, I would like to thank my main supervisor, dear friend, older brother and mentor **Per Stenberg**. You have not only educated me in molecular biology and bioinformatics but also taught me many lessons in life. I will always be thankful for your kindness, patience and wisdom. I could not in any way have wished for a better supervisor. Thanks for everything!

Next, I would like to thank the current and previous group members of Stenberg's lab, and I apologize for not mentioning everyone's names. **Philge**, I am grateful to you for patiently helping me learn how to code and encouraging me whenever I was feeling discouraged. **Jacob**, I am thankful to you for being an amazing guy, awesome colleague and trustworthy friend. I am very happy that we had you in the group, and I wish you all the best. **Anna-Mia**, thank you for being around with the charming smile you always wear. Without you the fly floor would be much gloomier. **Edvin**, I am thankful to you for being an amazing, calm and disciplined colleague.

Many thanks to our collaborator and my dear friend **Nico**. You have an interesting character, full of confidence, very strong and super friendly which transmits a great feeling to people around you. I certainly enjoyed talking with you whenever we met, regardless of whether it was in your office, or a restaurant or at a party.

I am very thankful to my former roommate and oldest friend in town, **Hasan**. You always inspire me. Thank you for being there where I needed help, especially near the end. I wish the best for you and your family.

Special thanks to our neighboring lab leader **Jan Larsson** who is kind, smart, knowledgeable and very humble. I enjoyed every conversation with you and I am proud that I had the opportunity to be an assistant on your courses for several years. I am also thankful to your group members (among others, who are not mentioned, but also not forgotten) **Masha** and **Henrik** for always creating a kind and friendly atmosphere.

I am also thankful to my co-supervisor **Yuri Schwartz**. I have learned a lot from our discussions. Thank you for taking the time to explain stuff to me every time I knocked on your door. I learned lessons from you that I couldn't learn from anyone else. Many thanks also to your group members (among others) **Tanya** and **Alex,** and also **Misha** who I miss a lot!

I would also like to thank my friends on the fly floor who contributed to the environment's friendliness. **Behzad**, I am thankful for your friendship during these years. The best times at work were sitting on the sofa in your office and talking about different stuff, mainly football and games. **Isaac**, thank you for your kind support every time I needed it at work. Best of luck for your dissertation (but I'm sure you won't need luck). **Marie-Line**, I greatly enjoyed every discussion we had in the office, lunchroom, lab, your house, my house and our apartment when we were housemates. I wish you and your family a life full of joy and happiness.

I am also thankful to my friends outside the department, **Kaviano** and **Rman** for many years of friendship and endless support.

I would also like to thank many people in the Molecular Biology department who helped me to progress during all these years. I apologize for not mentioning all your names. **Micke**, I am thankful to you for leaving your door open and always being willing to talk about stuff. I always felt calmer after talking with you. **Marek**, I am very thankful for all the times we talked about computers and stuff. **Mathew**, I am very thankful for all the times you spent chatting with me when I was feeling low. You were there in the beginning as my first teacher during my Master's and towards the end of the road. Thank you. **Marcus**, I always appreciate your company, it's always very comforting to talk to you. I wish the best for you. **Christer**, thanks for encouraging me every time I talked with you. Rock on!

Finally, I would like to thank my **JoJo** who was with me during most of this long path. You were with me every moment through ups and downs. We shared the best times and worst times, and you stayed by my side. I wish a happy future for both of us.

Last, but not least, I must thank my parents **Ata** and **Fifi**, especially my **mom**, for everything they have done for me. I am also thankful to my dear sisters **Hoda**, **Neda** and **Hosna** for their love and support, even when I am living far away from them. Love you all!

References

Ahn, S. H., Kim, M. and Buratowski, S. (2004) 'Phosphorylation of Serine 2 within the RNA Polymerase II C-Terminal Domain Couples Transcription and 3′ End Processing', *Molecular Cell*. Cell Press, 13(1), pp. 67–76. doi: 10.1016/S1097-2765(03)00492-1.

Aiken, C. E., Tarry-Adkins, J. L. and Ozanne, S. E. (2016) 'Transgenerational effects of maternal diet on metabolic and reproductive ageing', *Mammalian Genome*. Springer US, 27(7–8), pp. 430–439. doi: 10.1007/s00335-016-9631-1.

Akhtar, A. and Becker, P. B. (2000) 'Activation of Transcription through Histone H4 Acetylation by MOF, an Acetyltransferase Essential for Dosage Compensation in Drosophila', *Molecular Cell*. Cell Press, 5(2), pp. 367–375. doi: 10.1016/S1097-2765(00)80431-1.

Alexandrov, L. B. *et al.* (2013) 'Signatures of mutational processes in human cancer', *Nature*. Nature Publishing Group, 500(7463), pp. 415–421. doi: 10.1038/nature12477.

Allen, B. L. and Taatjes, D. J. (2015) 'The Mediator complex: A central integrator of transcription', *Nature Reviews Molecular Cell Biology*, pp. 155–166. doi: 10.1038/nrm3951.

Allfrey, V. G., Faulkner, R. and Mirsky, A. E. (1964) 'Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 51(5), pp. 786–94. doi: 10.1073/PNAS.51.5.786.

Azevedo, C. and Saiardi, A. (2016) 'Why always lysine? The ongoing tale of one of the most modified amino acids', *Advances in Biological Regulation*. Pergamon, 60, pp. 144–150. doi: 10.1016/J.JBIOR.2015.09.008.

