Chromatin-based memory of prolonged cold exposure in *Arabidopsis thaliana*

Scott Berry

September 2015

A thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy

ABSTRACT

All living organisms contain genes. Turning these genes on and off at the appropriate times controls much of an organism's development and its responses to environmental conditions. In recent years, chromatin has emerged as an important player in orchestrating gene regulation. This thesis focuses on the role of chromatin in the maintenance of gene expression states and their inheritance through cell division.

FLOWERING LOCUS C (*FLC*) in the plant *Arabidopsis thaliana* is repressed by the prolonged cold of winter, and repression is maintained in subsequent warm conditions. The molecular complexes involved in modulating *FLC* chromatin are vital for *FLC* regulation and are conserved among plants and animals, making *FLC* a paradigmatic system for understanding of the role of chromatin in gene regulation.

After cold, *FLC* chromatin adopts a distinct configuration. In this study, experiments are used to show that this local chromatin 'state' instructs its own inheritance through cell division in growing plants. Thus, memory of winter cold is stored in the chromatin of the *FLC* gene.

Mathematical models developed in this work focus on understanding how chromatin states are maintained and also re-established after DNA replication. Minimal models are used to investigate if a particular set of interactions between chromatin and chromatin-modifiers can give rise to the qualitative behaviours, and quantitative results that are observed experimentally. Models developed here make predictions for the *FLC* system, and more generally show how cis and trans determinants of gene expression can be integrated by chromatin.

The role of transcription in determining chromatin states is also examined experimentally by studying the chromatin-associated protein LHP1. LHP1 is required for *FLC* repression and binds to modified histones associated with repressed *FLC* chromatin. In this work, it is shown that LHP1 also binds RNA and that this is important for its *in vivo* function.

CONTENTS

PUBLISHED WORK

This thesis includes research material from the following published work.

Berry, S., Hartley, M., Olsson, T. S. G., Dean, C., and Howard, M. (2015). **Local chromatin environment of a Polycomb target gene instructs its own epigenetic inheritance**. *eLife* 4:e07205.

Material from the following review articles is also included.
* Indicates that these authors contributed equally.

Berry, S., and Dean, C. (2015). **Environmental perception and epigenetic memory: mechanistic insight through FLC.** *The Plant Journal*, 83:133-148.

Richards, D.*, Berry, S.*, and Howard, M. (2012). **Illustrations of mathematical modeling in biology: epigenetics, meiosis, and an outlook.** *Cold Spring Harb. Symp. Quant. Biol.*, 77:175–181.

ACKNOWLEDGEMENTS

The decision to move to the other side of the world to study a Ph.D. was a big one, which involved many sacrifices but also opportunities for great adventure. This is even more true for my wife Jacque, who moved to Norwich, East Anglia ultimately because of the opportunity that this provided for me. Jacque, I am proud of how you embraced life in England and I thank you for your incredible support and friendship throughout this four years. Despite all the effort, I am sure we will look back on our time in Norwich as happy years.

The John Innes Centre, and in particular the rotation Ph.D. programme, have provided a wonderful environment for my growth as a scientist. I thank my fellow rotation students: Ania, Annis, Richard and Olu, and also Nick Brewin and Mike Merrick who organised the rotation programme. This fantastic programme gave me an overview of JIC science and the opportunity to identify the most fascinating research on site and pursue it. This programme also instilled in me an appreciation of the importance, breadth, and potential impact of plant science. I thank Richard Morris for his supervision throughout an entertaining and fruitful rotation project.

I would like to thank all members of the Dean and Howard groups. The many interesting discussions and debates have been integral to this work. Several individuals deserve special thanks: Jie Song, for an introduction to working with plants and plant molecular biology; Andrew Angel, for developing the original *FLC* vernalisation model and helping me to understand constraints on modelling in chromatin-based epigenetics; Hongchun Yang, for being the go-to man for any and all technical problems and protocols, and for such beautiful ChIP data that has been invaluable for the modelling; Robert Ietswaart, for thought-provoking discussions about transcription and chromatin; Stefanie Rosa, for help on the confocal microscope and for teaching me about Arabidopsis de-

velopment; Zhe Wu (Jake) for the collaboration on RNA polymerase II ChIP and interesting discussions comparing the autonomous and vernalisation pathways; Oleg Raitskin for help with the FLC-Venus pull-down and mass spectrometry; Finally, Zhen Tao, Shuqin Chen (Chi Chi), Huamei Wang, and my undergraduate students Theo Portlock and Hannah Tainton for all their help with the plant work.

To others at the JIC/TSL: Robert Sablowski, for being on my supervisory committee; Nick Pullen, for all the help with R and ggplot, and for generously providing the beautiful LaTeX template for this thesis; Grant Calder, for microscope training and assistance, and for work on the FLC-Venus time-lapse project; Christian Rogers, Andy Breakspear and Andrey Korolev, for helping me to establish Golden Gate cloning in the Dean lab; Zane Duxbury, for assistance in transformation of tobacco plants; Matthew Hartley and Tjelvar Olsson, for computing assistance and for developing the automated image quantification procedure; Finally, horticultural services, particularly Lesley Philips and Timothy Wells, for looking after the plants.

I would like to thank the John Innes Foundation for generous financial sponsorship of my student fees, stipend, and lab expenses. Without this support, none of this work would have been possible. Experiments were also funded by the BBSRC (BB/J004588/1), and the ERC (ENVGENE and MEXTIM).

Part of this work was undertaken at the FMI in Basel. I would like to thank Marc Bühler and his group for hosting me, in particular Veronika Ostapcuk for so much help with my LHP1 project.

I have had the honour of having two great scientists as my supervisors: Martin Howard and Caroline Dean. Martin's philosophy in how to combine minimal mathematical models with experimental data has been deeply influential. It has been a privilege to be part of a theory group studying such diverse topics as plasmid segregation, cell size control, starch metabolism, hyphal branching, transcriptional gene regulation, and epigenetics. Caroline's intuition and ability to extract fun-

damental principles from disparate data, literature and modelling has consistently provided a clear road forward and enabled the ideas developed in this work to be distilled and communicated effectively. The informal Dean group meetings were sometimes tense, but a great training-ground for presentation and discussion of ideas.

The highly successful collaboration between the Dean and Howard groups will stay with me forever as a shining example of how interactions between theorists and experimentalists can lead to progress in understanding complex biological systems.

Scott Berry
John Innes Centre, Norwich
September 2015

1 INTRODUCTION

Cells in our bodies display a variety of different morphologies and behaviours. These differences are determined not by differences in the genetic material contained in these cells but rather by persistent differences in the set of genes which are expressed. The differences between these cells are therefore not genetic but epi-genetic (literally, *above* genetic).

What causes these differences in gene expression between cells? In some cases it can be specific signals in the environment of the cell. These can be chemical signals generated by other cells [1–3], or abiotic signals such as local crowding or mechanical strain [4, 5]. However in some cases, a gene can be expressed not because an environmental signal continually instructs it, but because there is an internal mechanism within the cell that acts to maintain the gene expression state - even when the initial signal has disappeared [6]. For example, a skin cell expresses a set of genes that make it a skin cell, yet the signals which initially turned these genes on, and caused the differentiation of that cell from a stem cell, have long since passed. This skin cell can even be taken out of the organism, and will continue to express many of the genes specific to that cell type. In this sense, the skin cell has a memory of its identity because of the cell-autonomous mechanisms that maintain the expression of its 'skin cell' genes. So we see that the concept of epigenetics is intimately related to that of memory. In fact, at its simplest level, epigenetics is just the memory of gene expression.

Genes do not automatically stay 'on' or 'off' after being activated or repressed. Instead, these expression 'states' must be maintained by the continual action of specific regulators [7, 8]. It has been known for many years that gene expression can be controlled at the level of transcription by regulators called transcription factors. These factors ulti-

mately cause a gene to be expressed or repressed by influencing how often a gene is transcribed into RNA. Transcription factors are themselves encoded by genes, and their expression can therefore regulate other genes. The genes encoding transcription factors can potentially also regulate themselves. This so-called 'feedback'-regulation is widespread in biological systems and is the fundamental characteristic that allows networks of transcription factors to store epigenetic memory. That is, it is possible for a system of transcriptional regulators with feedback to have more than one stable state of gene expression [9–11]. To clarify this, a simple example is shown in Fig. 1.1A. In this network, protein A represses the transcription of gene B and protein B represses the transcription of gene A. These proteins also feed back to positively regulate the expression of their corresponding genes. It is easy to see that if the system has high levels of protein A, then gene B will be repressed and gene A will be activated, producing more protein A. Thus 'A high / B low' is a stable state. Similarly, 'A low / B high' is a stable state. That is, if the system is set into one of these two states, it has the ability to maintain itself in this state. This simple system therefore exhibits memory. To exhibit memory, a system must have at least two stable states, which is referred to as bistability [12]. The two stable states in this model are also heritable, since when a cell divides the regulatory proteins will be divided roughly equally between the daughter cells via the cytoplasm (Fig. 1.1A). This type of epigenetic memory is sometimes referred to as 'cytoplasmic memory', or more commonly 'trans memory' [6]. This is because the epigenetic information is stored as a diffusible signal that acts in trans[a] to regulate expression of its targets. While this example is deliberately simple, it underlies the basic thinking behind networks of transcription factors, which are believed to specify and maintain cell types in multicellular organisms.

In recent years, a qualitatively different mechanism for maintaining gene expression states has been proposed, which does not rely on in-

[a] derived from the Latin trāns, meaning 'across, on the far side, beyond'.

