From Department of Biosciences and Nutrition
Karolinska Institutet, Stockholm, Sweden

INTERPLAY OF CHROMATIN REMODELING, TRANSCRIPTIONAL REGULATION, AND NUCLEAR ORGANIZATION

Babett Steglich

Stockholm 2015
Interplay of chromatin remodeling, transcriptional regulation, and nuclear organization

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Babett Steglich

Principal Supervisor:
Professor Karl Ekwall
Karolinska Institutet
Department of Biosciences and Nutrition

Co-supervisor(s):
Professor Bas van Steensel
Netherlands Cancer Institute
Division of Gene Regulation

Opponent:
Dr. Emmanuelle Fabre
Université Paris Diderot
Chromosome Biology and Dynamics Group

Examination Board:
Docent Piergiorgio Percipalle
Karolinska Institutet
Department of Cell and Molecular Biology

Professor Neus Visa
Stockholms Universitet
Department of Molecular Biosciences

Professor Ann-Kristin Östlund-Farrants
Stockholms Universitet
Department of Molecular Biosciences
“They say a little knowledge is a dangerous thing, but it is not one half so bad as a lot of ignorance.”

Terry Pratchett (1948-2015)
ABSTRACT

Transcription is regulated on different levels to ensure that genes are expressed at the correct time and in the amounts required. At the chromatin level, DNA is wound onto histone proteins, forming nucleosomes that influence accessibility of DNA elements. Modifications on those histones and interactions with other chromatin proteins can either encourage or inhibit recruitment of the transcription machinery. Genomic regions of similar character form chromatin domains, organizing the genome based on their transcription states. Within the nucleus, both individual loci and entire chromosomes assume non-random positions, based on their transcription levels and interactions with nuclear landmarks. This thesis examines the effects of the Fun30 chromatin remodeling enzymes on transcription regulation and nuclear organization, both on the local chromatin level as well as on a genome-wide scale.

Using the fission yeast Schizosaccharomyces pombe as a model organism, we mapped the interactions between the genome and two inner nuclear membrane proteins, Ima1 and Man1. We observed a preference for lowly expressed genes to associate with the nuclear envelope, similar to what had been observed in mammalian and fruit fly cells. When comparing Ima1 and Man1 binding patterns, we found both common and separate target sites, suggesting a role for inner nuclear membrane proteins in organizing the fission yeast genome.

Following up on these results, we went on to examine subtelomeric chromatin domains, which are regulated through the Fun30 remodeler Fft3. These domains contain repressed genes, whose transcription levels increase in cells carrying an ftt3Δ deletion. While the subtelomeres associate with the nuclear envelope through Man1 in wild-type cells, this interaction is lost in ftt3Δ cells. In these cells, we also observed changes in nucleosome occupancy at the subtelomeric borders. Interestingly, a strain carrying a catalytically inactive version of the Fft3 remodeler showed the same behavior as the deletion strain, with upregulation of subtelomeric genes and loss of Man1 interactions. Together, these results point to an active role of Fft3 in regulating subtelomeric chromatin, transcription, and nuclear periphery interactions.

In addition to their role at subtelomeres, Fun30 remodelers also control transcription in other parts of the genome. When we examined a strain lacking Fft2, a paralog of Fft3, we found increased transcription of the fission yeast Tf2 retrotransposons. This increase is accompanied by a shift in transcription start site (TSS) further upstream and is especially pronounced when both ffl2 and ftt3 are deleted. By mapping nucleosome positioning, we were able to establish that Fft2 and Fft3 collaborate in stabilizing a nucleosome over the upstream TSS, resulting in transcription initiation further downstream and production of an mRNA incapable of transposition. Expression of both remodelers is downregulated in stress conditions, allowing for production of the longer transcript under these circumstances. We propose that the shift in TSS choice allows for bursts of transposition in cells under environmental stress. This can enable cells to adapt to changed conditions through favorable insertion events altering expression of nearby genes.
LIST OF SCIENTIFIC PAPERS

I. Babett Steglich, Guillaume J Filion, Bas van Steensel, and Karl Ekwall.
The inner nuclear membrane proteins Man1 and Ima1 link to two different
types of chromatin at the nuclear periphery in *S. pombe*.
*Nucleus*, 2012, 3 (1) pp. 77-87.

II. Babett Steglich*, Annelie Strålfors*, Olga Khorosjutina, Jenna Persson,
Agata Smialowska, Jean-Paul Javerzat, and Karl Ekwall.
The Fun30 chromatin remodeler Fft3 controls nuclear organization and
chromatin structure of insulators and subtelomeres in fission yeast

III. Jenna Persson, Babett Steglich, Agata Smialowska, Mette Boyd, Jette
Bornholdt, Robin Andersson, Albin Sandelin, Olaf Nielsen, and Karl Ekwall.
Regulating retrotransposons via alternative transcription start sites
*Manuscript.*

*equal contribution

RELATED PUBLICATIONS

Babett Steglich, Shelley Sazer, and Karl Ekwall.
Transcriptional regulation at the yeast nuclear envelope.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin Immunoprecipitation</td>
</tr>
<tr>
<td>CTD</td>
<td>C-terminal domain</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>DSB</td>
<td>Double strand break</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence In-situ Hybridization</td>
</tr>
<tr>
<td>GFP</td>
<td>Green Fluorescent Protein</td>
</tr>
<tr>
<td>HAT</td>
<td>Histone Acetyltransferase</td>
</tr>
<tr>
<td>HDAC</td>
<td>Histone Deacetylase</td>
</tr>
<tr>
<td>HKMT</td>
<td>Histone Lysine Methyltransferase</td>
</tr>
<tr>
<td>INM</td>
<td>Inner Nuclear Membrane</td>
</tr>
<tr>
<td>LAD</td>
<td>Lamina-associated domain</td>
</tr>
<tr>
<td>LINC</td>
<td>Linker of Nucleoskeleton and Cytoskeleton</td>
</tr>
<tr>
<td>LTR</td>
<td>Long Terminal Repeat</td>
</tr>
<tr>
<td>MNase</td>
<td>Micrococcal Nuclease</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>NDR</td>
<td>Nucleosome depleted region</td>
</tr>
<tr>
<td>NPC</td>
<td>Nuclear Pore Complex</td>
</tr>
<tr>
<td>ONM</td>
<td>Outer Nuclear Membrane</td>
</tr>
<tr>
<td>PIC</td>
<td>pre-Initiation Complex</td>
</tr>
<tr>
<td>Pol I/II/III</td>
<td>RNA polymerase I/II/III</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>SPB</td>
<td>Spindle pole body</td>
</tr>
<tr>
<td>TAD</td>
<td>Topologically associated domain</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription Start Site</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

1.1 A CHROMATIN PRIMER

1.1.1 Basic building blocks of chromatin

Fitting several meters of DNA into a cell nucleus at micrometer scale is not an easy task. The negatively charged sugar and phosphate backbone prevents it from folding back on itself, resulting in a persistence length of about 50 nm for naked DNA (Garcia et al., 2007). Eukaryotes, but also prokaryotic and archaeal species have therefore evolved proteins to act as spools for the DNA thread (Grove, 2011; Sandman and Reeve, 2006), neutralizing the negative charge and allowing for compaction. Together with DNA, histone proteins form the smallest unit of chromatin, the nucleosome. Each nucleosome consists of DNA wrapped around a core of eight histone proteins: a tetramer of two H3 and two H4 proteins, as well as two dimers consisting of one H2A and one H2B protein, respectively (Luger et al., 1997; Richmond and Davey, 2003).

The proposition of histones as packaging material came from X-ray diffraction studies (Finch and Klug, 1976), which led to the discovery of the 30 nm fiber. In this solenoid structure, nucleosomes are arranged in a staggered helix with a diameter of 30 nm, allowing for further compaction of the DNA. However, the 30 nm fiber has not been detected in interphase nuclei (Eltsov et al., 2008; Fussner et al., 2011; Joti et al., 2012). Instead, chromatin resembles beads on a string and forms a 10 nm fiber in these cases, with locally folded areas forming fractal globules (Lieberman-Aiden et al., 2009). This loose structure brings several benefits. Firstly, the DNA sequence becomes much more accessible to other proteins compared to the very dense 30 nm fiber. Secondly, it provides the possibility for histone proteins to act not merely as packaging molecules, but to interact with other factors and regulate DNA related processes such as transcription (Dekker, 2008; Fussner et al., 2011).Independent of which conformation it takes, chromatinized DNA is much less vulnerable to damage like ionizing radiation compared to naked DNA (Elia and Bradley, 1992).

It is important to note that chromatin is not uniform, but rather occurs in different configurations. When looking at microscopy images of cell nuclei, two different kinds of chromatin are apparent: a darker stained, more compacted form called heterochromatin, and the lighter, less condensed euchromatin. While heterochromatin mainly harbors genes with low levels of transcription, euchromatin contains actively transcribed genes. On a molecular level, the two types of chromatin differ in their protein composition: both regarding the nucleosome itself and other proteins associating with it. In addition, levels of cytosine methylation on the DNA (present in metazoan species) vary between euchromatin and heterochromatin. Chromatin states are necessary for genome organization and functions.
1.1.2 Histone modifications

Within the nucleosome structure, histones are arranged with their C-terminus at the center and their N-terminal tails protruding toward the outside (Luger et al., 1997; Richmond and Davey, 2003). Early on in chromatin research, these tails were found to carry various kinds of posttranslational modifications (Allfrey and Mirsky, 1964; Allfrey et al., 1964; Murray, 1964; Stevely and Stocken, 1966). These modifications can influence how tightly DNA is wound around the nucleosome core, but also affect protein interactions. It is therefore not surprising that mutations in the N-terminal tails were found to affect transcription (Han et al., 1988; Johnson et al., 1990). Since the discovery of these histone marks, efforts have been made to understand their significance and interplay, most prominently through the hypothesis of a “histone code” (Strahl and Allis, 1999). This theory proposes that combinations of different histone modifications define chromatin properties, with combinations specific for repressive or active chromatin, respectively.

Among the first modifications to be described was the acetylation of lysine residues (Allfrey and Mirsky, 1964). This primarily alters the charge of the lysine from positive to neutral and thereby weakens interactions between the negatively charged DNA backbone and the histone tails. Hyperacetylation, i.e. the acetylation of multiple residues, disrupts chromatin structure (Tse et al., 1998) and increases nucleosome mobility on DNA (Ferreira et al., 2007). The enzymes which set these histone marks, histone acetyltransferases (HATs), can be divided into two classes: type A HATs acetylate histones in the nucleus, while type B HATs are active in the cytoplasm and target free histones (Parthun, 2007). For both classes, target choice depends on co-factors acting in the same complex. The type A HAT Gcn5 in Saccharomyces cerevisiae for example targets free histones only when acting by itself, but acetylates nucleosomal histones when assembled into the SAGA complex (Grant et al., 1997). The reverse reaction, histone deacetylation, is catalyzed by histone deacetylase enzymes (HDACs). Based on sequence similarities, HDACs can be grouped into three classes (Ekwall, 2005). Which residue is deacetylated depends largely on other factors forming protein complexes with the enzyme, with HDACs acting in multiple complexes with different targets (Lalonde et al., 2014). In general, HATs are considered transcriptional activators, since hyperacetylation correlates with higher transcription levels, while HDACs act as co-repressors.

The effects of histone methylation are much less clear cut. Addition of a methyl group, neutral by itself, to lysine and arginine residues does not alter the charge of the amino acids. Moreover, up to three methyl groups can be added per residue, resulting in mono-, di- and trimethylation marks. Depending on which residue is methylated and to what extent, different chromatin proteins recognize the pattern and bind to the nucleosome. This leads to vastly different outcomes for methylation marks, a prime example being the lysine 4 and lysine 9 residues on histone H3 (H3K4 and H3K9). While methylation of H3K4 is associated with active transcription, H3K9 di- and trimethylation is found in repressed chromatin (Bannister and Kouzarides, 2011).
In contrast to HATs, lysine methyltransferases (HKMTs) have specificity for a defined residue and degree of methylation, independent of co-factors. Most HKMTs contain a SET domain which catalyzes the reaction. Independent of their mode of action, all HKMTs require the presence of S-adenosylmethionine (SAM) as a methyl-donor. Histone demethylases were unknown for a long time, until LSD1 was discovered in 2004 (Shi et al., 2004). Since it can de-methylate both H3K4 and H3K9, LSD1 can act as both co-repressor and co-activator.

Histone phosphorylation predominantly occurs on serine, threonine and tyrosine residues. Kinases like Aurora B can add a phosphate group, while phosphatases like PPI remove it (Goto et al., 2002; Oki et al., 2007; Sugiyama et al., 2002). Like acetylation, phosphorylation adds a negative charge to the histone tail, opening up chromatin structure. This is used especially during the cell cycle, when H3S10 is phosphorylated and thereby disrupts heterochromatin (Dormann et al., 2006).