Bajusz, I. *et al.* (2018) 'From Flies to Mice: The Emerging Role of Non-Canonical PRC1 Members in Mammalian Development', *Epigenomes*. Multidisciplinary Digital Publishing Institute, 2(1), p. 4. doi: 10.3390/epigenomes2010004.

Bannister, A. J., Schneider, R. and Kouzarides, T. (2002) 'Histone Methylation: Dynamic or Static?', *Cell*. Cell Press, 109(7), pp. 801–806. doi: 10.1016/S0092-8674(02)00798-5.

Barboric, M. *et al.* (2001) 'NF-κB Binds P-TEFb to Stimulate Transcriptional Elongation by RNA Polymerase II', *Molecular Cell*. Cell Press, 8(2), pp. 327–337. doi: 10.1016/S1097-2765(01)00314-8.

Barlow, D. P. and Bartolomei, M. S. (2014) 'Genomic imprinting in mammals.', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press, 6(2), p. a018382. doi: 10.1101/cshperspect.a018382.

Barski, A. *et al.* (2007) 'High-Resolution Profiling of Histone Methylations in the Human Genome', *Cell*. Cell Press, 129(4), pp. 823–837. doi: 10.1016/J.CELL.2007.05.009.

Barth, T. K. and Imhof, A. (2010) 'Fast signals and slow marks: the dynamics of histone

modifications', *Trends in Biochemical Sciences*. Elsevier Current Trends, 35(11), pp. 618–626. doi: 10.1016/J.TIBS.2010.05.006.

Bauer, M., Trupke, J. and Ringrose, L. (2016) 'The quest for mammalian Polycomb response elements: are we there yet?', *Chromosoma*. Springer, 125(3), pp. 471–96. doi: 10.1007/s00412-015-0539-4.

Bedford, D. C. *et al.* (2010) 'Target gene context influences the transcriptional requirement for the KAT3 family of CBP and p300 histone acetyltransferases', *Epigenetics*. Taylor & Francis, 5(1), pp. 9–15. doi: 10.4161/epi.5.1.10449.

Bird, A. P. (1980) 'DNA methylation and the frequency of CpG in animal DNA.', *Nucleic acids research*. Oxford University Press, 8(7), pp. 1499–504. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6253938 (Accessed: 3 July 2018).

Blackledge, N. P. *et al.* (2014) 'Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation', *Cell*. Cell Press, 157(6), pp. 1445–1459. doi: 10.1016/J.CELL.2014.05.004.

Blackledge, N. P., Rose, N. R. and Klose, R. J. (2015) 'Targeting Polycomb systems to regulate gene expression: modifications to a complex story', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 16(11), pp. 643–649. doi: 10.1038/nrm4067.

Bonev, B. and Cavalli, G. (2016) 'Organization and function of the 3D genome', *Nature Reviews Genetics*. Nature Publishing Group, 17(11), pp. 661–678. doi: 10.1038/nrg.2016.112.

Budhavarapu, V. N., Chavez, M. and Tyler, J. K. (2013) 'How is epigenetic information maintained through DNA replication?', *Epigenetics & Chromatin*. BioMed Central, 6(1), p. 32. doi: 10.1186/1756-8935-6-32.

Cameron, S. R. *et al.* (2018) 'PTE, a novel module to target Polycomb Repressive Complex 1 to the human cyclin D2 (CCND2) oncogene.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 293(37), pp. 14342–14358. doi: 10.1074/jbc.RA118.005010.

Carrozza, M. J. *et al.* (2003) 'The diverse functions of histone acetyltransferase complexes', *Trends in Genetics*. Elsevier Current Trends, 19(6), pp. 321–329. doi: 10.1016/S0168-9525(03)00115-X.

Cedar, H. and Bergman, Y. (2012) 'Programming of DNA Methylation Patterns', *Annual Review of Biochemistry*. Annual Reviews , 81(1), pp. 97–117. doi: 10.1146/annurev-biochem-052610-091920.

Cheung, A. C. M. and Cramer, P. (2011) 'Structural basis of RNA polymerase II backtracking, arrest and reactivation', *Nature*. Nature Publishing Group, 471(7337), pp. 249–253. doi: 10.1038/nature09785.

Chittock, E. C. *et al.* (2017) 'Molecular architecture of polycomb repressive complexes.', *Biochemical Society transactions*. Portland Press Ltd, 45(1), pp. 193–205. doi: 10.1042/BST20160173.

Chrun, E. S., Modolo, F. and Daniel, F. I. (2017) 'Histone modifications: A review about the presence of this epigenetic phenomenon in carcinogenesis', *Pathology - Research and Practice*. Urban & Fischer, 213(11), pp. 1329–1339. doi: 10.1016/J.PRP.2017.06.013.

Coleman, R. A. *et al.* (1999) 'TFIIA Regulates TBP and TFIID Dimers', *Molecular Cell*. Cell Press, 4(3), pp. 451–457. doi: 10.1016/S1097-2765(00)80453-0.

Collings, C. K., Waddell, P. J. and Anderson, J. N. (2013) 'Effects of DNA methylation on nucleosome stability', *Nucleic Acids Research*. Oxford University Press, 41(5), pp. 2918–2931. doi: 10.1093/nar/gks893.

Conaway, R. C. and Conaway, J. W. (2011) 'Origins and activity of the Mediator complex', *Seminars in Cell & Developmental Biology*. Academic Press, 22(7), pp. 729–734. doi: 10.1016/J.SEMCDB.2011.07.021.

Core, L. J. *et al.* (2014) 'Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers', *Nature Genetics*. Nature Publishing Group, 46(12), pp. 1311–1320. doi: 10.1038/ng.3142.