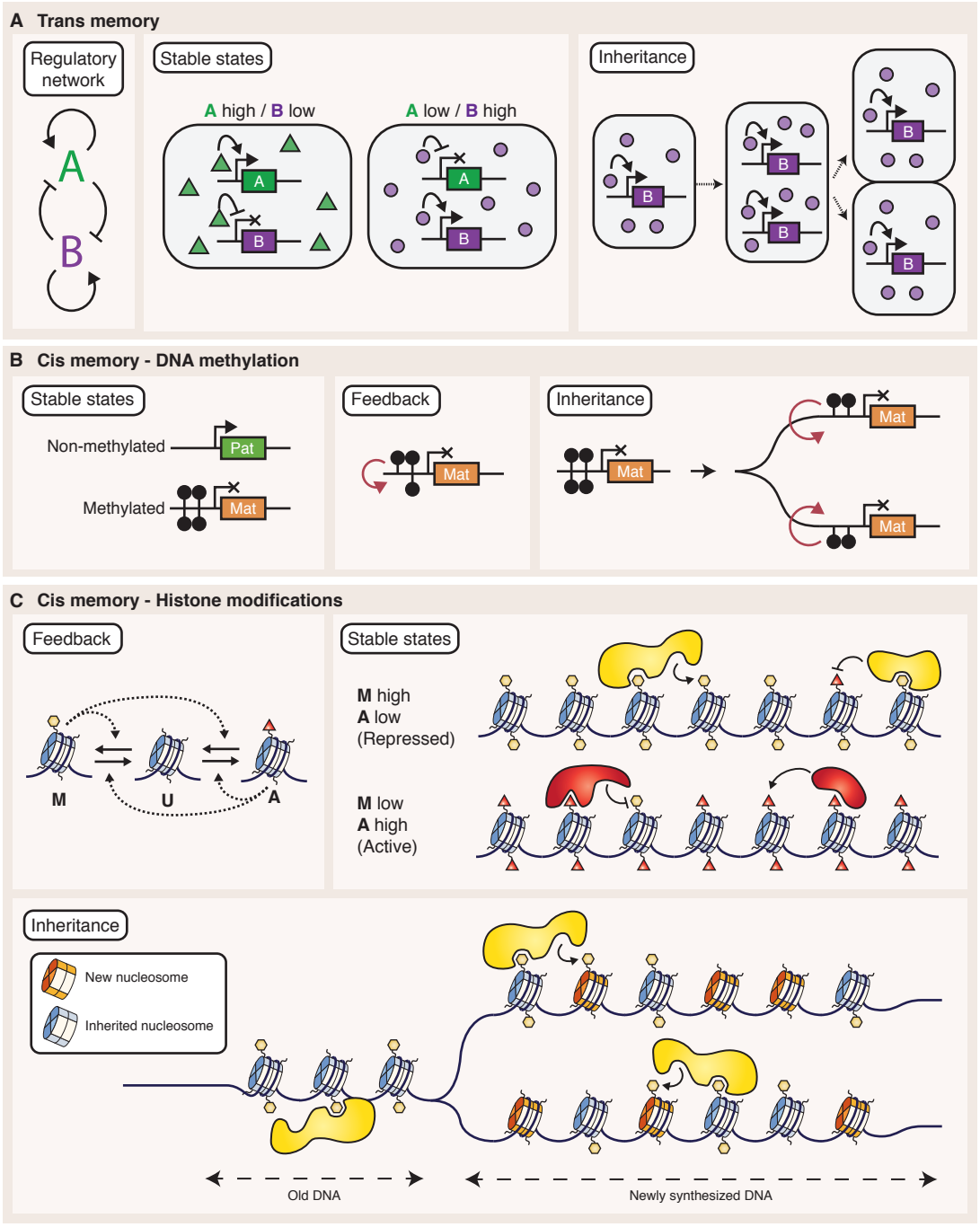


Figure 1.1: Mechanisms of maintaining gene expression states. (A) Trans epigenetic memory with transcription factor concentrations as the heritable information carrier. Pointed arrowheads represent activation, while flat arrowheads indicate repression. Green triangles and purple circles represent diffusible A and B proteins, respectively. (B) Cis epigenetic memory in genomic imprinting with DNA methylation (black circles) as the heritable information carrier. Pat = Paternal gene copy, Mat = Maternal gene copy. (C) Cis epigenetic memory with histone modifications as the heritable information carriers. Opposing M and A modifications on histone tails are depicted as yellow hexagons and red triangles, respectively. Core nucleosome particles made up of histone octamers are depicted as beads on a DNA string.

structions from diffusible regulators. This mechanism is referred to as 'cis memory', because the epigenetic information is physically located in cis[b], in the vicinity of the gene being regulated [6].

1.1 Cis epigenetic memory

How could epigenetic information be stored in cis? To understand more about cis memory it is necessary to first introduce chromatin: the proteins, RNA and DNA that make up chromosomes and their packaging in eukaryotic cells. The main proteins associated with DNA in chromatin are histones. The four 'core' histones H2A, H2B, H3 and H4 form an octomeric structure known as a nucleosome, around which 147 bp of DNA is wrapped [13, 14]. These nucleosomes assemble onto DNA approximately every 200 bp over the entire chromosome. Histones are highly basic proteins that bind tightly to negatively charged DNA. This generally impedes the accessibility of DNA to other factors. The location of histones relative to important regulatory DNA sequences and the strength of histone-DNA contacts could therefore play important roles in gene regulation. In chromatin, both histones and DNA can have chemical moieties such as methyl groups added to them [15–17]. These may act as molecular markers that could potentially cause changes to regulation of the underlying genes. Histones can be moved along the chromatin fibre by chromatin remodellers [18] and can also be exchanged for replacement histones throughout the cell cycle [19]. The latter process is known as histone turnover.

Chromosome structure can be divided into two broad classes: heterochromatin and euchromatin. Heterochromatin is a gene-poor region of a chromosome —rich in repetitive sequences, which surrounds the centromeres and telomeres in many eukaryotes. Heterochromatin remains relatively condensed throughout the cell cycle. The remainder of the genome is designated euchromatin. Euchromatin contains the majority of genes and is less condensed.

[b] From the Latin preposition cis, meaning 'on this side of'.

The most well-understood example of cis memory is that of genomic imprinting [20]. A gene is said to be 'imprinted' if it is silenced at either the maternal or paternal copy, but not both [20]. Imprinting results in the stable inheritance of a silenced gene expression state at only one gene copy out of two in the cell, in the absence of any sequence differences between the two gene copies[c]. The difference in expression is therefore epigenetic, because no differences in DNA sequence exist between the two copies. In this case, it is also clear that the epigenetic information must not be stored as a diffusible signal, otherwise the two copies of the gene would both be in the same expression state. In all cases known so far, epigenetic memory in imprinting is stored in patterns of DNA methylation [20]. Classically, a region known as the differentially methylated region is methylated at the silenced copy and unmethylated at the active copy. DNA methylation in the CG dinucleotide context can be copied from parental to daughter DNA strands by a so-called maintenance methyltransferase (Fig. 1.1B). After DNA replication of a symmetrically methylated CG dinucleotide, each of the daughter strands carries a hemi-methylated cytosine in the CG context. A maintenance methyltransferase specifically recognises this hemi-methylated CG and adds a methyl group to the unmethylated cytosine [22]. In this way, the DNA methylation pattern can be replicated and thereby carry epigenetic memory.

Another proposal for a cis-based epigenetic memory is based on post-translational modifications of histones (Fig. 1.1C) [23, 24]. In this case, the idea is that a particular histone modification can recruit an enzyme capable of adding the same type of modification to a nearby unmodified histone. In this way, a region of chromatin could become predominantly covered in a certain type of modification. Figure 1.1C shows a simplified model with two 'types' of histone modifications — labelled as M and A. DNA is replicated semi-conservatively. For DNA

[c]Repression of the mating-type cassettes in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (reviewed in [21]) is not normally considered imprinting due to the predominantly haploid nature of these yeasts and the differences in sequence between the two loci.

methylation this meant that each of the daughter strands inherited a methylated cytosine (Fig. 1.1B). However, histones are instead segregated conservatively between the two daughter DNA strands [25, 26]. Nucleosomes remain close to their pre-replication locations on daughter chromosomes [27] and the 'gaps' are filled by newly-synthesized nucleosomes. If the histone modifications remain in place as the nucleosomes are inherited, then they could recruit the enzymes necessary to modify the new histones and thereby perpetuate the chromatin state (Fig. 1.1C). Theoretical studies have suggested that histone-modification-based epigenetic memory is possible [28], but experimental evidence that inherited histone modifications can indeed act as inherited memory elements is so far inconclusive [29–32].

In summary, we have seen that differences between the cells of an organism arise because of differences in gene expression rather than differences in genetic material. The concept of epigenetics was introduced as the study of heritable changes in gene expression in the absence of initiating signals. Finally, how gene expression levels could be maintained by both cis and trans memory mechanisms was discussed. It is now time to introduce the specific model of epigenetic gene regulation that is the focus in this work.

1.2 Memory of cold exposure in Arabidopsis through *FLC* regulation

In plants, vernalisation is the acquisition of competence to flower following a period of prolonged cold exposure [33]. A requirement for vernalisation ensures that plants over-winter vegetatively and flower in the following spring [34]. Central to this process in *Arabidopsis thaliana* (hereafter, Arabidopsis) is regulation of the floral repressor *FLOWERING LOCUS C* (*FLC*) [35, 36]. Expression of *FLC* throughout the Arabidopsis life cycle is shown in Figure 1.2A. High expression of *FLC* before cold exposure ensures that key activators of the floral transition such as *FLOWERING LOCUS T* (*FT*) remain repressed [37]. Prolonged

cold exposure in winter, however, causes transcriptional repression of *FLC*, which is maintained after plants are returned to warm conditions [36, 38]. After cold —when *FLC* is repressed, the repression of flowering is relieved and Arabidopsis plants quickly make the transition from vegetative to reproductive development. Viewed in this way, vernalisation is a classic epigenetic phenomenon, whereby a transient environmental stimulus (i.e. cold) generates a stable change in gene expression (i.e. *FLC* repression) that is maintained in subsequent warm conditions. Furthermore, the memory of *FLC* repression is stable in growing plants, which are made up of non-dividing and dividing cells, so repression must be mitotically heritable. Unlike in animals, the germ line in plants arises from somatic cells late in development [39]. *FLC* expression is reset during embryogenesis to ensure a vernalisation requirement in the next generation.