In contrast to the other three histone marks in this chapter, ubiquitination is a large modification, involving the addition of a 76 amino acid polypeptide to the histone tail. Ubiquitin is added by E3-ligases to the tails of H2A and H2B. Among the modifications studied so far are H2AK119ub, which is involved in silencing, and H2BK123ub, which has been connected to transcriptional initiation and elongation (Kim et al., 2009; Lee et al., 2007a).

The effects of histone modifications are two-fold. Firstly, altering the charge of the histone protein can result in direct structural perturbations. Phosphorylation and acetylation decrease the charge of the residues they affect, leading to reduced interactions with nucleosomal DNA. This results in a looser nucleosome structure which is especially apparent in hyperacetylated histones. Regions with these high acetylation levels, such as promoter and enhancer regions, allow transcription factors and other proteins access to DNA (Kiefer et al., 2008; Wang et al., 2008).

Secondly, histone marks affect the binding of chromatin factors to the nucleosome. There is a myriad of proteins that interact with modified histones, which can be classified by the protein domains that facilitate this interaction. The largest group of domains is those that recognize methylation. Among these are chromo-, Tudor, MBT, and PHD domains (Musselman et al., 2012). While these domains all bind to the same modification, their specificity varies, with some recognizing more than one modified residue and multiple domains recognizing the same mark. Prominent examples are the chromodomains in CHD1, which recognizes H3K4me3 (Sims et al., 2005), and in HP1, which binds to H3K9me3 (Bannister et al., 2001; Lachner et al., 2001). A second important group are bromodomain proteins, which bind to acetylated residues. Among these are type A HATs, which contain bromodomains that help them recognize already mono-acetylated histones and add more acetylation marks (Dhalluin et al., 1999).
It is important to note that histone modifications do not occur in a vacuum, they are affected by other marks on the nucleosome and there is considerable crosstalk between them. Chemically, acetylation and methylation are antagonists, since lysine residues can either be acetylated or methylated, but not both. Some histone modifications depend on the presence of others, e.g. H3K4me and H3K79me depend on H2BK123ub (Kim et al., 2009; Lee et al., 2007a). In other cases, modification of one residue can impair protein binding to another, neighboring residue: e.g. HP1 binding to methylated H3K9 is prevented by phosphorylation of H3S10 (Fischle et al., 2005). There is also crosstalk between DNA methylation and histone modifications: some proteins bind to modified histones only when nucleosomal DNA is methylated or unmethylated. The PHD protein UHRF1 prefers binding to H3K9me3 on methylated DNA, while the lysine demethylase KDM2A only binds to H3K9me3 over unmethylated DNA (Bartke et al., 2010).

Away from the well-studied histone tail, modifications can also be placed on the histone core. These affect interaction with the DNA much more than the N-terminal tail modifications. H3K56 acetylation is very abundant in yeast and was shown to be involved in transcriptional regulation (Masumoto et al., 2005; Ozdemir et al., 2005; Xu et al., 2005). Since H3K56 is positioned at the entry point of DNA into the nucleosome, it allows for nucleosome ‘breathing’, i.e. transient DNA exposure from the nucleosome (Hyland et al., 2005; Masumoto et al., 2005; Xu et al., 2005).

### 1.1.3 Histone variants

In certain contexts, nucleosomes can contain various histone variants. These proteins each resemble one of the canonical histones and can take their place within a nucleosome. All eukaryotes contain CENP-A, a variant of H3 that is incorporated at centromeres and facilitates interaction with the kinetochore during cell division (Elsaesser et al., 2010; Malik and Henikoff, 2003). Two histone variants, H2A.Z and H3.3, are known to affect transcription. Both variants can occur together in hybrid nucleosomes, which are found at the transcription start site (TSS) of active genes (Jin et al., 2009). H3.3 can be acetylated on the histone core at K122, which disrupts histone-DNA interaction and helps de-stabilize the nucleosome (Simon et al., 2011; Tropberger et al., 2013).

Histone variants also play crucial roles in DNA repair pathways: H2A.X is phosphorylated by kinases activated through DNA damage. The resulting γ-H2A.X serves as a marker for the damaged site (Meyer et al., 2013). H2A.Z is also deposited at double strand break (DSB) sites by the p400 remodeler (Xu et al., 2012).
1.1.4 Chromatin remodeling enzymes

Although histone proteins have a natural affinity for DNA due to the opposite charges they carry, their positioning in vivo is controlled not only by the DNA sequence, but also a class of enzymes. These chromatin remodeling enzymes belong to the Snf2-family of proteins, which is part of the helicase-like SF2 superfamily (Flaus et al., 2006). To date, more than 1300 Snf2-enzymes have been identified, 30 alone in humans (Dürr et al., 2006; Flaus et al., 2006; Ryan and Owen-Hughes, 2011), that can be divided into at least 23 subfamilies. All of these enzymes contain an ATPase domain, which uses ATP hydrolysis to affect conformational changes in the enzyme and its substrates.

Chromatin remodelers can have different effects on the nucleosomes they interact with (Fig. 1). By sliding the nucleosome along the DNA or evicting a nucleosome from its position, they can make a DNA sequence elements accessible for other factors. Remodeling enzymes can also exchange histones, altering the composition of the nucleosome, e.g. by incorporation of a histone variant. By assembling nucleosomes from free histones or spacing nucleosomes on the DNA, chromatin remodelers also affect chromatin structure. The basic mechanism behind all these actions appears to based on structural changes in the nucleosome, followed by movement of DNA relative to the nucleosome core. Through pulling or pushing the DNA around the nucleosome, the enzyme is able to destabilize the nucleosome. This imbalance can then be used to exchange histones, affect disassembly/eviction or slide nucleosomes along the DNA (Lia et al., 2006; Saha et al., 2002; Singleton et al., 2007; Zhang et al., 2006).

Based on sequence similarity of their ATPase domain or the presence of additional domains in the protein sequence, Snf2-remodeling enzymes can be classified into subgroups (Clapier and Cairns, 2009; Flaus et al., 2006). The SWI/SNF group contains remodelers with a C-terminal bromodomain, enabling them to interact with acetylated histones. Correspondingly, the CHD remodelers can bind methylated histones through their N-terminal tandem chromodomains (Hauk et al., 2010). The ISWI group is characterized by a C-terminal SANT/SLIDE domain which localizes at linker DNA and helps position the remodeler for DNA translocation (Dang and Bartholomew, 2007; Ryan and Owen-Hughes, 2011). Finally, the INO80 group stands out because of its split ATPase domain.

It is worth noting that remodeling enzymes tend to act in complexes, although there are examples of enzymes acting alone. Among the co-factors occurring in complexes are histone chaperones, actin-related proteins and chromatin-binding proteins. Often one remodeling enzyme can act together with different sets of subunits in different complexes. The Brm remodeler in the fruit fly Drosophila melanogaster forms two complexes, BAP and PBAP, that affect distinct sets of targets (Mohrmann and Verrijzer, 2005). Furthermore, BAF-complexes in mammalian cells show cell-type specific compositions (Wang et al., 1996). In this manner, a relatively small set of remodelers can act in various different contexts and at different targets, with distinct outcomes.
Chromatin remodelers and their co-factors target specific regions through recognition of chromatin marks. The bromodomain present in e.g. RSC in yeast allows recruitment of the remodeler to actively transcribed genes through acetylated histone tails (Kasten et al., 2004). CHD1 is targeted in a similar manner: its chromodomain recognizes methylated H3K4 at the 5’ end of coding sequences (Sims et al., 2005). Some remodeling complexes contain subunits that can bind to methylated DNA sequences. The NuRD complex, which has both HDAC and chromatin remodeling activity (Xue et al., 1998), contains the methylCG binding MBD2/3 (Zhang et al., 1999) and is thought to contribute to DNA-methylation mediated repression. Histone variants, while being deposited by remodelers, can also aid or hinder recruitment of remodeling complexes (González-Romero et al., 2008).

Interestingly, chromatin remodelers can be regulated by the same enzymes that also modify histones. In yeast, the HAT Gcn5 can acetylate the Rsc4 subunit of the RSC complex, thereby inhibiting the interactions between RSC and histones acetylated on H3K14 (Choi et al., 2008). This has been proposed as a fine-tuning mechanism for the interaction between RSC and chromatin. Other examples are PARylation through PARP-1, which regulates ISWI remodeling (Sala et al., 2008), and phosphorylation of the RSC subunit Sfh1 through a kinase involved in DNA damage response (Cao et al., 1997).

When it comes to functionality, the biggest effects of chromatin remodelers are in activation and repression of transcription. Some transcription factors are able to recruit remodeling enzymes directly to a gene, as in the case of Swi/Snf (Fry and Peterson, 2001; Vignali et al., 2000). Alternatively, remodelers can bind to nucleosomes through recognition of histone acetylation, e.g. through the bromodomain of Gcn5 in the yeast SAGA complex (Li and Shogren-Knaak, 2009). In this way, Gcn5 can recruit SWI/SNF to inducible genes in yeast (Cosma et al., 1999; Natarajan et al., 1999) and to DNA damage sites in humans (Lee et al., 2010). It should be pointed out that there is considerable functional overlap between remodelers, since very few genes depend in their expression on only one remodeling enzyme. Chromatin remodelers also affect transcriptional elongation, through disassembly of nucleosomes ahead of RNA polymerase and assembly behind it.

An important function of chromatin remodelers is to establish and maintain proper nucleosome positioning. In yeast, CHD1 is involved in nucleosome spacing (Gkikopoulos et al., 2011; Hennig et al., 2012; Pointner et al., 2012; Shim et al., 2012), while Isw2 influences the position of the +1 nucleosome (Whitehouse et al., 2007; Yen et al., 2012). Several enzymes are involved in nucleosome eviction, such as RSC (Badis et al., 2008; Hartley and Madhani, 2009) and SWI/SNF (Dechassa et al., 2010; Engeholm et al., 2009; Ulyanova and Schnitzler, 2005).

By exchanging histones, chromatin remodeling enzymes can alter chromatin composition. Among the remodelers with histone exchange capabilities are Swr1, Fun30 and Ino80 (Awad et al., 2010; Mizuguchi et al., 2004; Papamichos-Chronakis et al., 2011). In vivo studies have shown that all three enzymes affect H2A.Z distribution (Durand-Dubief et al., 2012; Mizuguchi et al., 2004; Papamichos-Chronakis et al., 2011). The budding yeast Fun30
remodeler is also involved in re-establishment of histone modifications after DNA replication (Rowbotham et al., 2011) and affects rate and extent of strand resection in double strand break (DSB) repair (Chen et al., 2012; Costelloe et al., 2012; Eapen et al., 2012).

Since they fundamentally shape chromatin structure, it is not surprising that chromatin remodelers are important during development. One example are the BAF complexes in mammals, which feature different subunit compositions at different stages during development (Son and Crabtree, 2014). This results in activation of distinct genetic programs, most likely through interactions with cell-type specific transcription factors. Consistent with their role in development, BAF proteins act as tumor suppressors, as they facilitate differentiation (Wang et al., 2014). Mutations in remodelers have been found in various tumors, with mutations in SWI/SNF components occurring at a frequency of 19% across different human cancers (p53 is at 26%, for comparison) (Kadoch et al., 2013).

![Diagram of chromatin remodeling factors](image)

**Figure 1:** Effects of chromatin remodeling factors on nucleosome positioning and composition.

### 1.1.5 Nucleosome positioning

Two terms are generally applied to describe a nucleosome: its positioning and its occupancy. Positioning reflects where a nucleosome is located with respect to the underlying genomic DNA sequence. Depending on how defined this location is, nucleosomes can be described by perfect positioning (fixed nucleosome positions), partial positioning (nucleosomes that are loosely associated with DNA), and depletion (in case no nucleosome is present). Nucleosome occupancy is a measure averaged over a population of cells and describes the fraction of cells in which a given region is occupied by a histone octamer (Struhl and Segal, 2013). Both positioning and occupancy influence accessibility of DNA for all proteins interacting with it (Lam et al., 2008; Mavrich et al., 2008a; Raveh-Sadka et al., 2012; Shivaswamy et al., 2008).
There are several general nucleosome positioning patterns present in all eukaryotes. Most prominent among these is the nucleosome depleted region (NDR), which is positioned over the promoter region of genes and allows access for the transcription machinery. The depth of the NDR, i.e. the degree of nucleosome depletion, correlates with transcriptional activity, with highly transcribed genes showing a deeper NDR (Lee et al., 2007b; Schones et al., 2008; Valouev et al., 2011; Weiner et al., 2010). To increase accessibility for the interacting proteins, many enhancers, insulator sites and terminator regions are also depleted of nucleosomes (Fu et al., 2008; Heintzman et al., 2007; Mavrich et al., 2008a). Both genes and non-gene regions show preferred positioning patterns (Lee et al., 2007b; Mavrich et al., 2008a; 2008b; Schones et al., 2008; Shivaswamy et al., 2008; Valouev et al., 2011; 2008; Yuan et al., 2005). Especially the nucleosomes bordering the NDR at the -1 and +1 positions are strongly fixed in place, with nucleosomes arranged in arrays pointing outward in both directions (Mavrich et al., 2008a; Shivaswamy et al., 2008). Within a gene, exons tend to show higher nucleosome occupancy than introns, possibly due to the higher GC content in those sequences (Schwartz et al., 2009; Tilgner et al., 2009).