Core, L. J., Waterfall, J. J. and Lis, J. T. (2008) 'Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters', *Science*, 322(5909), pp. 1845–1848. doi: 10.1126/science.1162228.

Crean, A. J. and Bonduriansky, R. (2014) 'What is a paternal effect?', *Trends in Ecology and Evolution*. Elsevier Ltd, 29(10), pp. 554–559. doi: 10.1016/j.tree.2014.07.009.

Cremer, T. and Cremer, C. (2001) 'Chromosome territories, nuclear architecture and gene regulation in mammalian cells', *Nature Reviews Genetics*. Nature Publishing Group, 2(4), pp. 292–301. doi: 10.1038/35066075.

Dancy, B. M. and Cole, P. A. (2015) 'Protein lysine acetylation by p300/CBP', *Chemical Reviews*, pp. 2419–2452. doi: 10.1021/cr500452k.

Danko, C. G. *et al.* (2013) 'Signaling Pathways Differentially Affect RNA Polymerase II Initiation, Pausing, and Elongation Rate in Cells', *Molecular Cell*. Cell Press, 50(2), pp. 212–222. doi: 10.1016/J.MOLCEL.2013.02.015.

Day, D. S. *et al.* (2016) 'Comprehensive analysis of promoter-proximal RNA polymerase II pausing across mammalian cell types', *Genome Biology*. BioMed Central, 17(1), p. 120. doi: 10.1186/s13059-016-0984-2.

Deans, C. and Maggert, K. A. (2015) 'What do you mean, "epigenetic"?', *Genetics*. Genetics Society of America, 199(4), pp. 887–96. doi: 10.1534/genetics.114.173492.

Deaton, A. M. and Bird, A. (2011) 'CpG islands and the regulation of transcription.', *Genes & development*. Cold Spring Harbor Laboratory Press, 25(10), pp. 1010–22. doi: 10.1101/gad.2037511.

Dekker, J. *et al.* (2002) 'Capturing chromosome conformation.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 295(5558), pp. 1306–11. doi:

10.1126/science.1067799.

Dunwell, T. L. and Pfeifer, G. P. (2014) 'Drosophila genomic methylation: new evidence and new questions.', *Epigenomics*. NIH Public Access, 6(5), pp. 459–61. doi: 10.2217/epi.14.46.

Eslami-Mossallam, B., Schiessel, H. and van Noort, J. (2016) 'Nucleosome dynamics: Sequence matters', *Advances in Colloid and Interface Science*. Elsevier, 232, pp. 101–113. doi: 10.1016/J.CIS.2016.01.007.

Filion, G. J. *et al.* (2010) 'Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells', *Cell*. Cell Press, 143(2), pp. 212–224. doi: 10.1016/J.CELL.2010.09.009.

Fishburn, J. *et al.* (2015) 'Double-stranded DNA translocase activity of transcription factor TFIIH and the mechanism of RNA polymerase II open complex formation', *PNAS*, 112(13), pp. 3961–3966. doi: 10.1073/pnas.1417709112.

Flanagan, P. M. *et al.* (1991) 'A mediator required for activation of RNA polymerase II transcription in vitro', *Nature*. Nature Publishing Group, 350(6317), pp. 436–438. doi: 10.1038/350436a0.

Francis, N. J. *et al.* (2009) 'Polycomb Proteins Remain Bound to Chromatin and DNA during DNA Replication In Vitro', *Cell*. Cell Press, 137(1), pp. 110–122. doi: 10.1016/J.CELL.2009.02.017.

Gapp, K. *et al.* (2014) 'Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice', *Nature Neuroscience*. Nature Publishing Group, 17(5), pp. 667–669. doi: 10.1038/nn.3695.

Gardini, A. (2017) 'Global Run-On Sequencing (GRO-Seq).', *Methods in molecular biology (Clifton, N.J.)*. NIH Public Access, 1468, pp. 111–20. doi: 10.1007/978-1-4939-4035-6_9.

Gates, L. A., Foulds, C. E. and O'Malley, B. W. (2017) 'Histone Marks in the "Driver"s Seat': Functional Roles in Steering the Transcription Cycle', *Trends in Biochemical Sciences*. Elsevier Current Trends, 42(12), pp. 977–989. doi: 10.1016/J.TIBS.2017.10.004.

Geisler, S. J. and Paro, R. (2015) 'Trithorax and Polycomb group-dependent regulation: a tale of opposing activities.', *Development (Cambridge, England)*. Oxford University Press for The Company of Biologists Limited, 142(17), pp. 2876–87. doi: 10.1242/dev.120030.

Gellon, G. and McGinnis, W. (1998) 'Shaping animal body plans in development and evolution by modulation of Hox expression patterns', *BioEssays*, 20(2), pp. 116–125. doi: 10.1002/(SICI)1521-1878(199802)20:2<116::AID-BIES4>3.0.CO;2-R.

Ghavi-Helm, Y. *et al.* (2014) 'Enhancer loops appear stable during development and are associated with paused polymerase', *Nature*. Nature Publishing Group, 512(7512), pp. 96–100. doi: 10.1038/nature13417.

Gil, J. and O'Loghlen, A. (2014) 'PRC1 complex diversity: where is it taking us?', *Trends in Cell Biology*. Elsevier Current Trends, 24(11), pp. 632–641. doi:

10.1016/J.TCB.2014.06.005.