Temperature conditions can vary dramatically from week-to-week. For a vernalisation requirement to function correctly, plants must not only sense and remember cold, but also distinguish long and short periods of cold. For example, if a plant experiences a week of cold weather in autumn followed by several warmer weeks, it is important that this is not perceived as the end of winter. In other words, vernalisation should be quantitative. This was elegantly explained for the case of Arabidopsis, when it was found that *FLC* expression after cold depends quantitatively of the duration of prior cold exposure [38]. *FLC* expression level therefore constitutes a quantitative memory of cold.

FLC is up-regulated by *FRIGIDA* (*FRI*), which leads to a requirement for vernalisation and a winter annual growth habit [40] (Fig. 1.2C). Many independent mutations in the *FRI* gene have been isolated from natural populations [41]. This indicates that loss of *FRI* is a common evolutionary strategy for generating a rapid-cycling growth habit in which multiple generations can be achieved per year. One such *FRI* loss-of-function accession is the commonly used lab accession Columbia-0 (Col-0) [41]. Studies of vernalisation rely on high *FLC* expression be-

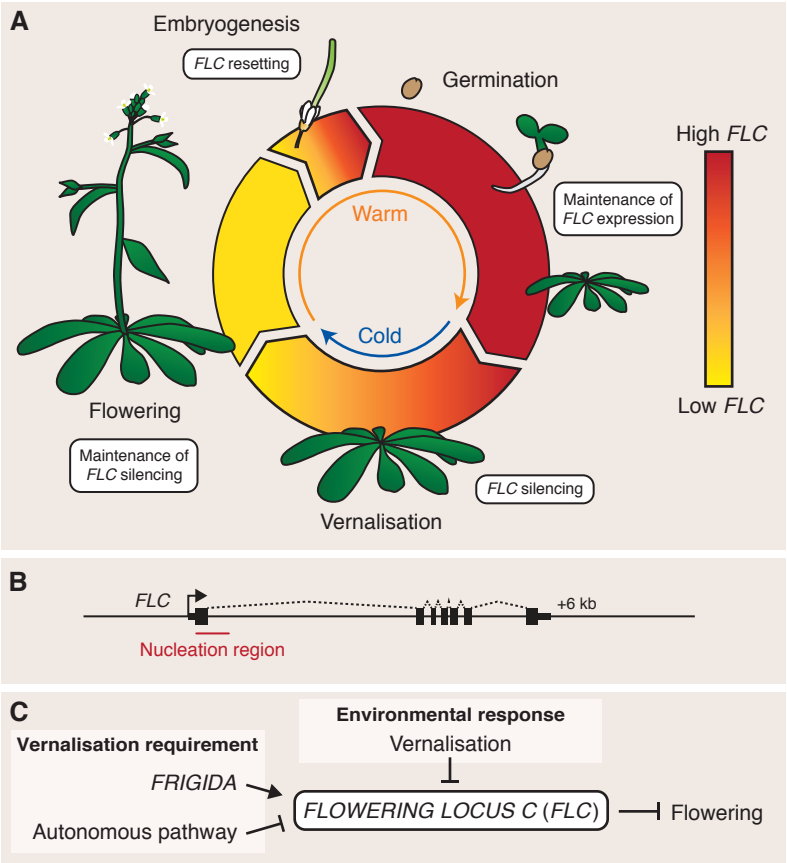


Figure 1.2: Expression of *FLC* throughout the life-cycle of winter-annual Arabidopsis thaliana. Growth before and after cold exposure are periods of stable *FLC* expression, whereas vernalisation and embryogenesis down-regulate and up-regulate *FLC* expression, respectively. (B) Diagram of *FLC* genomic DNA. Black boxes depict exons of *FLC* mRNA and dashed lines represent the splicing pattern. The 'nucleation region' is immediately downstream of the transcription start site. (C) *FLC* is constitutively activated by *FRIGIDA* and repressed by the autonomous pathway in non-vernalising conditions. In addition, high *FLC* expression can be reduced dynamically during development by vernalisation. High expression of *FLC* delays flowering.

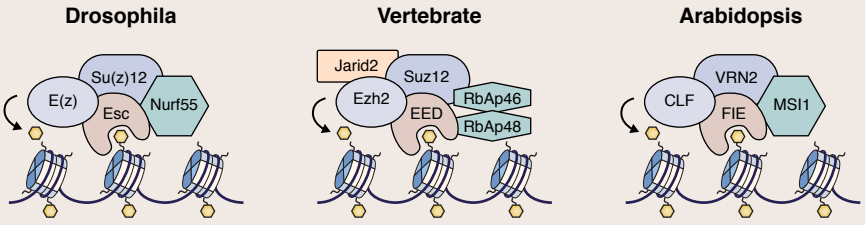
fore cold exposure, so it is common to use the accession Col-FRI —a near isogenic line in the Col-0 background which contains an active *FRI* allele from the San Feliu-2 (Sf-2) accession [40].

Another pathway acting antagonistically to *FRIGIDA* to regulate *FLC* expression is the autonomous pathway [42] (Fig. 1.2C). By definition, plants carrying mutations for genes in the autonomous pathway show high *FLC* expression (even in the absence of *FRI*), but respond to vernalising cold. The autonomous pathway therefore comprises a set of genes that act to repress *FLC*, independently of vernalisation [43].

1.2.1 *A role for chromatin in FLC repression*

Genetic screens have identified several of the key players necessary for high *FLC* expression in non-vernalising conditions and establishment and maintenance of *FLC* repression in response to cold exposure. The so-called *VERNALIZATION* (*VRN*) mutants were identified due to their inability to accelerate flowering after cold [44–47]. One of these, VRN2, was identified as a core component of the a highly conserved protein complex known as Polycomb repressive complex 2 (PRC2) [45]. Subsequent mechanistic studies identified *FLC* as a Polycomb target gene, with maintenance of *FLC* repression after vernalisation being dependent on PRC2 activity [48, 49].

Polycomb Repressive Complex 2. The Polycomb group of proteins (PcG) were originally identified in Drosophila. PcG mutants typically fail to maintain repression of key developmental regulators, such as the Hox genes, during embryogenesis and consequently display morphological defects [50]. PRC2 binds directly to its target genes in order to achieve repression. In Drosophila, PRC2 is made up of Enhancer of zeste (E(z)), Suppressor of zeste 12 (Su(z)12), Extra sex combs (Esc) and Nucleosome remodelling factor 55 kDa subunit (Nurf55) [24, 51]. E(z) is a methyltransferase, capable of methylation of histone H3 at lysine 27 (H3K27me) [52–55]. Su(z)12 is necessary for functional PRC2 complex



Drosophila	E(z)	Su(z)12	Esc	Nurf55
Vertebrate	Ezh2	Suz12	Eed	RbAp46 RbAp48
Arabidopsis	CLF SWN MEA	VRN2 EMF2 FIS2	FIE	MSI1

Figure 1.3: The conserved Polycomb Repressive Complex 2 (PRC2). Comparison of PRC2 from Drosophila, vertebrates and Arabidopsis. Core nucleosome particles made up of histone octamers are depicted as beads on a DNA string. Yellow hexagons represent H3K27me3. Each PRC2 complex binds to chromatin and methylates histone H3 at Lys-27. Table showing the various homologues of the Drosophila PRC2 components in Arabidopsis.

formation and chromatin targeting [56, 57]. Esc contributes to high affinity binding of PRC2 to histones [58, 59] and specifically recognises H3K27me2 and H3K27me3, leading to allosteric activation of E(z) [60]. Nurf55 makes contacts with histones and Su(z)12, and appears to play a role in sensing the chromatin state of potential target genes [61, 62]. Vertebrate and Arabidopsis equivalents of these Drosophila proteins are shown in Fig. 1.3.

The conserved role of PRC2 is to maintain repression of its target genes. However, the mechanism by which PRC2 acts to maintain transcriptional repression is not well understood. When repressed, Polycomb target genes are associated with high levels of H3K27me2/3 and low levels of histone acetylation. It is generally believed that Polycomb-repressed chromatin adopts a more compact structure which may act to prevent binding of specific transcription factors or the general transcription machinery [63]. H3K27me is associated with gene repression in all eukaryotes containing PRC2, and catalytic activity of E(z) is required for PRC2 function. However, until recently, it remained elusive as to

whether H3K27-methylation was actually required for gene repression by PRC2. To address this problem, all gene copies encoding histone H3 in Drosophila were replaced with histones containing a lysine-27 to arginine mutation [64]. This caused de-repression of Polycomb target genes, reproducing the phenotype of a PcG mutant. This finding confirms that H3K27 is the relevant physiological substrate of E(z) for Polycomb silencing.

Polycomb Repressive Complex 1. In addition to PRC2, other proteins in the Polycomb group (PcG) form a distinct complex known as Polycomb Repressive Complex 1 (PRC1) (reviewed in [24, 65]). PRC1 and PRC2 generally occupy common targets across mammalian and Drosophila genomes. PRC1 is able to catalyse mono-ubiquitination of histone H2A lysine 119 (H2Aub1) [66, 67], and also compact chromatin *in vitro* [68] and *in vivo* [69]. However, unlike PRC2, the enzymatic activity of PRC1 seems to be dispensible for its *in vivo* function [69, 70]. PRC1 contains chromodomain-containing protein Polycomb (Pc), which is believed to recognise H3K27me3 and mediate targeting of PRC1 to repressed PRC2 target genes [71, 72]. However, more recently it has been shown that PRC1 does not act entirely downstream of PRC2, and can even recruit PRC2 [73].

Until recently, PRC1 was thought to be absent in Arabidopsis. This was because initial bioinformatic analyses found no obvious subunit homologs [74] and also because H2Aub1 could not be detected by mass spectrometry [75]. However, subsequent database searches uncovered putative structural homologues for several PRC1 subunits [76] and mutants for these proteins were shown to have developmental defects [77]. Moreover, *in vitro* H2A ubiquitination activity was demonstrated for these factors [78]. It is important to note, however, that physical interactions between putative plant PRC1 components have not been confirmed *in vivo* [77–79], so the existence of a PRC1 'complex' in plants remains controversial.