While these general positioning patterns are largely dictated by the functionality of the underlying sequence elements, the DNA sequence itself influences nucleosome positioning on a local level. The interactions between DNA and histone proteins do not occur through precise base pair-protein binding, but rather rely on the ability of the DNA to bend around the histone core. Ideal wrapping of DNA around the histone core relies on bendable base dimers (AT/TT/AA) facing the core and stiffer dimers (CG/GG/CC) facing outward. Alternating these features periodically will result in a stable nucleosome array (Drew and Travers, 1985). In contrast, long stretches of poly(dA:dT) and poly(dG:dC) are very stiff and refractory to nucleosome formation (McCall et al., 1985; Nelson et al., 1987; Suter et al., 2000). This property is used in eukaryotic genomes, where poly(dA:dT) sequences are abundant (Dechering et al., 1998). These sequences are for example found in budding yeast promoters (Field et al., 2008; Struhl, 1985), where they are important for nucleosome depletion, promoter accessibility and transcriptional activity (Iyer and Struhl, 1995; Raveh-Sadka et al., 2012; Segal and Widom, 2009). In general, promoter and terminator regions in S. cerevisiae are intrinsically unfavorable for nucleosome formation (Iyer and Struhl, 1995; Kaplan et al., 2009; Sekinger et al., 2005; Zhang et al., 2009).

Apart from sequence features, many factors affect nucleosome positioning and occupancy. First and foremost are chromatin remodeling complexes (Hughes and Rando, 2014). With the exception of the RSC complex (Badis et al., 2008; Floer et al., 2010), these enzymes do not show sequence specificity. Instead, they are recruited to target regions through chromatin properties, such as histone modifications or DNA methylation. The Isw2 remodeler is involved in positioning the +1 nucleosome (Whitehouse et al., 2007), while different remodelers affect the downstream nucleosome array (Isw1/Chd1 in S. cerevisiae (Gkikopoulos et al., 2011), CHD-remodelers Hrp1/Hrp3 in S. pombe (Hennig et al., 2012; Pointner et al., 2012; Shim et al., 2012)). Ino80 and Swr1 both affect positioning (van Bakel
et al., 2013; Yen et al., 2012), although this might be due to downstream effects of H2A.Z insertion by dimer exchange (Li et al., 2005).

Recruitment of chromatin remodelers can also occur through pioneer transcription factors such as FoxA and GATA. By binding DNA directly on the nucleosome (Yu and Morse, 1999; Zaret and Carroll, 2011), they can serve as an anchor for remodeling enzymes, which open up the chromatin and free up space for further transcription factors. In addition, general regulatory factors like Abf1, Reb1 and Rap1 can evict histones in yeast, most likely also through recruitment of remodelers (Yarragudi et al., 2004; Yu and Morse, 1999).

As can be expected, the passing of RNA polymerase through a gene affects nucleosome positioning. Transcription can override any sequence-based positioning, since nucleosome positioning matches in silico models better in cells without active transcription (Moshkin et al., 2012; Weiner et al., 2010). Similarly, in vitro assembly without RNA polymerase and inactivation of the enzyme both result in a +1 nucleosome further downstream from the promoter than observed in vivo (Zhang et al., 2011). It has been proposed that nucleosomes are pushed upstream when RNA polymerase passes through the gene, possibly by a bubble of DNA that is pushed around the octamer (Bintu et al., 2011; Studitsky et al., 1994).

With the DNA sequence influencing nucleosome positioning in cis and proteins like remodelers affecting it in trans, the question arises which of these factors plays the stronger role. General patterns, such as the position of the +1 nucleosome, occur independent of the underlying DNA sequence (Kaplan et al., 2009; Zhang et al., 2009). While in vitro nucleosome positioning resembles the in vivo pattern, the addition of cell-free extract and ATP is necessary to obtain the same positioning (Zhang et al., 2011). The in silico models developed over the last decade perform best if they account not just for (dA:dT) periodicity, but also for (dA/dT)-rich sequences in promoters (Field et al., 2008; Tillo and Hughes, 2009; Yuan and Liu, 2008). It has therefore been suggested that sequence features like (dA:dT) periodicity fine-tune nucleosome positioning by +/- 5 bp, while trans factors regulate positioning on a larger scale (Brogaard et al., 2012). There is however some interaction between cis and trans factors, since some remodelers prefer pro- or anti-nucleosomal sequences. The Isw2 remodeler, for example, is able to deposit nucleosomes over anti-nucleosomal sequences (Whitehouse and Tsukiyama, 2006; Whitehouse et al., 2007), while SWI/SNF reinforces positioning at pro-nucleosomal sequences (Tolkunov et al., 2011).

Another angle to look at nucleosomes comes through nucleosome spacing, which refers to the length of linker DNA between nucleosomes in an array. Linker DNA length varies not only between species, but also between tissues and genes within a species (Hughes et al., 2012). Spacing correlates with and depends on transcription rate, with tighter spacing over highly expressed genes (Valouev et al., 2011; Weiner et al., 2010). In metazoans, the linker histone H1 is a major regulator of linker DNA length, since its presence increases the spacing between adjacent nucleosomes (Drew and Travers, 1985; Fan et al., 2005; Hashimoto et al., 2010; Öberg et al., 2012). This influence is reduced in yeasts, where H1 is structurally
different and expressed at very low levels compared to the core histones (Freidkin and Katcoff, 2001).

Overall, nucleosome positioning has far reaching consequences, not just for DNA elements in detail, but genome-wide in maintaining homeostasis. Depletion of histones leads to a genome-wide increase in expression levels, showing their importance in the regulation of transcription. Interestingly, some genes are down-regulated when histone levels are reduced. This could be due to lack of antisense RNA suppression (Martens et al., 2004), reduction in binding of transcription factors that prefer binding to DNA curved around a nucleosome (McPherson et al., 1993), or because nucleosomes can shorten the distance between two regulatory sequences that need to be close together to be recognized (Lu et al., 1995; Stünkel et al., 1997). Taken together, these observations highlight the importance of nucleosome positioning and spacing in genome function.

1.2 CHROMATIN DOMAINS AND THEIR BOUNDARIES

The traditional view divided chromatin into two types, heterochromatin and euchromatin. This was mainly based on cytological observations, but also on the assumption that a cell needed only two chromatin states, one actively transcribed and open, the other repressed and compacted. Several studies in metazoans have since shown that reality is more complex, with different ‘colors’ of chromatin rather than the black and white of heterochromatin and euchromatin (Ernst et al., 2011; Filion et al., 2010; Kharchenko et al., 2011; Liu et al., 2011; Ram et al., 2011; Zhu et al., 2013).

The number of existing chromatin states is still under debate, as the studies mentioned above found between three and nine different flavors. All of them find at least one active chromatin state, although it is sometimes subdivided further. This active chromatin is characterized by high expression levels, early initiation of DNA replication and a variety of histone modifications, such as methylation of H3K4, H3K36 and H3K79, as well as hyperacetylated histone tails (Filion et al., 2010; Kharchenko et al., 2011).

Repressed chromatin can be further divided into at least three types: Polycomb-repressed chromatin, ‘null’ chromatin and constitutive heterochromatin. While they all contain silenced genes, these types differ in protein composition and behavior. Polycomb-repressed chromatin contains high levels of the H3K27me3 mark and covers developmental genes, such as the Hox gene clusters, X-inactivation sites and imprinted regions (Schwartz and Pirrotta, 2013). In contrast, ‘null’ chromatin (or ‘black’ chromatin (Filion et al., 2010)) shows no enrichment for any particular histone modification. Its most evident feature is a strong enrichment for lamins and thereby a preference for the nuclear periphery. Lastly, constitutive heterochromatin covers those regions crucial for chromosome stability and propagation: centromeres and telomeres (Ekwall et al., 1995; Peters et al., 2001). These regions contain fewer genes and more sequence elements such as tandem repeats, satellite DNA and transposable elements. Constitutive heterochromatin is characterized by high levels of
H4K20me3, H3K9me2/me3 and HP1a (heterochromatin protein 1a) binding. While average transcriptional activity is lower than in active chromatin, it is still higher compared to the other two repressed chromatin types (Filion et al., 2010).

To date, no comparable study of chromatin flavors has been done in yeasts, but it is likely that at least three chromatin types would be found in fission yeast. Apart from an active chromatin state, the fission yeast genome contains constitutive heterochromatin at the centromeres, telomeres and the mating type locus (Grewal, 2000). In addition, the subtelomeric regions are characterized by a distinct chromatin type, lacking both methylated H3K9 (found in heterochromatin) and methylated H3K4 (found at actively transcribed genes) (Buchanan et al., 2009; Zofall et al., 2009).

Independent of how many types of chromatin exist, the question remains how chromatin domains are established and maintained. Theoretically, an enzyme freely diffusing through the nucleus could come in contact with any part of the genome and modify nucleosomes. In reality however, movement in the nucleus is restricted, with entire chromosomes moving very little, but small domains being able to move more quickly over short distances (Erdel et al., 2013). By measuring the reach of a chromatin-bound enzyme and through mapping of recombination events, studies were able to determine that most of these interactions happen on a scale of a few kilobases (Ringrose et al., 1999; van Steensel and Henikoff, 2000). In this manner, enzymes immobilized on chromatin or at the nuclear envelope can target only a small region of the genome. In some cases, this immobilization is even used to control activity, e.g. the guanine nucleotide exchange factor RCC1 is more active in its chromatin bound state than freely diffusing (Nemergut et al., 2001). Chromatin compaction affects accessibility, which in turn dictates which proteins can interact with genome regions. Very large protein complexes (in the MDa range) are excluded from compacted chromatin, while smaller complexes (kDa size) are able to gain access (Erdel et al., 2013).

To prevent the spreading of one chromatin type into another, it is necessary to define and enforce domain borders. This is achieved through insulators: DNA elements which, when removed, result in loss of boundary function (Ong and Corces, 2009). While they were identified in *Drosophila*, insulators have since been found in many other eukaryotes (Kirkland et al., 2012; Vogelmann et al., 2011), with sizes ranging from several hundred base pairs to more than 1 kb (Chetverina et al., 2014). Insulators can also be defined through their ability to block enhancer activity or silencing. To achieve these functions, insulator sites either show a particular chromatin structure or are bound by insulator proteins. Marks like histone acetylation, H3K9 monomethylation and H2A.Z incorporation have been associated with insulator activity (Barski et al., 2007). The most well-studied insulator protein is CTCF, which binds between 13,000 and 30,000 target sites, depending on cell line (Cuddapah et al., 2009; Kim et al., 2007) and is thought to be the only vertebrate insulator protein (Bell et al., 1999). As a possible mechanism, it has been proposed that CTCF collaborates with cohesin to organize chromatin domains into loops, thereby physically separating different chromatin types (Carretero et al., 2010).
1.3 TRANSCRIPTION

One of the most important processes in a cell is arguably transcription. A single mammalian cell typically contains around 200,000 mRNA molecules (Shapiro et al., 2013), which are being produced at speeds between 2 and 4 kbp/min (Fuchs et al., 2014; Tennyson et al., 1995). This task is accomplished through three RNA polymerase (Pol) complexes, with Pol I transcribing ribosomal RNA, Pol III responsible for small RNAs like tRNAs and Pol II covering mainly mRNA.

To perform the complex steps required to transcribe a DNA sequence into RNA, Pol II forms a 500 kDa complex consisting of 12 subunits (Armache et al., 2003; Bushnell and Kornberg, 2003). However, various transcription factors, co-activators and chromatin proteins are necessary for efficient transcription, justifying the term ‘transcription machinery’. The process of transcription can be divided into three steps: initiation, elongation and termination, each associated with its own regulation and co-factors.