Goodman, R. H. and Smolik, S. (2000) 'CBP/p300 in cell growth, transformation, and development.', *Genes & development*. Cold Spring Harbor Laboratory Press, 14(13), pp. 1553–77. doi: 10.1101/GAD.14.13.1553.

Gupta, K. *et al.* (2016) 'Zooming in on Transcription Preinitiation', *Journal of Molecular Biology*. Academic Press, 428(12), pp. 2581–2591. doi: 10.1016/J.JMB.2016.04.003.

Haberle, V. and Lenhard, B. (2016) 'Promoter architectures and developmental gene regulation', *Seminars in Cell & Developmental Biology*. Academic Press, 57, pp. 11–23. doi: 10.1016/J.SEMCDB.2016.01.014.

Hammond, C. M. *et al.* (2017) 'Histone chaperone networks shaping chromatin function.', *Nature reviews. Molecular cell biology*. NIH Public Access, 18(3), pp. 141–158. doi: 10.1038/nrm.2016.159.

Hantsche, M. and Cramer, P. (2017) 'Conserved RNA polymerase II initiation complex structure', *Current Opinion in Structural Biology*. Elsevier Current Trends, 47, pp. 17–22. doi: 10.1016/J.SBI.2017.03.013.

Heard, E. and Martienssen, R. A. (2014) 'Transgenerational epigenetic inheritance: Myths and mechanisms', *Cell*. Elsevier Inc., 157(1), pp. 95–109. doi: 10.1016/j.cell.2014.02.045.

Heintzman, N. D. *et al.* (2007) 'Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome', *Nature Genetics*. Nature Publishing Group, 39(3), pp. 311–318. doi: 10.1038/ng1966.

Henckel, A. and Arnaud, P. (2010) 'Genome-wide identification of new imprinted genes', *Briefings in Functional Genomics*. Oxford University Press, 9(4), pp. 304–314. doi: 10.1093/bfgp/elq016.

Holoch, D. and Margueron, R. (2017) 'Mechanisms Regulating PRC2 Recruitment and Enzymatic Activity', *Trends in Biochemical Sciences*, pp. 531–542. doi: 10.1016/j.tibs.2017.04.003.

Illingworth, R. S. *et al.* (2010) 'Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome', *PLoS Genet*, 6(9). doi: 10.1371/journal.pgen.1001134.

Immler, S. (2018) 'The sperm factor: paternal impact beyond genes', *Heredity*. Nature Publishing Group, 121(3), pp. 239–247. doi: 10.1038/s41437-018-0111-0.

Jabbari, K. and Bernardi, G. (2004) 'Cytosine methylation and CpG, TpG (CpA) and TpA frequencies', *Gene*. Elsevier, 333, pp. 143–149. doi: 10.1016/J.GENE.2004.02.043.

Jin, Q. *et al.* (2010) 'Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation', *The EMBO Journal*, 30, pp. 249–262. doi: 10.1038/emboj.2010.318.

Jones, P. A. (2012) 'Functions of DNA methylation: islands, start sites, gene bodies and

beyond', *Nature Reviews Genetics*. Nature Publishing Group, 13(7), pp. 484–492. doi: 10.1038/nrg3230.

Jonkers, I., Kwak, H. and Lis, J. T. (2014) 'Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons', *eLife*. eLife Sciences Publications Limited, 3, p. e02407. doi: 10.7554/eLife.02407.

Jonkers, I. and Lis, J. T. (2015) 'Getting up to speed with transcription elongation by RNA polymerase II', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 16(3), pp. 167–177. doi: 10.1038/nrm3953.

Kandiah, E. *et al.* (2014) 'More pieces to the puzzle: recent structural insights into class II transcription initiation', *Current Opinion in Structural Biology*. Elsevier Current Trends, 24, pp. 91–97. doi: 10.1016/J.SBI.2013.12.005.

Kanherkar, R. R., Bhatia-Dey, N. and Csoka, A. B. (2014) 'Epigenetics across the human lifespan.', *Frontiers in cell and developmental biology*. Frontiers Media SA, 2, p. 49. doi: 10.3389/fcell.2014.00049.

Kasper, L. H. *et al.* (2014) 'Genome-wide and single-cell analyses reveal a context dependent relationship between CBP recruitment and gene expression', *Nucleic Acids Research*. Oxford University Press, 42(18), pp. 11363–11382. doi: 10.1093/nar/gku827.

Kassis, J. A. and Brown, J. L. (2013) 'Polycomb Group Response Elements in Drosophila and Vertebrates', *Advances in Genetics*. Academic Press, 81, pp. 83–118. doi: 10.1016/B978-0-12-407677-8.00003-8.

Kassis, J. A., Kennison, J. A. and Tamkun, J. W. (2017) 'Polycomb and Trithorax Group Genes in Drosophila.', *Genetics*. Genetics, 206(4), pp. 1699–1725. doi: 10.1534/genetics.115.185116.

Kebede, A. F., Schneider, R. and Daujat, S. (2015) 'Novel types and sites of histone modifications emerge as players in the transcriptional regulation contest', *FEBS Journal*. Wiley/Blackwell (10.1111), 282(9), pp. 1658–1674. doi: 10.1111/febs.13047.

Kelleher, R. J., Flanagan, P. M. and Kornberg, R. D. (1990) 'A novel mediator between activator proteins and the RNA polymerase II transcription apparatus', *Cell*. Cell Press, 61(7), pp. 1209–1215. doi: 10.1016/0092-8674(90)90685-8.