Molecular events at the *FLC* locus in vernalisation. There are multiple homologues for some PRC2 subunits in Arabidopsis (Fig. 1.3). The specific PRC2 complex located at *FLC*, which is important for the vernalisation response, is comprised of core PRC2 components FIE, VRN2, MSI1 and SWN or CLF as well as plant homeodomain (PHD) proteins VRN5, VIN3 and VEL1 [47–49, 80, 81]. Before cold exposure, when *FLC* is highly expressed, H3K27me3 levels at *FLC* are low. During cold, expression of VIN3 is induced and quantitatively accumulates during longer cold, resulting in formation of a protein complex immediately downstream of the transcription start site of *FLC* —the so-called nucleation region (Fig. 1.2B). VIN3, VRN5 and VEL1 all contain PHD domains, so this PRC2 complex is designated PHD-PRC2. PHD-PRC2 results in deposition of H3K27me2/3 within the nucleation region. After cold, this 'nucleated' H3K27me3 can spread outwards to cover the entire *FLC* gene body [49, 82, 83]. 'Spreading' is hypothesized to be important for stable silencing in warm conditions following cold [83]. Other events at the *FLC* during the cold include up-regulation of the *FLC* antisense transcripts, named *COOLAIR* [84]. The role of *COOLAIR* in vernalisation is currently unclear. However, transgenic *FLC* lines with reduced *COOLAIR* expression show a slower reduction in *FLC* transcription during the cold [85].

PRC2 can bind to, and is activated by, H3K27me2/3 - the same histone mark that it deposits [60]. This suggests a locally-acting positive feedback mechanism, which could act to maintain high H3K27me3 levels at *FLC* after vernalisation. This is reminiscent of the model proposed earlier for cis memory based on post-translational modifications of histones (Fig. 1.1C, p. 19). That is, the H3K27me3 which covers *FLC* after vernalisation, could act to recruit PHD-PRC2 to reinforce the H3K27-methylation status of nearby histones. In this way, H3K27me3 could instruct its own maintenance and mitotic inheritance. Mathematical modelling based on this hypothesis showed that the observed H3K27-me3 levels at *FLC* were consistent with this hypothesis [83]. To quanti-

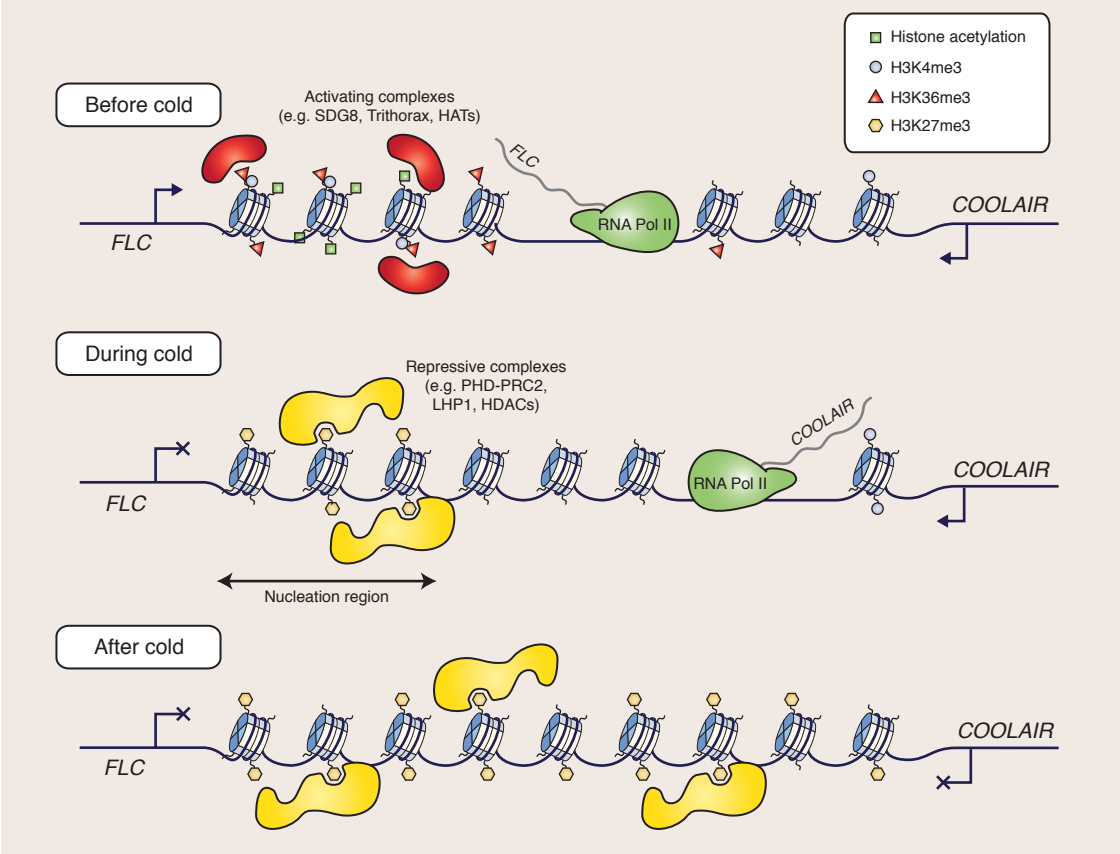


Figure 1.4: Molecular events at *FLC* during vernalisation. Before cold, *FLC* chromatin is associated with 'active' histone marks H3K4me3, H3K36me3 and acetylated histones. During cold, a PHD-PRC2 complex containing VIN3 accumulates at the nucleation region, leading to local deposition of H3K27me3 and depletion of H3K4me3, H3K36me3 and histone acetylation. After cold, H3K27me3 spreads out over the entire *FLC* gene body, concomitant with stable epigenetic repression. HAT = histone acetyltransferase, HDAC = histone deacetylase.

tatively fit the data, however, the model assumed that the population of cells assayed was actually heterogeneous —with some cells having high H3K27me3 at *FLC* (*FLC* silenced), and other cells having low H3K27-me3 (*FLC* expressed). This prediction was validated experimentally by examining expression of an *FLC* reporter in fixed root tissue after different lengths of cold [83]. Intriguingly, this suggested an answer to the long-standing problem of how to maintain a quantitative epigenetic memory of cold exposure. At the single-cell level, *FLC* expression may actually be just ON or OFF, with the number of cells that repress *FLC* after cold depending quantitatively on the duration of prior cold. Viewed in this way, each cell has a certain probability of switching from an *FLC*-ON to an *FLC*-OFF expression state during cold, a state which is then maintained and remembered through cell division after cold. The mathematical model of histone modifications at *FLC* also suggested that the 'nucleation' of H3K27me3 that is observed during cold exposure could actually drive the recruitment of PHD-PRC2 to the gene body after cold exposure, through the H3K27me3-based positive feedback mechanism outlined above [83]. Thus, 'nucleation' could cause 'spreading' of H3-K27me3 to the gene body.

The current view of *FLC* regulation through vernalisation can therefore be summarised as a state-switching mechanism: *FLC* is initially expressed highly in all cells, with cold stochastically inducing cells to switch to the *FLC*-repressed state. *FLC*-ON and *FLC*-OFF cells are then epigenetically maintained after cold [43]. This is referred to as 'digital' repression because the quantitative information about the length of cold exposure is encoded digitally in the number of *FLC*-OFF cells.

Cell-autonomous *FLC* repression may be converted back to an analogue flowering-induction signal at the level of a whole plant by floral integrators regulated by *FLC*. For example, one of the genes directly repressed by *FLC* is *FLOWERING LOCUS T* (*FT*) [37], which is expressed in the phloem companion cells in the vasculature of the leaves. FT subsequently moves from the leaves to the shoot apex to induce flower-

ing [86]. Movement of FT and other factors throughout the plant may act to average *FLC* expression between different parts of the plant and thereby provide an indicator of 'readiness-to-flower' at the whole-plant level.

1.3 Instructive versus responsive chromatin

Genetic and molecular biology experiments performed so far are consistent with the hypothesis that a histone-modification based cis epigenetic memory maintains *FLC* repression after prolonged cold. In this case, the *FLC* chromatin state would instruct expression of the underlying gene and the inherited H3K27-methylated histones would be the carriers of epigenetic memory. However, all available data are also consistent with an epigenetic memory stored in trans, provided that PRC2 is required by regulatory trans-factors to achieve *FLC* repression. It other words, it is not clear whether the local chromatin state determines gene repression or instead if *FLC* repression is determined by the concentrations of diffusible regulatory factors, which act by recruiting PRC2 to *FLC* to achieve repression. In the former case, *FLC* chromatin would be instructive, because the chromatin state instructs its own inheritance. In the latter case, *FLC* chromatin would be responsive, since the regulatory trans-factors network stores the epigenetic memory and the chromatin simply responds to the trans-factor concentration by adopting a particular configuration and expression level. A schematic contrasting instructive and responsive chromatin is shown in Fig. 1.5. The key difference between these two models is where the memory is stored. Responsive chromatin (trans memory) acts as a intermediary, enabling the trans-factors to achieve a certain gene expression state, whereas instructive chromatin (cis memory) actually stores epigenetic memory locally.

Experiments to distinguish cis and trans epigenetic memory for *FLC* regulation after vernalisation are presented in Chapter 2.

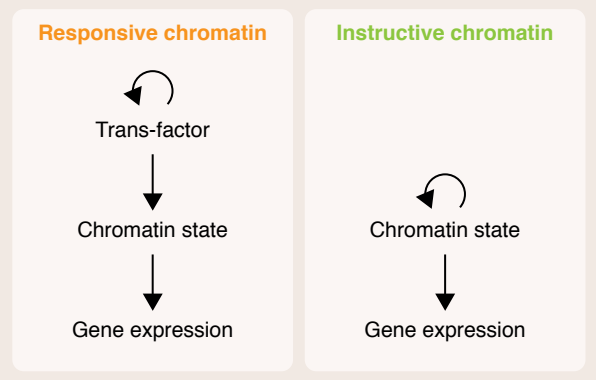


Figure 1.5: Responsive versus instructive chromatin. Reponsive chromatin responds to the epigenetic state specified by the concentration of a trans-acting factor but may still be necessary to achieve correct gene expression. Instructive chromatin instructs its own inheritance and determines the expression level of the underlying gene.