In the first step in transcription initiation, activator proteins bind upstream of the gene promoter, where they recruit co-activators such as Mediator and the SAGA complex. These co-activators alter the chromatin structure to allow access and assembly of the transcription machinery. Pol II can now come together with general transcription factors to form the pre-initiation complex (PIC). Among the general transcription factors, the TATA-Box binding protein (TBP) and TFIIB form the core of the PIC and are indispensable for transcription, whereas TFIIE is not required (Hausner et al., 1996; Parvin and Sharp, 1993; Qureshi et al., 1997). The direction of transcription is determined by TFIIB recognition elements upstream and downstream of the TATA box (Deng and Roberts, 2005; Lagrange et al., 1998; Littlefield et al., 1999; Qureshi and Jackson, 1998; Tsai and Sigler, 2000), although divergent transcription has been observed in yeast and mammals (Churchman and Weissman, 2011; Seila et al., 2008). To start transcribing, the PIC needs to move from a closed state, with DNA still double-stranded, to an open state, with a transcription ‘bubble’ forming and exposing about 15 bp of unwound, single-stranded DNA. This process requires ATP-dependent helicase activity and is described as ‘promoter melting’. The distance between the TATA box and the transcription start site (TSS) is around 30 bp in almost all genes that contain a TATA box (Smale and Kadonaga, 2003). In yeast cells, this distance is larger, between 40 and 120 bp (Struhl, 1989). It has been proposed that Pol II follows the same promoter melting process in yeast, but then scans for the TSS downstream of the TATA box (Giardina and Lis, 1993). This is supported by the observation that mutations in TFIIB, TFIIF, or Pol II subunits can affect TSS location (Ghazy et al., 2004; Hull et al., 1995; Pinto et al., 1994).

The last step of initiation and first step of elongation is called promoter escape. During this phase, TFIIB leaves the PIC, interactions between DNA and TBP are broken, and Pol II can start transcribing (Dvir, 2002). From about 25 nt of nascent RNA, Pol II forms a stable elongating complex. Having completed their task, most general transcription factors are released at this stage. At the same time, the nascent RNA receives its 7-methylguanylate cap,
increasing its stability. Pol II can now move stepwise through the coding-sequence, adding nucleotides complementary to the exposed DNA sequence in a ratchet-like manner. Potential errors can be corrected due to the proof-reading capability of Pol II, leading to a very low error rate of $\sim 4\times 10^{-6}$ (Gout et al., 2013).

Once the end of the coding sequence is reached, a polyadenylation signal is added into the mRNA, leading to termination of transcription (Lykke-Andersen and Jensen, 2007). This signal is recognized by cleavage and polyadenylation factors, which cut the nascent RNA and release it from the Pol II complex. A polyA-tail is now added and splicing occurs co-transcriptionally, resulting in an mRNA that can be translated into protein.

The process of transcription is tightly regulated, both on the chromatin level and on RNA polymerase itself. The C-terminal domain (CTD) of the Rpb1 subunit contains up to 52 repeats of a peptide sequence that can be phosphorylated at different positions (Meinhart and Cramer, 2004). While in the PIC, the CTD is unphosphorylated, however phosphorylation on Serine 5 (Ser5ph) is required for promoter escape. Throughout elongation, Ser5ph and Ser7ph are present, while Ser2ph is found when Pol II reaches the termination stage (Brookes and Pombo, 2012; Chapman et al., 2007; Corden, 1993; Komarnitsky et al., 2000; Marshall et al., 1996a). Histone modifications are also used to mark different sections within the coding sequence. While the 5’ end of the gene is dominated by acetylated histone tails and H3K4me3, the 3’ end is marked by H3K36me3 (Smolle and Workman, 2013). The H3K36me2 and H3K79me marks are present throughout the gene body. These modifications are thought to help define the different regions and prevent aberrant intragenic transcription.

### 1.4 RNA INTERFERENCE – A LINK BETWEEN TRANSCRIPTION AND SILENCING

The RNA interference (RNAi) machinery was originally discovered in the worm *Caenorhabditis elegans*, where it provides a way to regulate mRNA levels in the cytoplasm through complementary small RNAs (Mello and Conte, 2004). In this post-transcriptional gene silencing (PTGS) mechanism, the endonuclease Dicer cuts double stranded RNA into small RNAs (called siRNA), which can be bound by the Argonaute protein and turned into single stranded RNA. The small RNA recognizes its complementary mRNA in the cytoplasm, bringing it into contact with Argonaute and its binding partners, which form the RISC complex. In this close proximity, RISC can now cleave the target mRNA, preventing its translation.

Since PTGS had been described to act on the mRNA level, it came as a surprise that mutations in RNAi components alleviate silencing at pericentric heterochromatin in *S. pombe* (Volpe et al., 2002). This posed the question how RNAi is able to influence gene expression on the chromatin level. Interestingly, the pericentric repeats are not entirely silent, but show a burst of transcription during the S-phase of the cell cycle (Kloc et al., 2008). These low levels of transcription are required for pericentric silencing (Djupedal et al., 2005; Kato et al., 2005;
Volpe et al., 2002). The nascent RNA forms short double stranded sections, either through primal small RNA (Halic and Moazed, 2010) or through folding into secondary structures (Djupedal et al., 2009). These serve as a primer for an RNA-dependent RNA polymerase to synthesize a complementary RNA strand (Motamedi et al., 2004). The resulting dsRNA can be cleaved into small RNAs by Dicer, similar to its role in PTGS. Instead of targeting the RNA for cleavage, the small RNAs are incorporated into Argonaute and the RITS complex (Verdel et al., 2004). RITS then in turn recruits Clr4, the HKMT that methylates H3K9 in fission yeast, establishing heterochromatin over the pericentric repeats. The chromodomain of Chp1, a subunit of the RITS complex, can then bind methylated H3K9 and further anchor the complex. This mechanism of small RNA targeted heterochromatin formation has since been named transcriptional gene silencing (TGS) and parallels have been found in other organisms (Bourc'his and Voinnet, 2010; Huisinga and Elgin, 2009; Malecová and Morris, 2010; Matzke et al., 2009).

1.5 TRANSPOSABLE ELEMENTS

While protein coding genes and their regulation are the focus of most research efforts, the genome contains many other elements of interest. Transposable elements stand out among these: they can make up a large proportion of a species’ genome (e.g. 75% of the Zea mays genome (SanMiguel et al., 1996; Schnable et al., 2009)), but it remains a matter of debate whether they are actually useful for the cell. Transposons have also been described as ‘selfish DNA’, since the proteins they encode only affect their own movement and survival. There are two main types of transposable elements: DNA transposons and RNA transposons, also called retrotransposons (Levin and Moran, 2011). Compared to retrotransposons, DNA transposons do not propagate. Instead, they use a cut-and-paste mechanism, jumping from one location to another. In contrast, retrotransposons use an RNA intermediate for propagation. First, Pol II transcribes the transposon, with the resulting mRNA encoding for the reverse transcriptase and integrase enzymes. The reverse transcriptase then uses the mRNA to make cDNA, which can be inserted into a new locus by the integrase. Based on their structure and the priming mechanism used during reverse transcription, retrotransposons are divided into long terminal repeat (LTR) retrotransposons and non-LTR retrotransposons. Both classes are abundant in mammals, e.g. the non-LTR LINE-1 and SINE elements make up about 30% of the human genome (Lander et al., 2001).

The choice of integration site is critical for the survival of transposable elements. If the insertion causes a detrimental mutation, both the host cell and the transposons in its genome are eliminated. Therefore, transposable elements have evolved to prefer insertion sites that do not negatively affect the fitness of the host cell, but also ensure their own propagation. One strategy is to insert into heterochromatin domains or gene-poor regions, as in the case of the budding yeast Ty5 transposon (Zou and Voytas, 1997; Zou et al., 1996; 1995). Some transposons have evolved to target gene-rich regions, but insert outside of the coding sequence. In S. cerevisiae, the Ty1 and Ty3 transposon integrate upstream of Pol III.
transcribed genes (Bushman, 2003; Devine and Boeke, 1996; Lesage and Todeschini, 2005). Similarly, the fission yeast Tf1 transposon targets promoters of Pol II transcribed genes, especially stress response genes regulated by the transcription factor Atf1 (Behrens et al., 2000; Guo and Levin, 2010; Majumdar et al., 2011; Singleton and Levin, 2002). This strategy is thought to be beneficial for the host cell, since transposon integration can alter the expression of adjacent genes when they are inserted (Leem et al., 2008). It has been proposed that this provides a way to increase genetic diversity and adaptability to stress conditions, positively affecting the fitness of the host (McClintock, 1984). Some retroviruses, which are thought to be related to retrotransposons, show the same preferential integrase behavior, e.g. HIV-1 (Mitchell et al., 2004), although in this case likely to escape detection.

Transposon integration not only poses a threat through possible mutations, but also increases the amount of DNA to be copied in each cell cycle, slowing down growth. Host organisms have therefore evolved strategies to prevent uncontrolled spreading of the transposons throughout the genome. DNA methylation has proved to be useful for repression of retrotransposons in organisms with larger genomes (Selker et al., 2003; Yoder et al., 1997), as deletion of the DNA methyltransferase DNMT-3L in mice leads to reactivation of retrotransposons (Bourc'his and Bestor, 2004). RNAi provides a mechanism to combat retrotransposon spreading by digesting the transposon RNA before it can be turned into cDNA. In one pathway, the RNAi response is triggered by dsRNA stretches in the transposon mRNA, resulting in Dicer recruitment, siRNA production and degradation through RISC (Malone and Hannon, 2009). A second strategy is the use of piRNAs produced from clusters of transposon-derived sequences. These are then recognized by PIWI/Aubergine proteins which cleave the transposon mRNA. The resulting small RNAs can then be used to degrade further mRNAs, creating an amplification cycle (Aravin et al., 2007).

Irrespective of these strategies, transposons have penetrated nearly all species. However detrimental they are, there are more and more hints that transposable elements can also be beneficial for an organism. As mentioned above, transposons are thought to increase genome plasticity and adaptability to stress conditions. Some organisms allow limited transposition events in germ cells and early development, increasing genetic diversity in future generations. Transposition events in somatic cells are thought to contribute to brain plasticity (Muotri et al., 2010; Rehen et al., 2001; Westra et al., 2010). In the long run, it might be worth for a cell to handle the extra load of transposons during replication for the benefit of greater adaptability.
1.6 NUCLEAR ORGANIZATION

1.6.1 Basics of nuclear architecture

When discussing transcription, DNA repair, or DNA replication, the location where these processes take place is often neglected. The nucleus is not only the cellular compartment containing chromosomal DNA, surrounded by nuclear envelope. It is specialized to store, maintain, replicate and transcribe those chromosomes and contains a variety of substructures that support these processes.

Defining the borders of the nucleus, the nuclear envelope is formed by the inner (INM) and outer nuclear membrane (ONM), which are separated by the 40-50 nm wide perinuclear space (Burke and Stewart, 2014). The ONM is continuous with the endoplasmatic reticulum (ER) (Callan and Tomlin, 1950) and contains proteins required to position and stabilize the nucleus within the cell (Hetzer, 2010). The INM contains a multitude of transmembrane proteins (Schirmer and Foisner, 2007) that protrude into the nuclear interior. In addition, metazoan nuclei contain the nuclear lamina, a protein network lining the nuclear envelope which provides stability and organizes chromatin. The exchange between nucleus and cytoplasm occurs through nuclear pore complexes, which are embedded into the nuclear envelope.

Within the nucleoplasm, the most prominent landmark are the nucleoli, the compartments where rRNA is transcribed and assembled into ribosome subunits. In addition, many smaller structures have been described, such as paraspeckles, PML bodies and Cajal bodies, which are involved in RNA processing (Mao et al., 2011).

1.6.2 Levels of nuclear organization

The simplest organization of the genome becomes apparent by a look through the microscope. Heterochromatin was first described in the 1920s, as the material at the nuclear periphery stained darker than the rest of the chromatin (Heitz, 1928). The distribution of heterochromatin depends on species and cell type and does not reveal much about the functionality of chromatin apart from its position.

Moving down to individual chromosomes, a different picture emerges (Fig. 2). Rather than intermingling indiscriminately, chromosomes take up territories in the nucleus (Cremer et al., 2000). This arrangement favors intra-chromosomal over inter-chromosomal contacts. The degree of intermingling between territories is controversial, with some favoring the model of a chromatin-free space between the territories and others arguing that there is substantial contact between chromosomes (Branco and Pombo, 2006; Cremer et al., 2000; Markaki et al., 2012; Olivares-Chauvet et al., 2011). In addition, the relative position of individual chromosomes in the nucleus is not random. Instead, defined pairings of chromosomes are found in many cell types (Cremer et al., 2001; Dundr and Misteli, 2001; Kuroda et al., 2004; Parada et al., 2004). The position of chromosomes with respect to the nuclear envelope
depends on their gene density (Boyle et al., 2001; Cremer et al., 2001; Croft et al., 1999), with gene-poor chromosomes closer to the nuclear envelope. The level of transcription also matters, e.g. the active and inactive X-chromosomes in female mammals are positioned differently (Chaumeil et al., 2006).

On each chromosome, the telomeres and centromeres can have specific localization patterns, depending on the species. The centromeres in fission yeast cluster at the nuclear periphery, opposite the spindle pole body (SPB), throughout the cell cycle (Funabiki et al., 1993; Uzawa and Yanagida, 1992). In plants, centromeres move into this conformation during meiosis (Bass et al., 1997; Cowan et al., 2001). At the ends of the chromosomes, the telomeres also tend to prefer certain positions. While in yeasts telomeres attach permanently to the nuclear envelope (Galy et al., 2000; Scherthan et al., 2000), this occurs only transiently in metazoans during meiotic recombination. Both centromere and telomere tethering is thought to occur through SUN proteins (Bupp et al., 2007; Chikashige et al., 2006; Conrad et al., 2007; Hou et al., 2012; Schmitt et al., 2007). In fission yeast, the SUN protein Sad1 was shown to interact with the kinetochore protein Csi1, anchoring the centromeres (Hou et al., 2012).