Kellner, W. A., Bell, J. S. K. and Vertino, P. M. (2015) 'GC skew defines distinct RNA polymerase pause sites in CpG island promoters.', *Genome research*. Cold Spring Harbor Laboratory Press, 25(11), pp. 1600–9. doi: 10.1101/gr.189068.114.

Kharchenko, P. V. *et al.* (2011) 'Comprehensive analysis of the chromatin landscape in Drosophila melanogaster', *Nature*. Nature Publishing Group, 471(7339), pp. 480–485. doi: 10.1038/nature09725.

Khatter, H., Vorländer, M. K. and Müller, C. W. (2017) 'RNA polymerase I and III: similar yet unique', *Current Opinion in Structural Biology*. Elsevier Current Trends, 47, pp. 88–94. doi: 10.1016/J.SBI.2017.05.008.

Kingston, R. E. and Tamkun, J. W. (2014) 'Transcriptional regulation by trithorax-group proteins', *Cold Spring Harbor Perspectives in Biology*. Cold Spring Harbor Laboratory Press, 6(10), p. a019349. doi: 10.1101/cshperspect.a019349.

Kodzius, R. *et al.* (2006) 'CAGE: cap analysis of gene expression', *Nature Methods*. Nature Publishing Group, 3(3), pp. 211–222. doi: 10.1038/nmeth0306-211.

Komarnitsky, P., Cho, E. J. and Buratowski, S. (2000) 'Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription.', *Genes & development*. Cold Spring Harbor Laboratory Press, 14(19), pp. 2452–60. doi: 10.1101/GAD.824700.

Kornberg, R. D. (1974) 'Chromatin Structure: A Repeating Unit of Histones and DNA', *Science*, 184(4139). doi: 10.1126/science.184.4139.868.

Kornberg, R. D. (1974) 'Chromatin Structure: A Repeating Unit of Histones and DNA', *Science*, 184(4139), pp. 868–871. doi: 10.1126/science.184.4139.868.

Kouzarides, T. (2007) 'Chromatin modifications and their function.', *Cell*, 128(4), pp. 693–705. doi: 10.1016/j.cell.2007.02.005.

Kubitscheck, U. and Siebrasse, J.-P. (2017) 'Kinetics of transport through the nuclear pore complex', *Seminars in Cell & Developmental Biology*. Academic Press, 68, pp. 18–26. doi: 10.1016/J.SEMCDB.2017.06.016.

Kwak, H. *et al.* (2013) 'Precise maps of RNA polymerase reveal how promoters direct initiation and pausing.', *Science (New York, N.Y.)*. NIH Public Access, 339(6122), pp. 950–3. doi: 10.1126/science.1229386.

Lagrange, T. *et al.* (1998) 'New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB.', *Genes & development*. Cold Spring Harbor Laboratory Press, 12(1), pp. 34–44. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9420329 (Accessed: 9 July 2018).

Lambert, S. A. *et al.* (2018) 'The Human Transcription Factors', *Cell*. Cell Press, 172(4), pp. 650–665. doi: 10.1016/J.CELL.2018.01.029.

Lanzuolo, C. *et al.* (2011) 'PcG Complexes Set the Stage for Epigenetic Inheritance of Gene Silencing in Early S Phase before Replication', *PLoS Genetics*. Edited by A. Akhtar. Public Library of Science, 7(11), p. e1002370. doi: 10.1371/journal.pgen.1002370.

Lanzuolo, C., Lo Sardo, F. and Orlando, V. (2012) 'Concerted epigenetic signatures inheritance at PcG targets through replication', *Cell Cycle*. Taylor & Francis, 11(7), pp. 1296–1300. doi: 10.4161/cc.19710.

Lawrence, M., Daujat, S. and Schneider, R. (2016) 'Lateral Thinking: How Histone Modifications Regulate Gene Expression', *Trends in Genetics*. Elsevier Current Trends, 32(1), pp. 42–56. doi: 10.1016/j.tig.2015.10.007.

Lee, S. H. *et al.* (2018) 'Mapping the spectrum of 3D communities in human chromosome conformation capture data'. Available at: http://arxiv.org/abs/1810.01307 (Accessed: 28

October 2018).

Levine, M., Cattoglio, C. and Tjian, R. (2014) 'Looping Back to Leap Forward: Transcription Enters a New Era', *Cell*. Cell Press, 157(1), pp. 13–25. doi: 10.1016/J.CELL.2014.02.009.

Lewis, E. . (1947) 'New mutants: Reports of P. lewis. Drosoph.', *Inf. Serv.*, p. 21:69.

Lewis, E. B. (1978) 'A gene complex controlling segmentation in Drosophila.', *Nature*, 276(5688), pp. 565–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/103000 (Accessed: 15 October 2018).

Li, J. and Gilmour, D. S. (2013) 'Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor.', *The EMBO journal*. EMBO Press, 32(13), pp. 1829–41. doi: 10.1038/emboj.2013.111.

Lieberman-Aiden, E. *et al.* (2009) 'Comprehensive mapping of long-range interactions reveals folding principles of the human genome.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 326(5950), pp. 289–93. doi: 10.1126/science.1181369.

Liu, W. H. *et al.* (2012) 'CAF-1-induced oligomerization of histones H3/H4 and mutually exclusive interactions with Asf1 guide H3/H4 transitions among histone chaperones and DNA', *Nucleic Acids Research*. Oxford University Press, 40(22), pp. 11229–11239. doi: 10.1093/nar/gks906.

Lo, S. M. *et al.* (2012) 'A Bridging Model for Persistence of a Polycomb Group Protein Complex through DNA Replication In Vitro', *Molecular Cell*. Cell Press, 46(6), pp. 784–796. doi: 10.1016/J.MOLCEL.2012.05.038.