1.3.1 *Definitions of epigenetics*

This lack of understanding regarding the identity of the epigenetic memory elements is not isolated to *FLC* regulation. It is currently unknown whether any histone modification, in any organism, is the decisive factor that instructs inheritance of a gene expression state. This has led to much debate regarding the use of the term epigenetic to describe histone modifications. Many researchers have argued that the term epigenetic should be reserved for those factors which have been proven to transmit memory of an expression state [29–32]. Despite this objection, however, the term 'epigenetic' has almost become synonymous with 'chromatin modification', and the term 'epigenome' has come to describe genome-wide patterns of chromatin modifications. Therefore, there are fundamental differences in what different people mean by the term 'epigenetic'. In an attempt to bring clarity, some eminent researchers in the field suggested a new definition of epigenetics as the study of "… heritable phenotype(s) resulting from changes in a chromosome without alterations in the DNA sequence" [87]. They went on to suggest that epigenetic 'maintainers' may include "DNA methylation, histone modifications, histone variants, nucleosome positioning,

and others." However, with the exception of DNA methylation in certain contexts [16], none of these processes have been shown to be capable of maintaining gene expression states. Furthermore, trans memory (which has been shown in both natural [88] and synthetic [89] systems to be sufficient for generating bistable gene expression states) is omitted from this definition. Rather than clarify, this new definition therefore brings yet more confusion.

In this thesis, the term epigenetic is used in relation to the maintenance of gene expression states in non-dividing cells or through mitotic cell division in dividing cells. Post-translational modifications of histones are referred to as histone modifications, and the term chromatin modification refers, more generally, to covalent chemical modifications of DNA, histones, or chromatin-associated proteins and RNA.

1.4 Mathematical models of histone-modification-based memory

The first theoretical consideration of how epigenetic memory could be stored in patterns of histone modifications was based on heterochromatic silencing of the mating-type locus in the fission yeast *Schizosaccharomyces pombe* (*S. pombe*) [28]. In wild-type yeast, this locus is found in the silenced state, but through genetic perturbation, a system that exhibits bistability can be generated [90]. In a theoretical study, Dodd *et al.* developed a highly-simplified model of chromatin to explore the conditions required for chromatin states to self-perpetuate. It was assumed that histones could exist in three mutually exclusive states, which were labelled M, U and A (methylated, unmodified, acetylated) (M-U-A model) (Fig. 1.1C). Transitions between the M, U, and A states are driven by histone-modifying enzymes, and the process of histone turnover, in which histones are removed from DNA and replaced independently of DNA replication [19]. The model hypothesized positive feedback, so that the M and A modifications could each recruit modifications of the same type to other histones in the region, and mutual

antagonism, so that these marks could also remove modifications of the opposing type (Fig. 1.1C, p.19). Using stochastic Monte Carlo simulations, this model was shown to be capable of generating bistability. That is, the feedbacks in the model meant that a region of chromatin tended to be predominantly covered in either A or M marks, but not both at the same time.

To simulate the random partitioning of nucleosomes between daughter chromosomes at DNA replication[d] [25–27], Dodd *et al.* replaced half of the nucleosomes (chosen randomly) with unmodified U nucleosomes, regardless of their modification state. The cis-acting feedbacks allow both stable states to survive this perturbation because the inherited M or A histones can effectively re-recruit more modifications of the same type and thereby restore the state. Thus, in the model, the modified histones act as the inherited epigenetic memory elements. Importantly, this model does not require a copying mechanism to be specifically recruited during DNA replication. The internal dynamics that exist throughout the cell cycle, together with the inherited histone marks are sufficient to re-establish the chromatin state.

Three important theoretical results emerged from this model. First, it was shown that larger systems (i.e. more histones) are more stable. This is primarily because systems with few memory elements were unable to buffer the perturbation of DNA replication with high fidelity. (For example, a system of 3 nucleosomes has a $1/2^3 = 1/8$ probability of losing all nucleosomes at each DNA replication). Second, it was shown that the rate of transitions towards the M mark needs to increase in a more-than-linear fashion with the number of M marks present (and vice versa for the A mark). This can be achieved naturally in the three state M-U-A model since A to M conversions require two consecutive transitions in the direction of M, so the rate is proportional to the number of M marks squared. In other models without an intermediate U state, this can be achieved by adding explicit nonlinearity to the

[d] Described in Sec. 1.1. Illustrated in Fig. 1.1C.

model feedbacks[e] [94]. Third, the model required enzymes recruited by histone modifications to act beyond nearest neighbours. This was necessary to prevent 'islands' of one histone modification from invading a region of chromatin predominantly covered in the opposite state. These so-called 'long-range interactions' were assumed to be mediated through higher-order chromatin structure, which would allow nucleosomes physically distant along the chromatin fibre to be actually be in close proximity.

There have been many subsequent theoretical studies of epigenetic memory based on histone modifications in different biological contexts. Many of these have considered 'two-state' models [94–99]. In these cases, it has been necessary to introduce nonlinearity explicitly somewhere in the model. Typically, this has been explained in a mechanistic sense by multiple marks of the same type acting together to facilitate the recruitment of a single histone modifying enzyme [98].

Where specific biological systems have been considered, the majority of theoretical studies have focused on the heterochromatic regions of either *S. pombe* [28, 100], *Saccharomyces cerevisiae* (*S. cerevisiae*) [94, 97, 98, 101–103], or mouse [104]. Only three studies have considered Polycomb-dependent gene silencing [83, 105, 106], all for the case of *FLC*.

This existing model of state-switching of *FLC* chromatin during vernalisation [83] is used as a starting point for Chapter 3. This chapter presents results of experiments that challenge the original model and then turns back to modelling to understand the implications of these results for our mechanistic understanding of *FLC* repression.

[e]Nonlinearity is known to be necessary but not sufficient for bistability from studies of trans-regulatory networks [12, 91–93]

1.5 Interactions between transcription and chromatin in cis memory

So far, the process of transcription has not been mechanistically integrated into mathematical models of histone-modification-based epigenetic memory. This may be because much of the literature has focused on large regions of heterochromatin rather than the regulation of individual genes. Existing mathematical models typically contain an implicit assumption that transcription levels are determined by chromatin states.

How transcription and Polycomb silencing interact to specify chromatin states forms the basis of the second half of this thesis. Both theoretical and experimental approaches are used to address this question. In Chapter 4, mathematical modelling is used to examine the hypothesis that transcription acts to directly antagonise Polycomb silencing. Taking an experimental approach, Chapter 5 then focuses on how interactions between nascent RNA and chromatin-associated proteins contribute to determining chromatin states.

1.5.1 *Transcription as an opposing state to Polycomb silencing*

Models such as those illustrated in Sec. 1.4 generally assume the existence of an 'active' histone modification that is mutually exclusive with the histone modification associated with repression. By analogy with repression, these 'A-marks' are generally assumed to have their own mechanism of positive feedback.

The Trithorax group of proteins (TrxG) are possible candidates for the enzymatic complexes that deposit 'A-marks' at Polycomb group (PcG) target genes. These proteins have been historically associated with maintenance of active expression states at PcG target genes [24]. In fact, regions of DNA recognised by PcG proteins in Drosophila, called Polycomb response elements, are often also recognised by TrxG proteins, and are therefore also referred to as Trithorax response elements [24]. While mechanistic details regarding the antagonism of PRC2 by the

TrxG-associated 'A-marks' are emerging [61, 62], a mechanism of positive feedback for such marks similar to the binding of H3K27me3 by PRC2 (Fig. 1.3) [60, 107] has not been discovered. One possibility is that the positive feedback mechanism for A-marks is actually mediated through the process of transcription.

There are other reasons to suspect that the process of transcription itself can antagonise Polycomb silencing. In Drosophila and mammalian systems, H3K27me can accumulate 'by default' at Polycomb target genes as a result of their transcriptional repression [108–111]. Furthermore, the processes of histone turnover and H3K27-demethylation, both of which act in opposition to H3K27me3 accumulation, are known to be linked to transcription [112–118].

For *FLC*, it is known that high H3K27me3 levels are associated with low *FLC* expression in different mutant backgrounds and accessions [119–123]. In vernalisation time-course experiments, transcription levels at *FLC* are also anti-correlated with H3K27me3 levels (Fig. 1.6) [106, 123]. In addition, it has been shown that increasing transcription at *FLC* using an artificial inducible promoter leads to loss of H3K27me3, and conversely that reduction from a high transcription state leads to H3K27me3 accumulation [121]. Finally, *FLC* repression in early cold is independent of *VIN3* [84] —an important member of the PHD-PRC2 complex that accumulates at *FLC* during cold and which mediates deposition of H3K27me3-nucleation (Sec. 1.2.1). This suggests that down-regulation of transcription may be a pre-requisite for establishment of H3K27me3 at *FLC* during cold.

These results support the hypothesis that active transcription of *FLC* is antagonistic to Polycomb silencing. The observation that antagonism of Polycomb silencing by transcription exists in many species suggests that this is a general feature of the Polycomb system. This hypothesis is pursued in the theoretical models presented in Chapter 4.