Within the chromosome territories, chromatin folds itself into modular domains. These topologically associated domains (TADs) are formed from a continuous chromatin strand that folds into a globular structure, with more internal interactions than contacts to the outside (Dixon et al., 2012; Lieberman-Aiden et al., 2009; Sexton et al., 2012). The size of these domains ranges from 100 kb to 10 Mb in D. melanogaster (Sexton et al., 2012), and show a median size of 880 kb in mouse embryonic stem cells. TAD borders are enriched for CTCF binding sites and housekeeping genes (Dixon et al., 2012), and, although usually well-defined, can shift between cell types (Andrey et al., 2013; Noordermeer et al., 2011). It is unclear whether similar structures exist outside the animal kingdom, although there is some evidence that the S. pombe genome forms globular structures that are comparable to TADs and bordered by cohesin binding (Mizuguchi et al., 2014). TADs mainly have a uniform ‘flavor’ of chromatin – i.e. they consist either of only active chromatin or only repressive chromatin (Ciabrelli and Cavalli, 2014). However, it is important to note that while borders of TADs are generally well-defined by genetic features, the chromatin state of individual TADs can vary between cell types (Ciabrelli and Cavalli, 2014).

When it comes to transcription, individual genes tend to follow a radial organization. Genes at the nuclear periphery tend to have lower expression levels, replicate late and are marked by chromatin marks associated with silent chromatin (Kind and van Steensel, 2010). DamID experiments in fruit fly, human and mouse cells found genome regions associating with the nuclear lamina, called lamina-associated domains (LADs) (Guelen et al., 2008; Peric-Hupkes et al., 2010; Pickersgill et al., 2006). Varying in size between 0.1 and 1 Mb, LADs are characterized by low expression levels, repressive chromatin marks (low Pol II levels, low H3K4me3, high H3K9me2) and marked by CTCF binding at borders (Guelen et al., 2008; Pickersgill et al., 2006). This architecture is largely unchanged during development. However, individual genes that reside in LADs dissociate from the lamina when they are
activated during differentiation (Peric-Hupkes et al., 2010). The main factor defining chromatin-lamina interactions appears to be the H3K9me2 mark, since LADs displaced from the nuclear periphery lose H3K9me2 but do not increase in histone acetylation or transcription (Kind et al., 2013). It has therefore been suggested that H3K9me2 is required for peripheral localization (Towbin et al., 2012). In agreement with this hypothesis, a knockout of G9a, the HKMT responsible for setting H3K9me2, reduced LAD-nuclear lamina interactions (Kind et al., 2013). In addition to histone marks, DNA sequence also appears to play a role in defining LADs, as AT-rich sequences show a stronger preference for lamina association (Meuleman et al., 2013).

At the level of individual loci, loop formation helps to bring distal elements together to facilitate e.g. transcriptional initiation. In this way, enhancer elements and gene promoters can be brought in close proximity. On the whole, it appears that individual genes have their defined positions in the nucleus. Single loci can be localized reproducibly at similar distances to the nuclear envelope (Marshall et al., 1996b) and genes can be found within a defined territory in yeast (Berger et al., 2008). Tethering of reporter genes to nuclear membrane proteins has been used to examine the effects of nuclear envelope association on individual genes (Finlan et al., 2008; Kumaran and Spector, 2008; Reddy et al., 2008). Two of the loci studied showed reduced expression levels when tethered, which increased when released (Finlan et al., 2008; Reddy et al., 2008). One study found no change in expression (Kumaran and Spector, 2008), but used a strong inducible promoter for the reporter gene, which might explain why it was able to overcome the effects of peripheral positioning. In general, it appears that movement towards the nuclear periphery is accompanied by reduction in expression, while movement towards the interior allows for activation, but does not necessitate it (Stancheva and Schirmer, 2014).

All levels of nuclear organization are dynamic, although to varying extents. Entire chromosomes can change their position rapidly upon stimulation (Mehta et al., 2010). Genome organization takes some time to re-establish after mitosis, e.g. some LADs start out in the interior after cytokinesis and take up to 15 h to get back into contact with the nuclear periphery (Kind et al., 2013). While chromosome territories in one cell type stay as they were established early in G1, individual loci can move rapidly and dramatically (Müller et al., 2010). This movement requires ATP and happens 4x faster than diffusion (Levi et al., 2005). An inducible transgene moves towards the nuclear interior within 1-2 h after induction, but requires nuclear actin and nuclear myosin for translocation (Chuang et al., 2006).

When considering nuclear organization, it is important to think about the levels described above. The behavior of a single locus can depend on its own local structure, its position within a TAD, its distance to the nuclear periphery and its position within the chromosome territory.
Figure 2: Levels of nuclear organization. The chromosomes occupy territories in the nucleus (center). At the nuclear envelope, repressed chromatin interacts with the nuclear lamina and INM proteins, while active chromatin is located away from the periphery in the interior of the nucleus (A). Nuclear pore complexes form an exception, with nucleoporins interacting with active chromatin (B). In the nucleus, genes come together outside their chromosome territories in transcription factories (C). Within each chromosome, chromatin folds into topologically associated domains (TADs) organizing the genome into smaller units (D).

1.6.3 Components of the nuclear envelope – lamins and INM proteins

The nuclear envelope is a complex environment, with nuclear pores, the nuclear lamina and a multitude of transmembrane proteins forming a landscape of interaction points for chromatin (Cronshaw et al., 2002; Dreger et al., 2001; Rout et al., 2000; Schirmer et al., 2003). All these factors have their own function in organizing and regulating nuclear processes.

Underneath the nuclear envelope, the nuclear lamina forms a filamentous mesh, about 10-20 nm thick, which is pierced by and interacting with INM proteins (Höger et al., 1991). The network is built by polymerization of lamins, which belong to the family of type V intermediate filament proteins (Erber et al., 1999). Compared to the actin and tubulin filaments that form the cytoskeleton, these filaments are relatively elastic. To form a network, lamins first dimerize and then assemble head-to-tail to form protofilaments that can integrate into larger networks (Gruenbaum and Medalia, 2014). Importantly, yeasts seem to lack a nuclear lamina, since lamin genes have only been found in metazoan species. Mammals encode for two types of lamins: type A and type B, which differ in length and expression patterns. Apart from the nuclear lamina, a soluble fraction of both lamin types is also found in
the nucleoplasm (Dechat et al., 2000; Dyer et al., 1997; Moir et al., 1994). Lamin A interacts with the LEM-domain protein LAP2α, working together in cell-cycle regulation and tumor-suppression (Markiewicz et al., 2002). Lamin B is involved in stabilizing DNA damage sites for repair (Moir et al., 2000).

Close to and interacting with the nuclear lamina, INM proteins form another important component of the nuclear envelope. More and more of these proteins have been identified in the last decade (Korfali et al., 2010; Schirmer et al., 2003). Interestingly, their expression patterns differ between tissues (Korfali et al., 2010; 2012; Wilkie et al., 2011), which has been proposed to direct chromatin organization in different cell types (Korfali et al., 2010; Zuleger et al., 2013). Different INM proteins have been shown to interact with different chromatin-binding proteins, allowing for altered genome positioning depending on their expression levels. Some of these partners link to transcriptionally silent chromatin (Brown et al., 2008a; Makatsori et al., 2004; Ye and Worman, 1996), leading to distribution of heterochromatin at the NE (Solovei et al., 2013; Zullo et al., 2012).

One important group of INM proteins carry a LEM domain in their nucleoplasmic tails, such as Emerin, LEM2, MAN1 and LAP2β (Brachner and Foisner, 2011). Many of these proteins interact with the chromatin-associated BAF protein (Barrier-to-autointegration factor, not to be confused with the chromatin remodeling factor BAF!) (Dorner et al., 2007; Furukawa, 1999; Lee et al., 2001; Mansharamani and Wilson, 2005). BAF can bind to both naked DNA and chromatin (Margalit et al., 2007), but specifically interacts with the HKMT G9a (Montes de Oca et al., 2011) and transcription factors (Wang et al., 2002). Furthermore, LAP2β and emerin interact with HDACs (Holaska and Wilson, 2007; Somech et al., 2005), while MAN1 regulates SMAD transcription factors through interaction with Smad2 and Smad3, enhancing Smad transcriptional activity (Bourgeois et al., 2013; Lin et al., 2005).

The nuclear envelope also connects the genetic material to stimuli from the outside world, since proteins in the ONM connect to cytoskeleton and to INM proteins, which in turn connect to chromatin (Tapley and Starr, 2013). The main factor in this process is the LINC complex, which consists of transmembrane proteins linking cytoskeleton and chromatin. KASH-domain proteins in the ONM interact with all types of cytoskeletal networks (actin, intermediate filaments, microtubuli) (Burke and Stewart, 2014). In the perinuclear space, KASH proteins interact with SUN-domain proteins, forming LINC complexes. Inside the nucleus, SUN proteins protrude into the nucleoplasm and interact with lamins and other factors (Crisp et al., 2006; Haque et al., 2006). In this way, LINC is thought to provide a physical connection between forces affecting the cytoskeleton (internal and external) and chromatin, allowing for signaling across the nuclear envelope.

1.6.4 **Components of the nuclear envelope – Nuclear pore complexes**

The nuclear pore complexes (NPCs) are the gatekeepers of the nucleus, controlling the exchange with the cytoplasm. More than 30 proteins, called nucleoporins, come together in
multiple copies to form this large, multi-protein complex, with a molecular weight >40 MDa in yeast and >60 kDa in mammals (Alber et al., 2007; Cronshaw et al., 2002; Rout et al., 2000). The nucleoporins are arranged at an eight-fold symmetry around a central axis, with spikes extending into the cytoplasm and a basket-like structure reaching into the nucleoplasm. These formations help to keep the pore channel entry and exit points free from chromatin and cytoplasmic material, ensuring efficient transport through the pore. On the inside of the NPC are core nucleoporins, which are well conserved throughout evolution, building up the structure and interacting with transported material. Passage through the NPC can occur passively for material up to a certain size limit, but requires assistance of importin and exportin complexes for larger material. In addition to the core nucleoporins, there are also NPC components that shuttle from the pore to the nuclear interior (Dilworth et al., 2001; Rabut et al., 2004).

The first point of interaction with chromatin occurs at the nuclear basket, which is comprised of the coiled-coil Tpr protein (Mlp1/Mlp2 in yeast), and Nup50 (Guan et al., 2000; Jarnik and Aebl, 1991). In budding yeast, Mlp1 and Mlp2 are involved in telomere tethering and required for silencing of subtelomeric genes (Feuerbach et al., 2002; Galy et al., 2000; Mailet et al., 2001; Scherthan et al., 2000). In general, NPCs are in contact with less dense chromatin compared to the rest of the nuclear envelope (Schermelleh et al., 2008). Accordingly, nucleoporins were found to interact mainly with actively transcribed chromatin (Capelson et al., 2010; Kalverda et al., 2010; Vaquerizas et al., 2010). In yeast, Mlp1 interacts with SAGA and the mRNA export complex TREX2 (Cabal et al., 2006; Luthra et al., 2007; Pascual-Garcia et al., 2008; Rodriguez-Navarro et al., 2004), both found at actively transcribed genes. It is however still unclear if the interactions between nucleoporins and active genes are occurring at the NPC or in the nuclear interior, with nucleoporins acting as recruitment factors (Kalverda et al., 2010).