Lomvardas, S. and Thanos, D. (2002) 'Opening Chromatin', *Molecular Cell*. Cell Press, 9(2), pp. 209–211. doi: 10.1016/S1097-2765(02)00463-X.

Long, H. K. *et al.* (2013) 'Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates', *eLife*. eLife Sciences Publications Limited, 2, p. e00348. doi: 10.7554/eLife.00348.

Long, H. K., Prescott, S. L. and Wysocka, J. (2016) 'Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution', *Cell*, pp. 1170–1187. doi: 10.1016/j.cell.2016.09.018.

Loya, T. J. and Reines, D. (2016) 'Recent advances in understanding transcription termination by RNA polymerase II.', *F1000Research*. Faculty of 1000 Ltd, 5. doi: 10.12688/f1000research.8455.1.

Lu, X. *et al.* (2008) 'The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure', *Nature Structural & Molecular Biology*. Nature Publishing Group, 15(10), pp. 1122–1124. doi: 10.1038/nsmb.1489.

Luger, K. *et al.* (1997) 'Crystal structure of the nucleosome core particle at 2.8 Å resolution', *NATURE*, 389. Available at: https://www.nature.com/articles/38444.pdf (Accessed: 30

June 2018).

Macek, B. *et al.* (2006) 'Top-down protein sequencing and MS3 on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer.', *Molecular & cellular proteomics : MCP*. American Society for Biochemistry and Molecular Biology, 5(5), pp. 949–58. doi: 10.1074/mcp.T500042-MCP200.

Macleod, D. *et al.* (1994) 'Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island.', *Genes & development*. Cold Spring Harbor Laboratory Press, 8(19), pp. 2282–92. doi: 10.1101/GAD.8.19.2282.

Margueron, R. *et al.* (2009) 'Role of the polycomb protein EED in the propagation of repressive histone marks', *Nature*. Nature Publishing Group, 461(7265), pp. 762–767. doi: 10.1038/nature08398.

Maston, G. A., Evans, S. K. and Green, M. R. (2006) 'Transcriptional Regulatory Elements in the Human Genome', *Annual Review of Genomics and Human Genetics*, 7(1), pp. 29–59. doi: 10.1146/annurev.genom.7.080505.115623.

Mayer, A., Landry, H. M. and Churchman, L. S. (2017) 'Pause & go: from the discovery of RNA polymerase pausing to its functional implications', *Current Opinion in Cell Biology*. Elsevier Current Trends, pp. 72–80. doi: 10.1016/j.ceb.2017.03.002.

Morgan, H. D. *et al.* (2005) 'Epigenetic reprogramming in mammals', *Human Molecular Genetics*. Oxford University Press, 14(suppl_1), pp. R47–R58. doi: 10.1093/hmg/ddi114.

Müller, J. *et al.* (2002) 'Histone Methyltransferase Activity of a Drosophila Polycomb Group Repressor Complex', *Cell*. Cell Press, 111(2), pp. 197–208. doi: 10.1016/S0092-8674(02)00976-5.

Müller, J. and Bienz, M. (1991) 'Long range repression conferring boundaries of Ultrabithorax expression in the Drosophila embryo.', *The EMBO journal*. European Molecular Biology Organization, 10(11), pp. 3147–55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1680676 (Accessed: 22 July 2018).

Nekrasov, M. *et al.* (2007) 'Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes.', *The EMBO journal*. EMBO Press, 26(18), pp. 4078–88. doi: 10.1038/sj.emboj.7601837.

Nora, E. P. *et al.* (2012) 'Spatial partitioning of the regulatory landscape of the X-inactivation centre', *Nature*. Nature Publishing Group, 485(7398), pp. 381–385. doi: 10.1038/nature11049.

Olins, A. L. and Olins, D. E. (1973) 'Spheroid Chromatin Units', *Heart And Lung*, 183(8), pp. 330–332. Available at: http://science.sciencemag.org/content/sci/183/4122/330.full.pdf (Accessed: 1 July 2018).

van Otterdijk, S. D. and Michels, K. B. (2016) 'Transgenerational epigenetic inheritance in mammals: how good is the evidence?', *The FASEB Journal*. Federation of American Societies for Experimental BiologyBethesda, MD, USA, 30(7), pp. 2457–2465. doi:

10.1096/fj.201500083.

Papantonis, A. and Cook, P. R. (2010) 'Genome architecture and the role of transcription', *Current Opinion in Cell Biology*. Elsevier Current Trends, 22(3), pp. 271–276. doi: 10.1016/J.CEB.2010.03.004.

Parry, T. J. *et al.* (2010) 'The TCT motif, a key component of an RNA polymerase II transcription system for the translational machinery.', *Genes & development*. Cold Spring Harbor Laboratory Press, 24(18), pp. 2013–8. doi: 10.1101/gad.1951110.

Peterlin, B. M. and Price, D. H. (2006) 'Controlling the Elongation Phase of Transcription with P-TEFb', *Molecular Cell*. Cell Press, 23(3), pp. 297–305. doi: 10.1016/J.MOLCEL.2006.06.014.

Philip, P. *et al.* (2015) 'CBP binding outside of promoters and enhancers in Drosophila melanogaster', *Epigenetics & Chromatin*. BioMed Central, 8(1), p. 48. doi: 10.1186/s13072-015-0042-4.

Plaschka, C. *et al.* (2015) 'Architecture of the RNA polymerase II–Mediator core initiation complex', *Nature*. Nature Publishing Group, 518(7539), pp. 376–380. doi: 10.1038/nature14229.