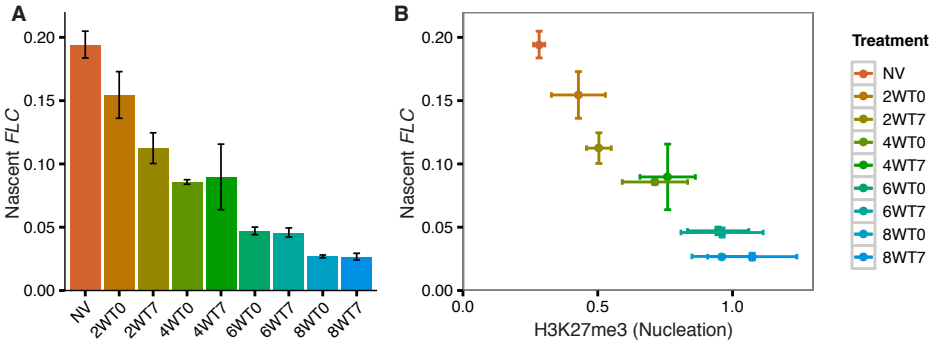


Figure 1.6: *FLC* repression is correlated with H3K27me3 accumulation. Nascent *FLC* transcript levels in Col-FRI plants throughout vernalisation plotted by (A) treatment or (B) average H3K27me3 level in the *FLC* nucleation region. Treatment labels *c*WT*x* represent *c* weeks cold treatment followed by *x* days growth in warm conditions, NV represents non-vernalised plants. Published data obtained from Hongchun Yang. Reproduced from [123] with permission. Details on experimental procedures provided in Secs. 2.5.6, 2.5.9.

1.5.2 *RNA as a modulator of chromatin states*

Over the last decade, it has been appreciated that transcription is not limited to protein coding genes [124–127]. In fact, it has been proposed that initiation of transcription is often imprecise [128, 129] and may occur wherever chromatin is depleted of nucleosomes, or where nucleosomes are not tightly bound to DNA [130].

Genome-wide studies have led to the discovery of thousands of RNA transcripts that do not encode for protein —so-called 'non-coding RNA' (ncRNA). Long ncRNA (>200 bp) can be highly abundant and processed similarly to protein-coding mRNA. The biological functions of these transcripts are mostly unknown but the concept of ncRNAs as contributing to the regulation of chromatin states is emerging from several isolated cases (reviewed in [131]).

ncRNA are proposed to act in specifying chromatin states in cis (at their site of transcription) [132, 133] and also in trans (independent of where they are transcribed) [134, 135]. ncRNA may act to recruit [136, 137] or remove [138, 139] chromatin-associated proteins, or to act as molecular scaffolds to facilitate protein complex formation [135, 140]. The discovery of thousands of highly-conserved non-coding transcripts

in mammalian genomes [127] has led to much excitement about their potential regulatory functions and implications for human development and disease [141]. One outcome of this excitement is that many studies in epigenetics have begun to consider how RNA influences the activity of chromatin modifiers.

In particular, it was shown that the enzymatic subunit of mammalian PRC2, Ezh2, as well as an accessory component, JARID2, can interact non-sequence-specifically with RNA *in vitro*, and both are associated with thousands of RNAs *in vivo* [142–145]. RNA-binding by PRC2 inhibits its enzymatic activity [139, 144]. This may provide another mechanism by which actively-transcribed genes are prevented from being silenced by PRC2 [144].

It has also been reported in the fission yeast *S. pombe* and in mammals, that proteins of the Heterochromatin Protein 1 (HP1) family are able to bind RNA [146, 147]. These findings are especially relevant for *FLC* regulation because the homologue of HP1 in Arabidopsis, called LIKE HETROCHROMATIN PROTEIN 1 (LHP1) is required for maintenance of *FLC* repression after vernalisation [148]. Although LHP1 is structurally related to members of the HP1 family, it is not found in heterochromatin in Arabidopsis, but instead localises to silenced Polycomb target genes [149–151]. In contrast to classical HP1 proteins in other organisms, which bind H3K9-methylated histones, LHP1 binds histones methylated at H3K27. For this reason, LHP1 has been suggested to play a role similar to PRC1-component Polycomb (Pc) in Drosophila. However, the existence of a PRC1 'complex' in plants remains controversial (Sec. 1.2.1, p. 27). HP1 family proteins are characterised by a chromodomain which recognises methylated histones and a related chromoshadow domain involved in protein-protein interactions. These two domains are separated by a flexible 'hinge' region, which is less well conserved in both length and primary sequence.

In both the yeast and mammalian cases, the hinge region has been implicated in RNA-binding [146, 147]. However, the reported biolog-

ical roles of HP1 RNA-binding in the two systems differ. In human cells, RNA-binding affected localisation of HP1 to heterochromatin [146, 152], while in *S. pombe*, RNA-binding did not affect HP1 localisation or heterochromatin integrity but rather affected the ability of heterochromatin to repress transcription [147].

LHP1 recognises H3K27me3; is localised to Polycomb target genes; is required for maintenance of the repressed transcription state at *FLC*; interacts with PRC2 [153]; and has the potential to bind RNA. This protein therefore sits at the interface between transcription and epigenetics and therefore provides an opportunity to investigate how RNA-binding of a chromatin-associated protein contributes to maintenance of a specific chromatin state.

Molecular phenotypes at the *FLC* locus in plants lacking LHP1 are presented in Chapter 3, and Chapter 5 contains an investigation of LHP1 RNA-binding.

1.6 Patterns of cell growth and division in Arabidopsis

The biological system employed experimentally throughout this thesis is the plant *Arabidopsis thaliana*. A basic introduction to Arabidopsis development is therefore useful for understanding the requirements for epigenetic memory to persist through cell division in a growing plant.

Arabidopsis is predominantly self-pollinated [154]. Embryos arise from fertilised egg cells in seed pods known as siliques. Upon completion of embryogenesis, mature seed is formed and is released from siliques [155]. Germination requires a combination of moisture, temperature and light, resulting in outgrowth of vegetative tissues from the seed.

Meristems are the origin of all post-embryonic cell lineages and contain undifferentiated cells (Fig. 1.7). Both the shoot and root apical meristems are specified in the embryo [39], and contain actively dividing cells. These cell divisions replenish the meristem and also produce cells that give rise to the non-meristematic tissues of the plant. Cells

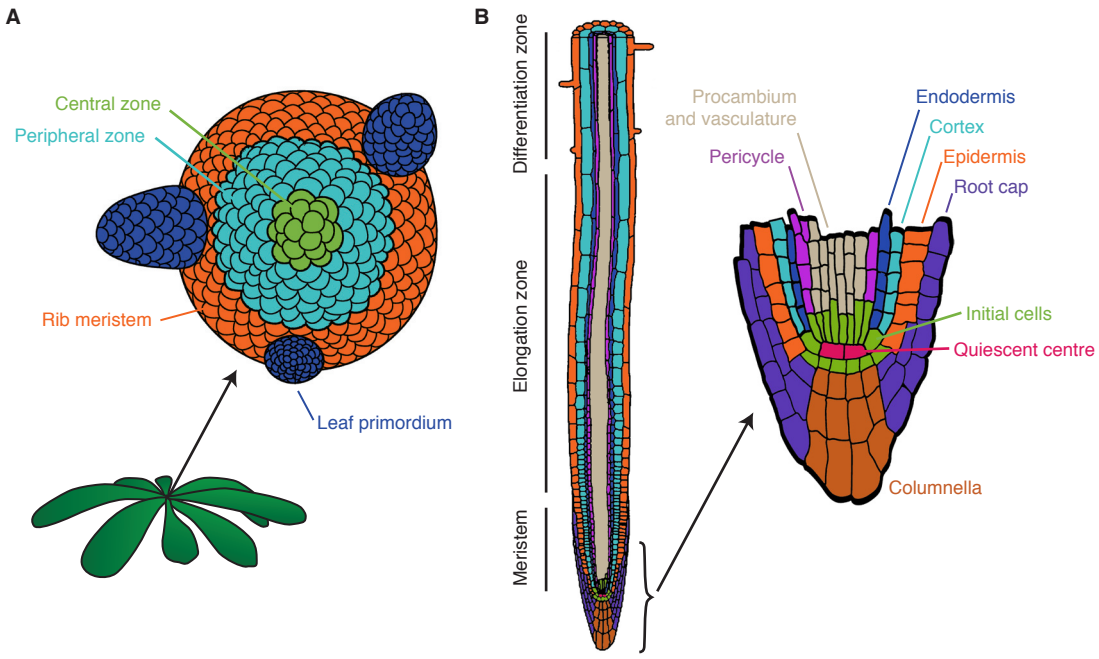


Figure 1.7: Apical meristems in Arabidopsis. (A) Schematic illustration of a vegetative shoot apical meristem, viewed from above. Cells in the central zone and peripheral zone are rapidly dividing. This displaces older cells, which move into the rib meristem or initiate lateral organs such as leaf primordia. Adapted from [39]. (B) Schematic illustration of a root showing the distinct cell layers. Initial cells in the root apical meristem are rapidly dividing and give rise to all other root cells through cell division. Adapted from [39, 160].

that are displaced from the meristem undergo only a few cell divisions before entering a non-canonical cell cycle in which DNA is replicated without mitotic cell division [156, 157]. These cell cycles are known as endocycles, and the phenomenon is called endoreduplication (or endoreplication). After several endocycles, cells cease DNA replication and remain in a quiescent state [158]. Leaf and stem epidermal cells typically endocycle only a few times before cell-cycle arrest and therefore contain 2-16 copies of each chromosome [159]. Nuclear DNA content in these tissues is correlated with cell size [159]. In contrast to endoreduplicated nuclei in differentiated cells, meristems contain diploid cells.

Because plant cells do not move with respect to their neighbours and cell division primarily occurs in meristems and developing primordia,

the position of cells in an organ is related to their 'age' (time since last division). 'Older' cells are displaced by 'newer' cells that arise through cell divisions at the apex. Older cells differentiate into root or shoot tissues as they become more distant from the apex [161]. Therefore, the average age of of cells increases as a function of distance from the apex [39].

1.6.1 *Shoot organisation*

The shoot apex comprises the apical meristem, the proximal meristem in which leaf primordia form, and a subapical region known as the rib meristem in which the shoot widens and primordia extend [162] (Fig. 1.7A). Within the apical meristem, a population of initial cells[f] in the 'central zone' give rise to the cells in the 'peripheral zone', through cell divisions [39]. These divisions also act to maintain the pool of undifferentiated initial cells in the central zone. Peripheral zone cells go on to generate leaf primordia and ultimately leaves. Therefore, the undifferentiated apical meristem initial cells can be thought of as the source of all leaf and stem tissue. Endoreduplication cycles commence as cells are displaced from the meristem [163].