Decades ago, it had been proposed that NPCs work in gene gating, by interacting with transcriptionally active genes to ensure rapid mRNA export and target mRNA to specific parts of the cell (Blobel, 1985). Since then, studies from various species have come to different conclusions regarding this hypothesis. In budding yeast, many inducible genes move from the nuclear interior to the nuclear periphery upon activation (Brickner et al., 2007; Taddei et al., 2006). This activation involves several nucleoporins and the SAGA complex (Cabal et al., 2006; Dieppois et al., 2006; Luthra et al., 2007; Taddei et al., 2006). In addition, some genes carry gene recruitment sequences (GRS) in their promoters, which are required for targeting them to nuclear pores (Ahmed et al., 2010; Brickner et al., 2012; Brickner and Walter, 2004). In contrast, nucleoporins have also been shown to be involved in silencing, e.g. Nup170 is involved in repression of subtelomeric genes (Van de Vosse et al., 2013). In C. elegans, the opposite shift was observed, with developmentally induced genes moving from periphery to the interior upon activation (Meister et al., 2010b). However, stress-induced genes in worms were found to move laterally along the nuclear perimeter towards NPC when activated (Rohner et al., 2013). Similar results were obtained in D. melanogaster, where nucleoporins associate with actively transcribed regions, e.g. in male X-chromosome
dosage compensation. Nucleoporins were found to interact with developmental and stress-induced genes (Capelson et al., 2010; Kalverda et al., 2010), but this interaction is thought to take place in the interior (Kalverda et al., 2010). In mammals, developmentally induced genes were again observed to move from the periphery to the nucleoplasm (Ragoczy et al., 2006; Takizawa et al., 2008; Williams et al., 2006). Here, nucleoporins are involved in transcriptional regulation in the nuclear interior, e.g. Nup98 shuttles back and forth between the NPC and the nucleoplasm (Chatel et al., 2012; Griffis et al., 2002; Liang et al., 2013; Singer et al., 2012) and works in transcriptional memory (Light et al., 2013). In summary, the gene gating hypothesis holds true in yeasts, with inducible genes moving to the NPC upon activation. Although nucleoporins are also involved in transcriptional regulation in metazoan systems, they appear to fulfill this function away from the NPC, in the nuclear interior.

1.6.5 Spatial organization of nuclear processes

Not only the genome, but also the processes associated with it show non-random positioning patterns. Many of these functions occur in foci, which allows for an efficient and controlled use of enzymes and co-factors.

A prime example for these foci are transcription factories, which were first observed through incorporation of halogenated nucleotides (Jackson et al., 1993; Wansink et al., 1993). All three RNA polymerases form transcription factories, with a small number of Pol I foci at the nucleolus, and many more Pol II and Pol III foci in the nucleoplasm (Iborra et al., 1996; Jackson et al., 1993; Wansink et al., 1993). The number of transcription factories varies between species and cell types, but is roughly proportional to size of the nucleus (Faro-Trindade and Cook, 2006; Osborne et al., 2004). Transcription factories are defined as nuclear sites with at least two distinct transcription units coming together and being transcribed (Cook, 2010; Osborne et al., 2004), although there are on average 8-10 transcription units per factory (Dekker et al., 2002; Jackson et al., 1993; Kimura et al., 1999; Osborne et al., 2004; Schoenfelder et al., 2010). Each factory contains transcription factors, RNA polymerase complexes, proteins of the splicing machinery, and histone modifiers (Melnik et al., 2011), allowing all steps of the process to occur in one place.

Transcription factories also provide spaces for genes from different chromosomes to intermingle (Branco and Pombo, 2006; Mehta et al., 2010), especially since active genes tend to lie outside their chromosome territory (Chambeyron and Bickmore, 2004; Gilbert et al., 2004; Mahy et al., 2002). Additionally, induced genes tend to move from the inside of their territory to the outside, toward transcription factories (Osborne et al., 2007; Volpi et al., 2000). This process has a certain degree of specificity, with gene pairs coming together in the same transcription factory. This is one of the reasons why translocations of specific gene pairs are common in diseases: although they are on different chromosomes they come together to be transcribed in transcription factories, where translocations can occur (Branco and Pombo, 2006; Küppers and Dalla-Favera, 2001; Osborne et al., 2007; Parada et al., 2004;
Interestingly, co-expressed genes have been observed to share transcription factories, possibly due to limited transcription factor availability in the nucleus (Eskiw and Fraser, 2011; Fullwood et al., 2009; Kang et al., 2011; Osborne et al., 2004; Papantonis et al., 2010; Splinter et al., 2011).

Replicating the human genome is an enormous task, with three billion base pairs that have to be faithfully copied at a speed of 33 nucleotides per second (Fangman and Brewer, 1992). Since starting at the chromosome ends and working inward would take a very long time, replication instead starts in parallel at replication origins, of which there are about 30,000 in the human genome (Cayrou et al., 2011). At these sites, replication starts by binding of the origin recognition complex (ORC), which is able to recruit other components to assemble a pre-replication complex (pre-RC). Initiation occurs through two cell cycle-regulated kinases, DDK and CDK, which are activated at the G1-to-S transition. By phosphorylating key components, these kinases allow for assembly of the complete replisome and initiation of DNA synthesis (Aparicio, 2013).

Replication origins are defined by sequence motifs (Stinchcomb et al., 1979), but their efficiency and timing vary depending on chromatin context: depending on its location on the chromosome, an origin might start replication earlier or later. In yeast, centromeric sequences replicate early, while telomeric and subtelomeric sequences replicate later (Ferguson and Fangman, 1992). The positioning of an origin within the nucleus also affects its replication timing. Origins tethered to the nuclear envelope through Taz1 in yeast replicate later than those in the nuclear interior (Tazumi et al., 2012). It has been proposed that this difference in timing is due to limiting of the replication complex component Cdc45 to foci in the nucleus. These structures are held together by dimerization of the Fox transcription factors Fkh1/2 and form factory-like structures (Ostrow et al., 2014). Origins close to replication foci experience earlier activation of replication origins, while loci tethered to the nuclear envelope are sequestered away from these foci and fire later (Aparicio, 2013). Interestingly, replication timing is organized by TADs, with some domains replicating earlier and some later. The differences in replication timing tend to coincide with TAD boundaries, effectively forming ‘replication domains’ (Pope et al., 2014).

Repair of double strand breaks (DSBs) is immensely important to avoid translocations and chromosome loss, as well as ensure faithful cell division. Two pathways used to mend DSBs: Homologous recombination repair (HRR) uses the homologous chromosome to repair the break. Non-homologous end-joining (NHEJ) repairs the break by resecting and ligating the two ends. While HRR results in faithful repair of the break site, NHEJ can lead to mutations since chromosomal material is lost in the process. Both processes occur in repair foci, which form around a DNA damage site where DNA-repair factors accumulate. These foci can extend up to 1 Mb from the site of damage (Lukas et al., 2005), which has been proposed to stabilize the DNA break in space keeping the ends from drifting apart. In yeast, repair centers can recruit and mend multiple DSBs, even when they occur away from the repair center (Lisby and Rothstein, 2004; Lisby et al., 2003a; 2003b). An increase in the number of lesions
will therefore not increase the number of repair foci. In some cases, DNA repair might require tethering to the NE (Mekhail et al., 2008; Oza et al., 2009). In mammals, broken ends are largely immobile (Soutoglou et al., 2007), therefore repair factors need to be recruited from the nucleoplasmic pool to come together at the repair site. Close proximity between homologs is important during HRR, since genome regions with overlapping territories recombine more efficiently compared to regions more distant (Agmon et al., 2013).

1.6.6 Nuclear organization in development and disease

With spatial positioning affecting all levels of genome organization and all major nuclear processes, it is not surprising that nuclear organization is important during development. In mammalian cells, lamin expression patterns change during differentiation. While at least one B-type lamin is expressed in all nucleated cells, A-type lamin expression can vary considerably. A-type lamins are not expressed in early embryonic stages and are lacking in some tissues even after birth, e.g. hematopoietic cells (Röber et al., 1990). It has been proposed that Lamin B receptor (LBR) can substitute some of their functions in these tissues (Solovei et al., 2013). Nuclear organization varies between cell types, presumably reflecting the changes in gene expression during differentiation. Both chromosome positions and chromatin domains such as LADs are altered, with developmentally induced genes moving away from the nuclear envelope and stem cell genes moving toward it (Peric-Hupkes et al., 2010). An interesting example of tissue-specific re-organization are rod cells in the eye, where the typical radial organization is reversed, with active chromatin at the nuclear periphery and heterochromatin in the center of the nucleus (Solovei et al., 2009). It has been proposed that this unusual structure allows the nuclei to act as lenses during light-sensing. This type of organization is thought to be achieved by different expression patterns of LBR and lamins A/C (Solovei et al., 2013).

Alterations in nuclear architecture are both hallmark and cause of various diseases. Mutations in lamins cause a group of syndromes called laminopathies, which range from muscular dystrophy to progeria (Worman et al., 2010). Similarly, mutations in INM proteins have been linked to various disorders affecting muscle, brain, bone and immune cells (Stancheva and Schirmer, 2014). One example is Emery-Dreifuss-Muscular-Dystrophy (EDMD) (Brown et al., 2008b), most often caused by mutations in EMD, encoding emerin, or LMNA, encoding lamin A and C. On the cellular level, this results in dense chromatin breaking away from the nuclear envelope (Fidziańska, 1996). Patients suffer from muscle weakness in arms and legs, later on also affecting cardiac muscles and therefore impairing mobility and causing heart problems.

A second example of laminopathies is Hutchinson-Gilford-Progeria-Syndrome (shortened to progeria) (Korf, 2008). Children born with the disease fail to thrive, show hardened skin, hair loss, degenerating eye sight and develop arteriosclerosis over time, few surviving past the age of 13. Progeria is caused by a mutation in LMNA (Eriksson et al., 2003) leading to a
precursor form of lamin A, progerin, which cannot be released from the nuclear membrane and accumulates there instead of moving into the nucleoplasm. Morphologically, this mutation causes nuclei to appear deformed, with bulges instead of a smooth, circular shape. Additionally, fibroblasts from progeria patients show reduced levels of H3K9me2 and H3K27me3, which characterize facultative heterochromatin, and increased levels of H4K20me3, typical for constitutive heterochromatin (Shumaker et al., 2006).

Mutations in INM proteins, NPC components and lamins have been linked to different cancers, although they are likely not the direct cause for tumorigenesis in these cases (Stancheva and Schirmer, 2014). Expression levels of INM proteins and lamins are altered in certain cancer types, with reduced levels of emerin observed in ovarian cancer (Capo-chichi et al., 2011) and increased expression of lamin A/C found in ovarian, prostate and colorectal cancers (Skvortsov et al., 2011; Willis et al., 2008). Again, this could be a consequence of gene expression changes of other factors. However, chromosome positions can be altered in tumor cells (Cremer et al., 2003) and DNA methylation levels are altered in LADs in colorectal cancer (Berman et al., 2012).

Taken together, these observations point to a strong interplay between genome organization and function, both during differentiation and in disease states.
2 METHODS

2.1 FISSION YEAST AS A MODEL SYSTEM FOR CHROMATIN BIOLOGY AND NUCLEAR ORGANIZATION

The fission yeast *Schizosaccharomyces pombe* was first found in millet beer brewed in East Africa and later isolated from French wine (Jeffares et al., 2015). The unicellular eukaryote forms rod-shaped cells 3-4 µm in diameter and 7-14 µm in length (Mitchison and Nurse, 1985). In contrast to *Saccharomyces cerevisiae*, which forms buds during cell division, *S. pombe* divides symmetrically into two daughter cells of equal size. While the diameter stays constant, the cell elongates, forms a furrow along the middle and finally separates into two cells. Of the 2.5 hours it takes to progress through the cell cycle, fission yeast spends about 70% in G2 phase (Fig. 3).

The *S. pombe* genome sequence was published in 2002 and confirmed a genome size of 14.1 Mb and fewer than 5,000 protein-coding genes (Wood et al., 2002). Fission yeast has been used as a model organism for decades, since it is easy to cross, manipulate and grow, even in large scale. In over 50 years of fission yeast research, a large portfolio of techniques has been developed, including rapid forward and reverse genetics, biochemistry, proteomics and genome-wide analysis. Sir Paul Nurse used fission yeast to study cell cycle regulation, which proved to be highly conserved in eukaryotes and earned him and others the Nobel Prize for Physiology or Medicine in 2001 (Nurse, 2002).

Fission yeast is especially suitable for chromatin research, since its chromatin structure and centromere organization is similar to that of more complex eukaryotes. The *S. pombe* genome also encodes for single copies of the enzymes involved in the RNAi machinery, which are missing in budding yeast.

![Figure 3](image-url): The vegetative life cycle of *Schizosaccharomyces pombe*. 
2.2 MAPPING THE CHROMATIN LANDSCAPE: CHIP, DAMID, AND MNASE

Much can be learned by studying the interactions between DNA and the myriad of proteins it comes into contact with in the cell. These interaction patterns can help to understand how DNA related processes work, to identify DNA sequence features and to study chromatin structure.

Chromatin-immunoprecipitation (ChIP) commonly requires fixation as a first step, which covalently crosslinks DNA and protein in the cell (Durand-Dubief and Ekwall, 2009). Next, the cell lysate is sonicated to break up the DNA into smaller pieces. Specific antibodies are then used to pull down the protein of interest and the DNA sequences it was interacting with at the time of fixation. After reversal of crosslinking and DNA purification, these sequences can identified by sequencing or quantified by qPCR with site-specific primers. ChIP can yield insights into where a protein binds in the genome at a given time, but requires optimization for each antibody used and careful controls to make quantitative comparisons possible.