Poleshko, A. *et al.* (2017) 'Genome-Nuclear Lamina Interactions Regulate Cardiac Stem Cell Lineage Restriction.', *Cell*. Elsevier, 171(3), p. 573–587.e14. doi: 10.1016/j.cell.2017.09.018.

Pombo, A. and Dillon, N. (2015) 'Three-dimensional genome architecture: players and mechanisms', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 16(4), pp. 245–257. doi: 10.1038/nrm3965.

Poynter, S. T. and Kadoch, C. (2016) 'Polycomb and trithorax opposition in development and disease.', *Wiley interdisciplinary reviews. Developmental biology*. NIH Public Access, 5(6), pp. 659–688. doi: 10.1002/wdev.244.

Pradeepa, M. M. (2017) 'Causal role of histone acetylations in enhancer function', *Transcription*. Taylor & Francis, 8(1), pp. 40–47. doi: 10.1080/21541264.2016.1253529.

Proudfoot, N. J. (2016) 'Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 352(6291), p. aad9926. doi: 10.1126/science.aad9926.

Rahl, P. B. *et al.* (2010) 'C-Myc regulates transcriptional pause release', *Cell*, 141(3), pp. 432–445. doi: 10.1016/j.cell.2010.03.030.

Rao, S. S. P. *et al.* (2014) 'A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping', *Cell*. Cell Press, 159(7), pp. 1665–1680. doi: 10.1016/J.CELL.2014.11.021.

Ringrose, L. (2007) 'Polycomb comes of age: genome-wide profiling of target sites', *Current Opinion in Cell Biology*. Elsevier Current Trends, 19(3), pp. 290–297. doi: 10.1016/J.CEB.2007.04.010.

Robertson, K. D. (2005) 'DNA methylation and human disease', *Nature Reviews Genetics*. Nature Publishing Group, 6(8), pp. 597–610. doi: 10.1038/nrg1655.

Sarma, K. *et al.* (2008) 'Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo.', *Molecular and cellular biology*. American Society for Microbiology, 28(8), pp. 2718–31. doi: 10.1128/MCB.02017-07.

Sawicka, A. and Seiser, C. (2014) 'Sensing core histone phosphorylation — A matter of perfect timing', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*. Elsevier, 1839(8), pp. 711–718. doi: 10.1016/J.BBAGRM.2014.04.013.

Saxonov, S., Berg, P. and Brutlag, D. L. (2006) 'A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters', *Proceedings of the National Academy of Sciences*, 103(5), pp. 1412–1417. doi: 10.1073/pnas.0510310103.

Schaaf, C. A. *et al.* (2013) 'Genome-Wide Control of RNA Polymerase II Activity by Cohesin', *PLoS Genetics*. Edited by B. Ren. Public Library of Science, 9(3), p. e1003382. doi: 10.1371/journal.pgen.1003382.

Schmitt, A. D., Hu, M. and Ren, B. (2016) 'Genome-wide mapping and analysis of chromosome architecture', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 17(12), pp. 743–755. doi: 10.1038/nrm.2016.104.

Schotta, G. *et al.* (2002) 'Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing.', *The EMBO journal*. EMBO Press, 21(5), pp. 1121–31. doi: 10.1093/emboj/21.5.1121.

Schotta, G. *et al.* (2004) 'A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin.', *Genes & development*. Cold Spring Harbor Laboratory Press, 18(11), pp. 1251–62. doi: 10.1101/gad.300704.

Schreieck, A. *et al.* (2014) 'RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7', *Nature Structural & Molecular Biology*. Nature Publishing Group, 21(2), pp. 175–179. doi: 10.1038/nsmb.2753.

Schuettengruber, B. *et al.* (2017) 'Genome Regulation by Polycomb and Trithorax: 70 Years and Counting', *Cell*. Cell Press, 171(1), pp. 34–57. doi: 10.1016/J.CELL.2017.08.002.

Segal, E. *et al.* (2006) 'A genomic code for nucleosome positioning', *Nature*. Nature Publishing Group, 442(7104), pp. 772–778. doi: 10.1038/nature04979.

Seisenberger, S. *et al.* (2013) 'Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers.', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. The Royal Society, 368(1609), p. 20110330. doi: 10.1098/rstb.2011.0330.

Sevier, S. A. and Levine, H. (2018) 'Properties of gene expression and chromatin structure with mechanically regulated elongation', *Nucleic Acids Research*. Oxford University Press, 46(12), pp. 5924–5934. doi: 10.1093/nar/gky382.

Sharif, J. *et al.* (2013) 'Embracing change to remain the same: conservation of polycomb functions despite divergence of binding motifs among species', *Current Opinion in Cell Biology*. Elsevier Current Trends, 25(3), pp. 305–313. doi: 10.1016/J.CEB.2013.02.009.

Sharma, A. (2013) 'Transgenerational epigenetic inheritance: focus on soma to germline information transfer.', *Progress in biophysics and molecular biology*, 113(3), pp. 439–46. doi: 10.1016/j.pbiomolbio.2012.12.003.

Shiraki, T. *et al.* (2003) 'Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 100(26), pp. 15776–81. doi: 10.1073/pnas.2136655100.

Shogren-Knaak, M. *et al.* (2006) 'Histone H4-K16 acetylation controls chromatin structure and protein interactions.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 311(5762), pp. 844–7. doi: 10.1126/science.1124000.