1.6.2 *Root organisation*

Similarly to the shoot meristem, a population of undifferentiated initial cells in the root meristem ultimately gives rise to all cells of the root [160] (Fig. 1.7B). The centre of the root meristem is called the quiescent centre (QC), which contains several cells that have low mitotic activity in mature roots [164]. Rapidly-dividing diploid initial cells surround the QC on all sides. Lateral root cap and columnella cells are derived from initials on the root tip side of the QC, whereas the epidermis, cortex, and endodermis are derived from initials on the lateral and non-root-tip side of the QC. Finally, the pericycle and procambium cells that surround the central vasculature of the root are derived from

[f] Initial cells are analogous to stem cells in animals but the term 'stem' is avoided due to potential confusion with cells of the plant stem.

initial cells closer to the centre of the root [160] (Fig. 1.7B). With the exception of the pericycle and procambium, cells in the root undergo endoreduplication several cell divisions after being 'born' in the meristem [39, 160].

Further away from the QC, cells lose their ability to divide, and subsequently increase in length [165]. This is known as the elongation zone. It is in this zone that endoreduplication begins [166]. Further still from the stem cell niche is the differentiation zone, in which cells complete acquisition of cell-type-specific characteristics. Daughters of initial cells generate 'files' of cells that are clonally related [167]. These extend along the longitudinal axis of the root.

1.7 Conventions and nomenclature

1.7.1 *Genes, genotypes and alleles*

This thesis follows the Arabidopsis convention for genetic nomenclature. Taking *VERNALIZATION 1* for example, the wild-type gene name is designated using upper case italics *VERNALIZATION 1* (*VRN1*). In contrast, a mutant allele and also plants and genotypes containing this mutant allele are designated in lowercase, *vrn1*. Different mutant alleles of the same gene are designated by a number following the gene name, i.e. *vrn1-4*. Unless otherwise specified, the genetic background in which the original mutation was obtained is implicit in the numbering system. For example the *vrn1-4* allele is a mutation in *VRN1* in the Columbia Col-0 accession, whereas *vrn1-1* is a mutation in the *fca-1* background (Landsberg *erecta* (Ler-0) accession). Where an allele is introgressed into a different genetic background this is usually made explicit. One example is the introgression of active *FRIGIDA* from Sf-2 into Col-0 to generate Col-0 $FRI^{Sf\text{-}2}$. In this case, Col-0 $FRI^{Sf\text{-}2}$ is so commonly used that it is considered an accession in its own right and is therefore non-italicised and abbreviated to Col-FRI.

RNA is designated using the gene name, therefore is also in upper case italics (e.g. *VRN1*). Typically only abbreviated names are used to

designate transcripts. In contrast, proteins are upper case but plain text (e.g. VRN1).

Traditionally, genes in Arabidopsis have been named for their loss of function phenotypes (e.g. *CURLY LEAF* (*CLF*), *EARLY FLOWERING IN SHORT DAYS* (*EFS*), *SHOOT MERISTEMLESS* (*STM*)), however since the advent of molecular biology, they have sometimes also been named according to homology with proteins or domains with known function in other systems (e.g. *LIKE HETEROCHROMATIN PROTEIN 1* (*LHP1*), *ARABIDOPSIS TRITHORAX-RELATED 7* (*ATXR7*)).

1.7.2 *Transgenic plants*

Transgenic plants are used throughout this thesis. By analogy with the filial (F) generations F1, F2, etc., the generations of transgenic (T) plants are denoted T1, T2, etc. Transgenic plants are generated by transforming the flowers of a T0 plant [168]. The seed harvested from this generation gives rise to T1 plants. If transformation is successful, these contain a hemizygous transgene insertion, analogous to the heterozygosity of F1 plants. Normally, populations segregate for the transgene at the T2 generation, and homozygous individuals can be further propagated to give homozygous T3 populations, which are no longer segregating for the transgene.

1.7.3 *Histone modifications*

Discussion of the role of histone modifications forms a major part of this work. The Brno nomenclature for histone modifications [169] and the phylogeny-based nomenclature for histone variants [170] will be followed throughout the thesis. In this system, H3K4me3 represent histone H3 lysine 4 tri-methylation while H2BK123ub1 represents histone H2B lysine 123 mono-ubiquitination. In addition to the Brno system, the suffix me (e.g. H3K27me) will be used to indicate general methylation without distinction between mono-, di- and tri-methylation.

1.8 Outline of the thesis

Chapter 2 contains the results of experiments investigating whether *FLC* expression after vernalisation is maintained by a cis or trans epigenetic memory system. It is shown that the epigenetic memory is physically located in the local chromatin environment of the *FLC* gene. This demonstrates that a Polycomb target gene can store epigenetic memory in cis, and supports the hypothesis that H3K27me3 could act as a heritable determinant of gene expression. Following this chapter, detailed investigations of various aspects of the molecular mechanism underpinning this cis memory are presented. In Chapter 3, mathematical modelling of histone-modification-based memory is introduced. Modelling of the *FLC* nucleation region leads to two qualitatively distinct models for the chromatin dynamics in the nucleation region, each with experimentally testable predictions. In Chapter 4, the focus shifts to the role of transcription in the antagonism of Polycomb silencing. The mathematical formulation of a model of histone-modification-based epigenetic memory incorporating transcription leads to a conceptual synthesis of cis and trans epigenetic memory. Continuing the theme of interactions between transcription and chromatin, Chapter 5 returns to experimental studies, to examine interactions between LHP1 and RNA. After showing that LHP1 can bind to RNA *in vitro*, targeted mutagenesis is used to generate LHP1 mutants that separate the RNA- and H3K27me3-binding functions of this protein. This allows dissection of the role of these two functions *in vivo*. Finally, a general discussion in Chapter 6 summarises these results in the context of the field of chromatin-based epigenetics and contains an outlook for integrating the mathematical models of the *FLC* nucleation region (Chapter 3) with the generalised model of transcription and Polycomb silencing developed in Chapter 4.

Several pieces of experimental data were obtained by members of the Dean lab other than myself. This is acknowledged in each case when the corresponding data is presented.

Materials and methods are included at the end of each results chapter (Chapters 2-5). Where a method is used again in a subsequent chapter, a reference to the original description is included in the methods section of that chapter. Primers and antibodies used are listed in Chapter 7. Abbreviations and acronyms are defined where they are initially used, and a comprehensive list is provided for reference on p. 241. Arabidopsis gene numbers (At*xgxxxxx*) are also provided with the list of abbreviations.

EPIGENETIC MEMORY OF *FLC* EXPRESSION IS STORED IN CIS

2

FLC is repressed by prolonged cold exposure. This repression correlates with changes in chromatin structure at the *FLC* locus. In this chapter, it is shown that local changes to chromatin determine *FLC* expression after cold. Furthermore, it is the chromatin state that determines its own inheritance through mitotic cell divisions. Thus, the memory of winter cold is a cis memory.

The chapter begins with a description of the development and validation of a fluorescent reporter for *FLC* expression, *FLC-Venus*. This is used to examine *FLC* expression at the level of single cells in living plant tissue after different lengths of cold exposure. These experiments provide confirmation of the hypothesis that *FLC* expression states are bistable and mitotically inherited. The *FLC-Venus* reporter paves the way for the development of a second fluorescent reporter, *FLC-mCherry*, which can be distinguished from *FLC-Venus* by confocal microscopy. These distinguishable reporters are then combined in the same plants and used to show that the key inherited epigenetic memory elements that determine *FLC* expression are physically associated with the *FLC* gene itself.

2.1 Development of a fluorescent reporter for *FLC* expression

2.1.1 *Design of transgenic FLC constructs*

The key DNA sequence elements necessary for cold-induced *FLC* repression are believed to reside at the 5′ end of intron 1 [122, 171]. However, there is also evidence for other important regulatory sequences further downstream within intron 1 [172], and the antisense *COOLAIR* promoter occupies the region immediately downstream of the *FLC* poly-

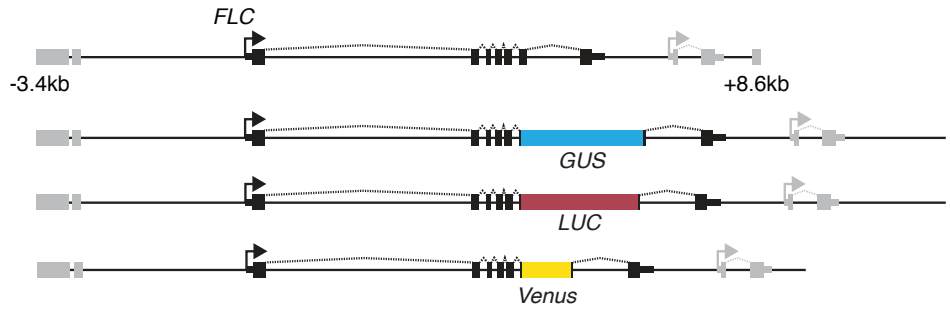


Figure 2.1: Schematic of *FLC* genomic DNA used to generate *FLC* reporters. Exons are represented by black boxes. Transgenes extend from 3.4 kb upstream to 8.6 kb downstream of the *FLC* transcription start site. Neighbouring genes are depicted in grey. In ease case, the fusion protein is inserted into exon 6.

adenylation site [84]. Previous work in the lab focused on developing both N- and C-terminal translational FLC fusion constructs with enhanced green fluorescent protein (eGFP). These attempts failed due to excessive transgene silencing or mis-regulation of the expressed fusion product[a]. This may be because the 5′ and 3′ regions of *FLC* contain important regulatory sequences, or unusual nucleic acid structures [173] that were disrupted in these transgenic constructs.