DamID is a method similar to ChIP, but relies on marking protein targets in vivo rather than pulling them out with antibodies (van Steensel and Henikoff, 2000). To achieve this, a fusion protein of the DNA adenine methyltransferase Dam and the protein of interest is constructed and expressed in a cell line or strain. Any DNA sequence that comes into contact with this fusion protein will be methylated on the N6 position of adenine in GATC sequences. The genomic DNA extracted from these cells carries the methylated sites, which can be enriched through digestion with methylation specific restriction enzymes, adapter ligation and PCR.

ChIP and DamID have their advantages and applications. DamID is especially useful for proteins that interact only transiently with the genome. The study of histone marks is only possible through ChIP with antibodies that recognize these modifications. In contrast to ChIP, DamID does not require fixation and is therefore less prone to artifacts. Antibody specificity is also not an issue. However, DamID maps always present the average of at least a few hours of interactions between protein and DNA, while ChIP can provide something more like a snapshot of interactions. Furthermore, DamID can only be used on cell lines that can be easily transfected and cultured.

The mapping of nucleosome positioning presents a particular challenge. While ChIP with histone-specific antibodies has been used, the resulting maps tend to lack resolution (Bernstein et al., 2004). To achieve higher resolutions, methods are used which rely on differences in how accessible free DNA and nucleosomal DNA are for nucleases. Most common is the use of the endonucleases DNaseI and Micrococcal Nuclease (MNase), which will cut DNA between nucleosomes (Lantermann et al., 2009). As in ChIP, DNA-nucleosome interactions are preserved by covalent crosslinking. The following nuclease digestion step then has to be timed carefully to obtain chromatin fragments with a few nucleosomes, but avoid degradation. This step is followed by reverse-crosslinking and purification of mono-nucleosomal DNA fragments through gel electrophoresis. The isolated DNA can then be analyzed through qPCR, microarrays or sequencing and used to produce nucleosome occupancy maps.
2.3 GENOME-WIDE ANALYSES: MICROARRAYS, HIGH THROUGHPUT SEQUENCING, AND BIOINFORMATICS

To understand the effect of a deletion or the DNA binding pattern of a protein, it is important to look beyond the single locus at a genome-wide level. With more and more genomes sequenced and annotated, maps from ChIP or DamID can be put into context, so patterns and correlations become apparent.

DNA microarrays provide a way to quickly access expression levels or protein interactions over a multitude of genome sites (Yazaki et al., 2007). Based on the known sequence of the genome, DNA probes are generated that map to loci of interest (e.g. exon arrays) or to the entire genome (in case of high resolution tiling arrays). These probes are then spotted onto a solid support and form a microarray-chip, which is commercially available from companies like Affymetrix or Agilent. The material to be analyzed, be it a cDNA or ChIP sample, is labeled with a fluorescent dye and then hybridized to the microarray. Depending on the abundance of a sequence within the sample, probe spots will now appear brighter or dimmer. The fluorescent signals are then scanned and converted into a numeric value for each probe. In this way, microarrays can provide a fast and relatively inexpensive way to assess transcription levels or protein binding, especially when probes are limited to a number of known targets. However, specificity is limited by the fidelity of the hybridization process: point mutations and sequences with high similarity are difficult to distinguish in this manner.

The development of next-generation sequencing methods has provided strong competition to microarray technologies over the last decade, with sequencing now well underway to outclass arrays (Liu et al., 2012). In contrast to traditional sequencing, these approaches rely on thousands or even millions of sequencing reactions occurring in parallel. Similar to microarray technology, DNA samples are immobilized on solid supports, either slides or beads. As the samples are highly diluted in this step, each template forms the seed for local clonal ‘colonies’, which are formed through PCR amplification. In the sequencing reaction, the support is flooded with one type of nucleotide after another (454 pyrosequencing) or a pool of nucleotides individually labeled with a different fluorescent dye (Solexa/Illumina sequencing). After the correct base has been added, unincorporated nucleotides are washed away and a camera detects the signal for the incorporated base. This cycle is repeated until the possible read length is reached (commonly between 50 and 1000bp). The sequentially acquired images are then processed and reveal the sequence for each ‘colony’ on the support. Over the last years, speed of next-generation sequencing has increased dramatically, while costs continue to decrease. These technologies can be used for many applications, most commonly transcriptome analysis, ChIP-seq and whole genome sequencing.

Independently of whether microarrays or sequencing are used, thorough bioinformatic analysis is required to extract, evaluate and interpret the large datasets produced by employing these methods. Sequencing data in particular contains millions or billions of reads, which need to be pre-processed, aligned with a reference genome and converted to quantitative data. In the case of microarray data, probe scores need to be normalized between...
arrays to account for differences in hybridization efficiency and image acquisition. Sophisticated programs and pipelines have been developed over the last decade to perform these analyses, most prominently through the Bioconductor project (Gentleman et al., 2004) for the R statistical analysis language. The platform provides tools for microarray and sequencing analysis, as well as the possibility to generate graphical representations of complex data.

2.4 VISUALIZING THE NUCLEAR INTERIOR: IMMUNOFLUORESCENCE, FISH, AND LIVE CELL IMAGING

While binding maps of chromatin proteins and nucleosome occupancy maps can provide a wealth of information, they lack context on location and timescale of nuclear processes. Microscopy-based methods can help make these connections (Meister et al., 2010a). Immunofluorescence and Fluorescence In-Situ Hybridization (FISH) are used to visualize proteins and DNA, respectively. By using an antibody recognizing a chromatin protein or a probe complementary to a chromosome locus, these targets can be made visible using fluorescent dyes. The genome and the nuclear envelope can be stained with other dyes and antibodies, allowing to determine the place of the protein or locus of interest within nuclear structures.

Although FISH and immunofluorescence microscopy can provide useful insights, it is important to note that these methods require a fixation step to preserve 3D-positioning throughout permeabilization, antibody incubation/hybridization, washes and imaging. Since the nucleus is relatively fragile, this can lead to artifacts. Expression of fusion constructs between nuclear proteins and fluorescent proteins like GFP allows observing nuclear morphology in vivo and over extended time periods. Even chromosomal loci can be visualized in this manner through insertion of lac-operon repeats into the DNA sequence and expression of lac-inhibitor fused to a fluorescent protein. However, fusing a fluorescent tag to a protein can alter their function and localization, photobleaching can impair study over longer timeframes, and care must be taken to prevent phototoxicity to cells due to prolonged exposure with strong laser light.
3  RESULTS AND DISCUSSION

3.1  PAPER I - THE INNER NUCLEAR MEMBRANE PROTEINS MAN1 AND IMA1 LINK TO TWO DIFFERENT TYPES OF CHROMATIN AT THE NUCLEAR PERIPHERY IN S. POMBE

As gene regulation occurs in three dimensions, there is a need for a closer look at chromosome organization. Specifically in fission yeast, only the rough contours of nuclear architecture were known previously. Both centromeres and telomeres bind to the nuclear envelope, with centromeres clustering opposite the SPB (Funabiki et al., 1993; Uzawa and Yanagida, 1992) and telomeres localizing away from the centromeres. Chromosome territories are thought to exist, as recombination frequencies were found to be higher within a chromosome than between chromosomes (Molnar and Kleckner, 2008). Furthermore, S. pombe cells lack a nuclear lamina, posing the question whether genome organization can be compared to metazoans.

Starting from these observations, we set out to map which regions of the fission yeast genome come in contact with the nuclear envelope in logarithmically growing cells. The DamID methodology had been successfully applied in fruit fly, mouse and human cells towards the same end (Guelen et al., 2008; Peric-Hupkes et al., 2010; Pickersgill et al., 2006). Therefore we constructed fusion proteins of the Dam-methyltransferase and two INM proteins, reasoning that we would be able to obtain a more complete picture of the nuclear landscape in this manner. The first, Ima1, had been implicated in anchoring centromeres at the nuclear envelope (King et al., 2008) and is the homolog of the human Samp1 (Gudise et al., 2011). The second, Man1, contains a helix-extension-helix domain and is related to LEM-domain proteins (Brachner and Foisner, 2011). We carried out DamID experiments with both constructs and used high-resolution tiling microarrays to map the enriched fragments to the S. pombe genome. Using a Hidden Markov Model and bioinformatics analysis, we were able to get a detailed view into chromosome-nuclear envelope interactions.

We observed that about a third of the fission yeast genome interacts with the nuclear envelope through Ima1 and Man1. Similar proportions have been observed in other eukaryotic systems (Guelen et al., 2008; Ikegami et al., 2010; Pickersgill et al., 2006). Importantly, we observed an inverse relationship between gene expression and INM protein interaction: lowly expressed genes show a higher enrichment of Man1 and Ima1 compared to highly expressed genes. Despite absence of a nuclear lamina, fission yeast shows a similar radial organization of expression states as have been shown in other systems (Guelen et al., 2008; Ikegami et al., 2010; Pickersgill et al., 2006). We also found other hallmarks of repressed chromatin at Man1 and Ima1 targets, such as low occupancy of Pol II and increased levels of H2A.Z, which is enriched at promoters of repressed genes in fission yeast (Buchanan et al., 2009).

Aside from these common interaction patterns, the two INM proteins also bind separate targets. We discovered that Ima1 interacts with regions that are also bound by Dcr1 and
Rdp1, two components of the RNAi machinery. Man1 does not show this preference, but instead associates with loci that interact with the heterochromatin protein Swi6. This overlap is especially visible over the subtelomeric chromatin domains. These findings are somewhat surprising, since one might assume the nuclear envelope to be uniform, with INM proteins diffusing freely and interacting with chromatin in a similar manner. Instead, our results point to microenvironments in the nuclear envelope that vary in protein composition and chromosome interactions.

In summary, this paper provides a foundation for studying nuclear organization in *S. pombe*, allowing to put epigenetic and genomic features into a 3D context.

### 3.2 PAPER II - THE FUN30 CHROMATIN REMODELER FFT3 CONTROLS NUCLEAR ORGANIZATION AND CHROMATIN STRUCTURE OF INSULATORS AND SUBTELOMERES IN FISSION YEAST

Following up on the results in paper I, this study further explores the interplay between nuclear positioning and transcriptional regulation. We had observed that subtelomeric chromatin domains show a strong enrichment of Man1. These domains contain genes with low expression levels and repressive chromatin marks (Buchanan et al., 2009; Zofall et al., 2009). Their chromatin state is regulated by the chromatin remodeler Fft3, which binds to subtelomeric borders (Strålfors et al., 2011). In its absence, subtelomeric genes are upregulated and active chromatin marks increase over the entire domain.

Based on these observations, we set out to determine if these changes in transcription and chromatin state are accompanied by changes in interactions with the nuclear envelope through Man1. Indeed, cells lacking Fft3 showed decreased interactions with the INM protein in DamID experiments. We were able to verify these changes in peripheral positioning by live cell microscopy, showing that Fft3 and the telomere anchoring protein Bqt4 collaborate to anchor subtelomeric chromatin to the nuclear envelope.

Since we had observed pronounced changes in chromatin, we were curious if the chromatin remodeling activity of Fft3 is required for its function at subtelomeric borders. We mapped nucleosome occupancy using MNase digestion, followed by next generation sequencing and found that nucleosome occupancy is reduced at several positions in the subtelomeric borders. We also created a catalytically inactive version of Fft3 through a point mutation in its ATPase domain. While the mutated version of Fft3 binds to the same targets as native Fft3, it mimics the phenotype of the *fft3Δ* deletion strain: subtelomeric gene expression and active chromatin marks are increased, while interaction with Man1 is reduced. Therefore, we conclude that chromatin remodeling activity of Fft3 is required to maintain proper nucleosome positioning at subtelomeric borders and to preserve transcriptional repression throughout the subtelomeric domain.
Fft3 binds to various other loci throughout the genome, not only at the subtelomeric borders. We were especially interested in RNA Polymerase III transcribed loci, since Fft3 is strongly enriched over tRNA and 5S rRNA genes. This is particularly interesting, since tRNA genes function as insulators at *S. pombe* centromeres (Scott et al., 2006). A yeast-two-hybrid screen showed a physical interaction between the remodeler and Sfc4, a subunit of the transcription factor TFIIIC. We were able to verify this interaction by co-immunoprecipitation of Fft3 and Sfc4. Importantly, we also observed a reduction in nucleosome occupancy at tRNA genes in the *fft3Δ* deletion strain, pointing to a role for Fft3 in chromatin structure at these loci.

Taken together, these results describe Fft3 as a versatile regulator of chromatin structure at insulating elements in the *S. pombe* genome.

### 3.3 Paper III - Regulating Retrotransposons via Alternative Transcription Start Sites

In this study, we wanted to further expand our understanding of Fun30 chromatin remodeler function in fission yeast. While paper II examined the functions of Fft3, the effects of its homolog, Fft2, were still to be determined. Since the budding yeast genome encodes only one Fun30 remodeler, it is interesting to investigate how the three homologs in fission yeast are distinguished from each other in target preference and function.