Smale, S. T. and Kadonaga, J. T. (2003) 'The RNA Polymerase II Core Promoter', *Annual Review of Biochemistry*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA, 72(1), pp. 449–479. doi: 10.1146/annurev.biochem.72.121801.161520.

Soubry, A. *et al.* (2014) 'A paternal environmental legacy: Evidence for epigenetic inheritance through the male germ line', *BioEssays*, 36(4), pp. 359–371. doi: 10.1002/bies.201300113.

Spitz, F. and Furlong, E. E. M. (2012) 'Transcription factors: from enhancer binding to developmental control', *Nature Reviews Genetics*. Nature Publishing Group, 13(9), pp. 613–626. doi: 10.1038/nrg3207.

Squazzo, S. L. *et al.* (2006) 'Suz12 binds to silenced regions of the genome in a cell-type-specific manner.', *Genome research*. Cold Spring Harbor Laboratory Press, 16(7), pp. 890–900. doi: 10.1101/gr.5306606.

Stegemann, R. and Buchner, D. A. (2015) 'Transgenerational inheritance of metabolic disease', *Seminars in Cell & Developmental Biology*. Academic Press, 43, pp. 131–140. doi: 10.1016/J.SEMCDB.2015.04.007.

Stevens, M. *et al.* (2013) 'Estimating absolute methylation levels at single-CpG resolution from methylation enrichment and restriction enzyme sequencing methods.', *Genome research*. Cold Spring Harbor Laboratory Press, 23(9), pp. 1541–53. doi: 10.1101/gr.152231.112.

Struhl, K. (1999) 'Fundamentally Different Logic of Gene Regulation in Eukaryotes and Prokaryotes', *Cell*. Cell Press, 98(1), pp. 1–4. doi: 10.1016/S0092-8674(00)80599-1.

Suzuki, S., Murakami, Y. and Takahata, S. (2017) 'H3K36 methylation state and associated silencing mechanisms.', *Transcription*. Taylor & Francis, 8(1), pp. 26–31. doi: 10.1080/21541264.2016.1246076.

Talasz, H. *et al.* (2005) 'Histone H4-lysine 20 monomethylation is increased in promoter

and coding regions of active genes and correlates with hyperacetylation.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 280(46), pp. 38814–22. doi: 10.1074/jbc.M505563200.

Tang, L., Nogales, E. and Ciferri, C. (2010) 'Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription.', *Progress in biophysics and molecular biology*. Howard Hughes Medical Institute, 102(2–3), pp. 122–8. doi: 10.1016/j.pbiomolbio.2010.05.001.

Tomkova, M. and Schuster-Böckler, B. (2018) 'DNA Modifications: Naturally More Error Prone?', *Trends in Genetics*. Elsevier Current Trends. doi: 10.1016/J.TIG.2018.04.005.

Valouev, A. *et al.* (2011) 'Determinants of nucleosome organization in primary human cells', *Nature*. Nature Publishing Group, 474(7352), pp. 516–520. doi: 10.1038/nature10002.

Veloso, A. *et al.* (2014) 'Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications.', *Genome research*. Cold Spring Harbor Laboratory Press, 24(6), pp. 896–905. doi: 10.1101/gr.171405.113.

Vidal, M. and Starowicz, K. (2017) 'Polycomb complexes PRC1 and their function in hematopoiesis', *Experimental Hematology*. Elsevier, 48, pp. 12–31. doi: 10.1016/J.EXPHEM.2016.12.006.

Waddington, C. H. (1942) 'The epigenotype', *Endeavour*, 1, pp. 18–20.

Wang, Z. *et al.* (2008) 'Combinatorial patterns of histone acetylations and methylations in the human genome', *Nature Genetics*. Nature Publishing Group, 40(7), pp. 897–903. doi: 10.1038/ng.154.

Weake, V. M. and Workman, J. L. (2010) 'Inducible gene expression: diverse regulatory mechanisms', *Nature Reviews Genetics*. Nature Publishing Group, 11(6), pp. 426–437. doi: 10.1038/nrg2781.

Weaver, I. C. G. *et al.* (2004) 'Epigenetic programming by maternal behavior', *Nature Neuroscience*. Nature Publishing Group, 7(8), pp. 847–854. doi: 10.1038/nn1276.

Wolf, J. B. and Wade, M. J. (2009) 'What are maternal effects (and what are they not)?', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. The Royal Society, 364(1520), pp. 1107–15. doi: 10.1098/rstb.2008.0238.

Wykes, S. M. and Krawetz, S. A. (2003) 'The structural organization of sperm chromatin.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 278(32), pp. 29471–7. doi: 10.1074/jbc.M304545200.

Xhemalce, B., Dawson, M. A. and Bannister, A. J. (2011) 'Histone Modifications', in *Encyclopedia of Molecular Cell Biology and Molecular Medicine*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. doi: 10.1002/3527600906.mcb.201100004.

Zeng, W., Ball, A. R. and Yokomori, K. (2010) 'HP1: Heterochromatin binding proteins working the genome', *Epigenetics*. NIH Public Access, 5(4), pp. 287–292. doi:

10.4161/epi.5.4.11683.

Zhang, T., Cooper, S. and Brockdorff, N. (2015) 'The interplay of histone modifications - writers that read.', *EMBO reports*. European Molecular Biology Organization, 16(11), pp. 1467–81. doi: 10.15252/embr.201540945.

Zlatanova, J. and Leuba, S. H. (2003) 'Chromatin Fibers, One-at-a-time', *Journal of Molecular Biology*. Academic Press, 331(1), pp. 1–19. doi: 10.1016/S0022-2836(03)00691-0.