Starting out with an examination of a *fft2Δ* deletion strain, it became clear that Fft2 plays a role in retrotransposon regulation: transcription of all 13 copies of the LTR retrotransposon *Tf2* was increased. When we confirmed this increase by Northern Blot, we were surprised to find this upregulation not only in *fft2Δ*, but also in *fft3Δ* and, most strongly, in the *fft2Δ fft3Δ* double deletion strain. We also observed a size shift in the length of the *Tf2* transcript, with slightly longer transcripts in the deletion mutants compared to the wild-type strain. By 5’ RACE (Rapid Amplification of cDNA Ends), we mapped the TSS of this longer transcript to the LTR upstream of the *Tf2* coding sequence, while the shorter, wild-type transcript originates further downstream, outside the LTR. We also confirmed these results with CAGE, showing that the shorter *Tf2* transcripts are capped and not digestion products.

To understand the significance of these positions, one needs to turn to the transposition cycle of LTR retrotransposons in fission yeast, as described by (Levin, 1995). To allow for synthesis of the second strand, part of the LTR contains a self-primer and a corresponding primer binding site (PBS). After transcription of the *Tf2* mRNA, the LTR can fold back on itself, bringing the self-primer in contact with the PBS. This section now forms a short stretch of double stranded RNA that can be used to prime second strand synthesis by reverse transcriptase. The resulting second strand can then be used to start reverse transcription from the 3’ end of the *Tf2* mRNA. Therefore, only long transcripts extending to the self-primer sequence are able to form the cDNA needed for transposition. Coming back to the TSS we found through 5’ RACE, this means that the longer transcript produced in the *fft2Δ fft3Δ*
strain is capable of second strand synthesis. The shorter wild-type transcript, however, does not contain the LTR sequences required to complete this cycle and is incapable of transposition.

Our next question was then how Fft2 and Fft3 influence the choice of TSS. To this end, we first mapped binding of the two remodelers using ChIP-chip. Both enzymes target the LTR regions on either side of the *Tf2* coding sequence and can substitute for each other when one is missing. We then mapped nucleosome positions over the entire region by MNase-Seq to see if the two enzymes affect local chromatin structure. In wild-type cells, we found a well-positioned nucleosome close to the self-primer sequence. In the *fft2Δ fft3Δ* double deletion strain, occupancy of this nucleosome is significantly reduced, allowing for initiation of transcription to occur from within the LTR and therefore producing the longer, transposition-capable mRNA.

This provides interesting possibilities for retrotransposon regulation. Instead of having to silence the locus, low levels of transcription can occur in wild-type cells without resulting in transposition events. Why is this beneficial for the cell? The fission yeast genome contains more than 200 solo LTRs, remnants of previous transposition events inserted between genes. We found that nucleosome occupancy over these solo LTRs is reduced in the *fft2Δ fft3Δ* strain as well. Furthermore, genes downstream from solo LTRs showed an increase in transcription when Fft2 and Fft3 were missing, while genes without this feature were unchanged. We propose that limited transposition allows integration of these elements, which leads to potentially advantageous changes in gene expression patterns.

Adaptability of transcript levels is especially crucial when changes in environmental conditions occur. We therefore wondered if similar effects on TSS choice and chromatin structure occur when cells are stressed. We exposed *S. pombe* cells to heat and osmotic stress, respectively, and found an increase in transcription of the longer mRNA as well as a reduction in nucleosome occupancy. At the same time, expression of Fft2 and Fft3 is reduced in these stress conditions (Marguerat et al., 2012). By down-regulating their expression, the choice of *Tf2* TSS can be influenced so that the longer transcript is produced and transposition events occur. This provides an elegant mechanism for the cell to quickly adapt to changes in its environment, since transcription at the *Tf2* loci already occurs from the TSS downstream. A contribution of retrotransposons to genome plasticity had been previously suggested by Nobel laureate Barbara McClintock (McClintock, 1984).

As in paper II, we wanted to find out whether remodeling activity of the two Fun30 enzymes is required for their function in retrotransposon regulation. We constructed a catalytically inactive version of Fft2, which binds to the *Tf2* LTR like native Fft2. However, the point mutations in the ATPase domains of Fft2 and Fft3 result in increased transcription from the alternative TSS and reduced nucleosome occupancy over the LTR, comparable to the single deletion strains. Interestingly, a strain carrying both catalytically inactive variants of the enzymes shows only a moderate increase in transcription from the upstream TSS compared to the *fft2Δ fft3Δ* double deletion strain. As Fft2 and Fft3 still bind to their
target site, we assume that their presence blocks access to the TSS for the transcription machinery to a certain extent.

Finally, we were able to show that deletion of Fft2 and Fft3 also affects subnuclear positioning of the Tj2 retrotransposons. In wild-type cells, the Tj2s cluster into so-called Tj-bodies (Cam et al., 2008), similar to the fruit fly gypsy transposons clustering to facilitate chromatin organization (Gerasimova et al., 2000). Using a FISH probe against the Tj2 sequence, we were able to visualize these foci. In the fj2Δ fj3Δ double deletion strain, the Tj-bodies are broken up into multiple smaller spots, suggesting that changes in chromatin structure and transcription also affect Tj2 clustering.

With this paper, we provide insight into a novel mechanism of retrotransposon regulation. By allowing low level production of a transcript unable to complete transposition, local chromatin structure at the Tj2 loci is kept open and poised to be upregulated quickly. When cells are stressed, the Fun30 remodeling enzymes are down-regulated, resulting in de-stabilization of the nucleosome over the Tj2 LTR and transcription initiation from this alternative TSS. The resulting long mRNA can then complete the transposition cycle and, by inserting into gene promoters, possibly alter gene expression patterns to cope with the changing environment.
4 CONCLUSIONS AND OUTLOOK

This thesis explores the connections between transcriptional regulation, chromatin structure and nuclear organization. The results presented in paper I suggest that radial organization based on transcription level might be a feature conserved throughout eukaryotic evolution. Even relatively simple, unicellular organisms like *S. pombe* show an enrichment of lowly expressed genes at the nuclear periphery, even though they lack the nuclear lamina involved in this organization in metazoans.

Another interesting result is the observation that two INM proteins can interact with different targets in the fission yeast genome. This suggests that the nuclear envelope is not uniform, but rather consists of micro-environments with different protein compositions. This separation could be either cause or consequence of the underlying genome organization. Many other questions remain open, such as whether the *S. pombe* genome encodes for a yet unknown lamin-like protein, as has been found in the unicellular *Dictyostelium discoideum* (Krüger et al., 2012). Furthermore, the interaction partners of Man1 and Ima1 are still unknown, as no ortholog of BAF has been found in yeasts. Discovery of these interaction partners could possibly explain differences in binding patterns between Ima1 and Man1.

In paper II, we investigated the function of the Fft3 chromatin remodeler in the *S. pombe* genome. One interesting observation is the regulation of the subtelomeric chromatin domains through Fft3 remodeling activity at their borders. The fact that altering the local chromatin structure in regions a few kbp across can affect a domain ~100 kb in size speaks for the importance of insulators in chromatin regulation. Since the subtelomeres form a TAD-like domain (Mizuguchi et al., 2014), it would be interesting to perform a HiC-study in an fft3∆ strain. Fft3 could be a factor in maintaining boundaries between these globular domains.

Continuing in our study of Fun30 chromatin remodelers, we investigated the role of Fft2 and Fft3 in retrotransposon regulation. With both enzymes stabilizing a nucleosome over the *Tf2* LTR and shifting the TSS, we present an interesting way to harness retrotransposon activity for the benefit of the host organism. Production of an mRNA incapable of transposition provides a secure mechanism to keep chromatin open and allow for short bursts of transposition-capable transcripts in times of stress. The way in which the remodeling enzymes stabilize the LTR nucleosome is still unclear, though *in vitro* studies would be helpful to shed light on this issue. Through the use of catalytically inactive variants of chromatin remodelers, we were able to zoom in further on the mechanism underlying transcriptional regulation by Fft3. It is important to establish whether the actions of a remodeling enzyme are due to catalytic activity or merely caused indirectly by recruiting another factor. In both paper I and paper II, we were able to show that ATPase activity is required for Fun30-remodeler function at subtelomeric borders and retrotransposons.

Lastly, this thesis highlights the importance of high-resolution, genome-wide methods in chromatin biology. Through the use of tiling array and next-generation sequencing data, we were able to observe both genome-wide connections and small, localized changes.
5 ACKNOWLEDGEMENTS

This thesis would not have been possible without the help of many people. First and foremost, sincere thanks to my supervisor, Karl, for encouraging me to stay curious and supporting me along the way, even when yet another experiment failed. Your passion for science is infectious and I have thoroughly enjoyed our discussions over the years.

Many thanks also to my co-supervisor, Bas, for the help with setting up DamID in yeast and introducing me to the joy of coding in R. Your comments and advice on my project have been very much appreciated.

I’m grateful to KEK group members past and present, who helped me when lab results frustrated me yet again, joined in interesting discussions (both deeply scientific and not at all scientific), and lightened the week with Friday fikas. Jenna, it has been a joy playing tag team with you in the lab, geeking out over the Dresden files novels, and exploring Stockholm, Helsinki and Japan together. Thanks for always being there for a chat. Michelle, you taught me to knit, remember? Thank goodness you did, since I wouldn’t have made it through my PhD without wool. In all earnestness, thank you for the delicious fika, for fun adventures in London and for hugs when I needed them. Annelie, thank you for helping me through endless rounds of troubleshooting, your optimism and perseverance are inspiring. It’s been fun exploring Boston and New York together – yay for humpback whales and Target! Olga, you are so kind and helpful, always there for a chat and keeping lab morale up with delicious chocolate and cookies. Thank you for all the good advice and lab wisdom you shared. Ulrika, it has been great finishing our PhDs in parallel and having you there to commiserate and plan together. I’m inspired by the energy and thoughtfulness you put into everything you do. Alex, it has been wonderful meeting someone who shares my love for gardening, På spåret and sarcasm. Thanks for the fun lunch conversations and the thorough proofreading of my thesis! Andreas, I have enjoyed our Swedish/German conversations over the years a lot. Keep your curiosity and positive attitude! Galina, your stories have been the highlight of Friday fika since you joined the lab. Thank you for all your help in the lab and the great sewing inspiration. Birgitta, thanks for taking care of all the little things we need in the lab, and for reminding me that it is important to laugh and be silly sometimes. Florian and Joman, thanks for interesting lunch conversations about our different cultures. Laia, it has been lovely having you as my office mate, sharing tea and stories, and watching little birds learn to fly. I admire your patience and perseverance. Lee, thanks for all the good advice and comments, as well as for inserting some common sense and well-placed sarcasm into our lunch conversations. Lina, thank you for the fresh wind you brought to the lab. It has been wonderful to see you bring some well-needed organization to our mess and you are always so helpful. Punit, much appreciated for teaching me about chromatin remodelers, Indian cuisine, and cricket. It has been great fun sharing the office with you, thanks for all the laughs and lots of good advice. Peter, thanks for the companionship during the weekends when we both had yet another deadline to meet. Victoria, your laughter and sunny disposition has brightened up the lab so much, thank you for reminding me to have fun with
science. **Wenbo**, your determination and creativity are inspiring, thanks for the laughs and help in the lab.

**Agata**, my (bench) partner in science and knitting, it is hard to find someone equally well-versed in small RNAs and sheep breeds. It’s been wonderful working with you, thanks for helping me troubleshoot when things didn’t work as they should. **Carolina**, we miss you a lot in the lab, thank you for all your kindness and help over the years. **Christos**, it was great fun having you as my desk neighbor for a while, I hope you keep your passion for science. **Indranil, Ingela** and **Marianna**, thanks for fun lunch conversations.

Many thanks to the people at Bionut: **Sylvie** at LCI, for introducing me to the wonderful world of confocal microscopy. Your passion for science and cool new techniques is contagious. Thanks to **Fredrik** and **Patrick** at BEA for help with microarray and sequencing samples. Thank you to **Monica**, for help with all administrative troubles and questions.

Thanks to my friends, who made their own way towards a PhD at the same time as me. **Kathi, Kristin, Franz** and **Luzie**, although we lived so far apart during the last few years, I really enjoyed meeting up when we had the chance, to chat about the joys and sorrows of grad student life. Many thanks to **Dominika**, who started this whole journey through PhD land with me back in Edinburgh. We went through the highs and lows together, and I am so glad to have had your support and encouragement along the way.

Last but not least, a big ‘Dankeschön!’ to **my family**, who put up with us moving so far away for six years and only coming home a few times a year. Thanks to my parents in-law, **Sonja** and **Manfred**, for all the good discussions and encouragement. I am so grateful to my parents, your Skype calls kept my spirits up. Thank you for opening up the world to me and supporting me wherever I go. Finally, thanks to **Thomas**, without whom none of this would have been possible. Thank you for sharing this adventure with me.
6 REFERENCES


Devine, S.E., and Boeke, J.D. (1996). Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase II. Genes & Development 10, 620–633.


Heitz, E. (1928). Das Heterochromatin der Moose.