Prior to these attempts with N- and C-terminal FLC fusions, *FLC-GUS* [48] and *FLC-LUCIFERASE* (*FLC-LUC*) [174] reporters had been successfully constructed in the lab and both showed similar regulation in response to cold as endogenous *FLC*. These constructs used an internal NheI restriction site in *FLC* exon 6 to insert the coding sequence of the reporter (Fig. 2.1). When transformed into a *flc* mutant background, these constructs express a fusion protein that is not functional in delaying flowering, but nonetheless shows wild-type regulation [48, 174]. In fact, *FLC-LUC* was successfully used for genetic screens that identified several key *FLC* regulators [174].

Inspired by the success of these constructs, *FLC-Venus* was generated using a similar cloning strategy (Sec. 2.5.4). Venus is a fast-maturing yellow GFP-variant with a reported brightness 50% higher than eGFP

[a] Jie Song, personal communication.

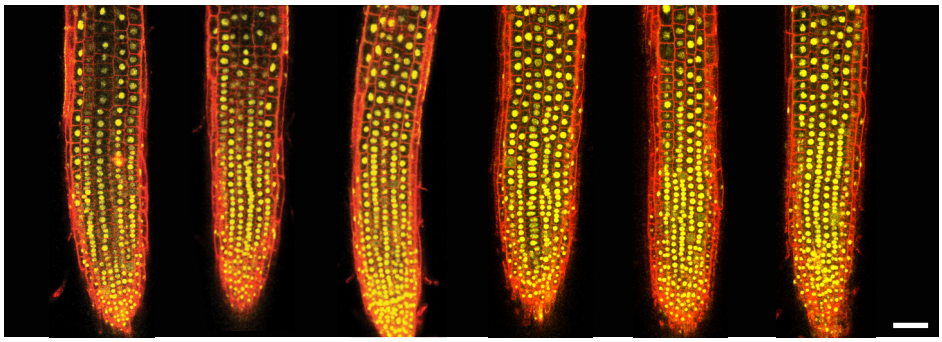


Figure 2.2: FLC-Venus expression in non-vernalised plants. FLC-Venus channel (yellow) is a maximum intensity projection over 2-3 z-planes. Propidium iodide (red) was used to stain the cell wall; a single z-plane corresponding to the centre of the projection was overlaid on the FLC-Venus image. 3 roots are shown for each of two independent *FLC-Venus* lines. Scale bars, 50 µm

[175]. Other than brightness, this fluorophore was chosen based on anecdotal evidence that it performs well in Arabidopsis tissues[b]. The *Venus* coding sequence was cloned by PCR and inserted into exon 6 (Sec. 2.5.4). This construct was transformed into an *flc* mutant background known as *flc-2*, which is early-flowering without vernalisation due to lack of functional *FLC* [40] (Sec. 2.5.4).

2.1.2 *Selection of transgenic lines.*

More than 50 T1 transgenic plants[c] were obtained. Surprisingly, the flowering time of these plants varied from very early (like parental *FRI flc-2*) to very late (like Col-FRI) (data not shown). This suggested that the *FLC-Venus* construct was capable of delaying flowering and therefore that the FLC-Venus fusion protein was functional. This was unexpected based on the previous lack of complementation observed for *FLC-GUS* and *FLC-LUC*.

Preliminary experiments indicated that FLC-Venus could be visualised in the roots of non-vernalised T2 *FLC-Venus* plants by confocal microscopy (Fig. 2.2). FLC-Venus showed nuclear localisation and reasonably uniform expression throughout the root. Visualisation was

[b] Jie Song, personal communication.
[c] See Sec. 1.7.2, p. 44 for a description of T1, T2, etc.

more difficult in the leaves, due to increased autofluorescence, however FLC-Venus was detected (data not shown).

It was observed by microscopy that FLC-Venus intensity varied between different transgenic lines. Such variability between different lines is typical for *FLC* transgenic plants[d]. In other studies, this variability has been overcome by using 'pools' of many different transgenic lines to extract average results [85, 122, 176]. For the present study, one or two individual lines which show regulation equivalent to endogenous *FLC* are required. Since flowering time is quantitatively related to *FLC* expression level in non-vernalised plants [40, 177], this provided a convenient assay with which to choose *FLC-Venus* lines showing a similar expression level to endogenous *FLC*. Flowering time data was recorded for T2 populations with and without vernalisation (Fig. 2.3A). In addition, a subset of lines were selected for measurement of transgene expression level by reverse-transcriptase quantitative PCR (RT-qPCR) (Fig. 2.3B,C). The following criteria were used to select *FLC-Venus* lines for further analysis:

1. The ability to delay flowering similarly to Col-FRI, and to accelerate flowering in response to cold. (Fig. 2.3A).

2. The consistency of flowering time results within a single-parent population (Fig. 2.3A).

3. Similar expression level to endogenous *FLC* in Col-FRI (Fig. 2.3B)

4. Repression of *FLC-Venus* at the mRNA level (Fig. 2.3C).

Based on these criteria, three lines were selected for further analysis: *FLC-Venus* 27, 33, and 45 (Fig. 2.3). Although these lines all showed 3:1 segregation for herbicide resistance in T2 populations (indicating single-locus transgene insertion), it was later determined using a qPCR-based copy number assay that only lines 27 and 33 contain a single copy

[d] C. Dean, J. Questa, P. Li, personal communication.

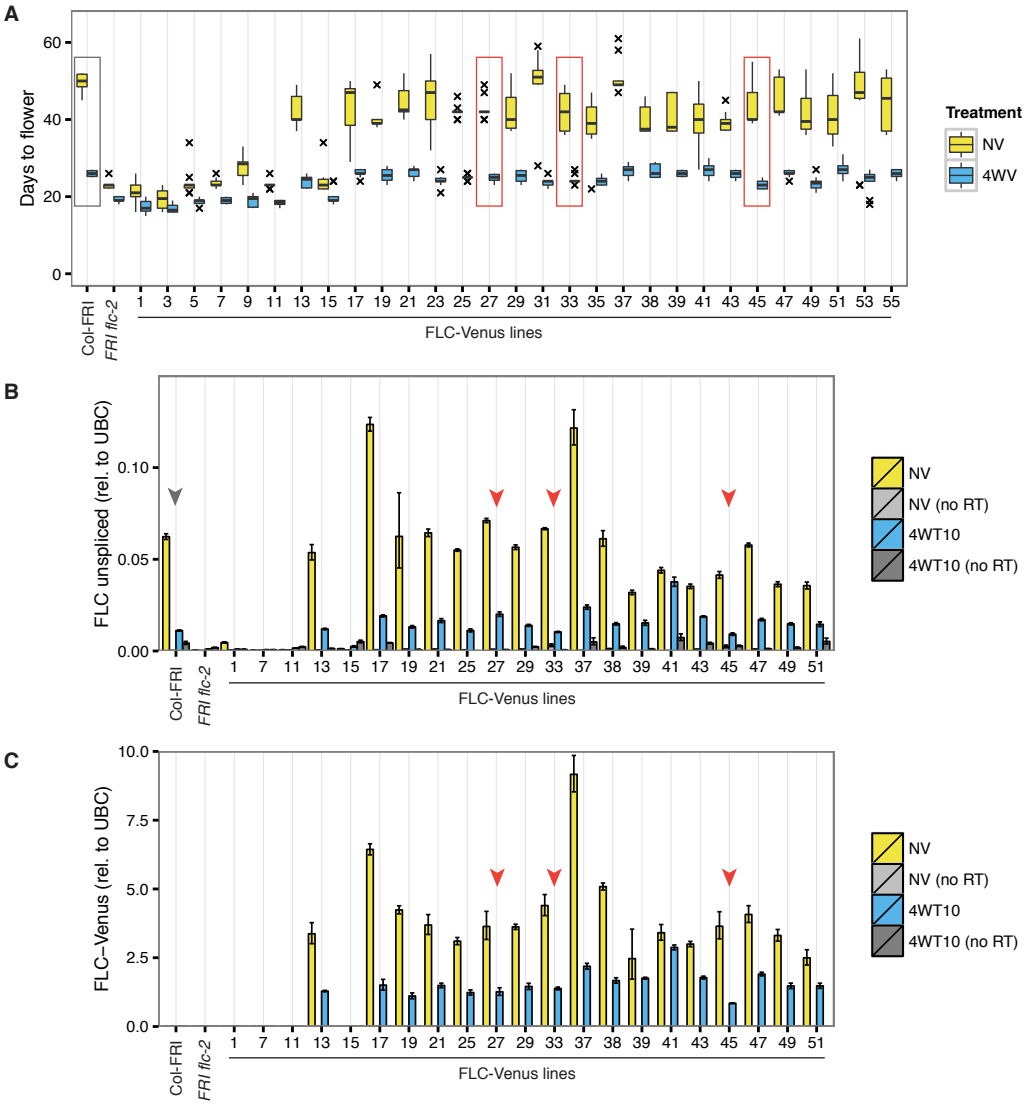


Figure 2.3: Characterisation of *FLC-Venus* transgenic lines. (A) Flowering time for T2 *FLC-Venus* plants compared to Col-FRI and parental *FRI flc-2* plants. Transgenic plants were isolated from T2 populations using a selectable herbicide-resistance marker on plates before transfer to soil. Vernalised plants were pre-grown for 1 week at 22°C and spent 4 weeks at 5°C. Glasshouse growth conditions (Norwich, April-July 2013). (B) Unspliced *FLC* and (C) *FLC-Venus* expression relative to control gene *UBC* for a selection of *FLC-Venus* lines (T2 generation), as quantified by RT-qPCR (Sec. 2.5.6). 4WT10 indicates samples were harvested 10 days after a 4-week cold treatment. cDNA samples prepared without reverse-transcriptase are labelled 'no RT' and represent the low-level contamination with genomic DNA. Red boxes/arrowheads represent lines selected for further experiments. Error bars represent standard deviation of technical qPCR replicates (n = 3) from a single RNA sample.

of *FLC-Venus*, while line 45 contains 3 copies[e] (Sec. 2.5.4). All results presented hereafter are for the two single-copy lines, 27 and 33.

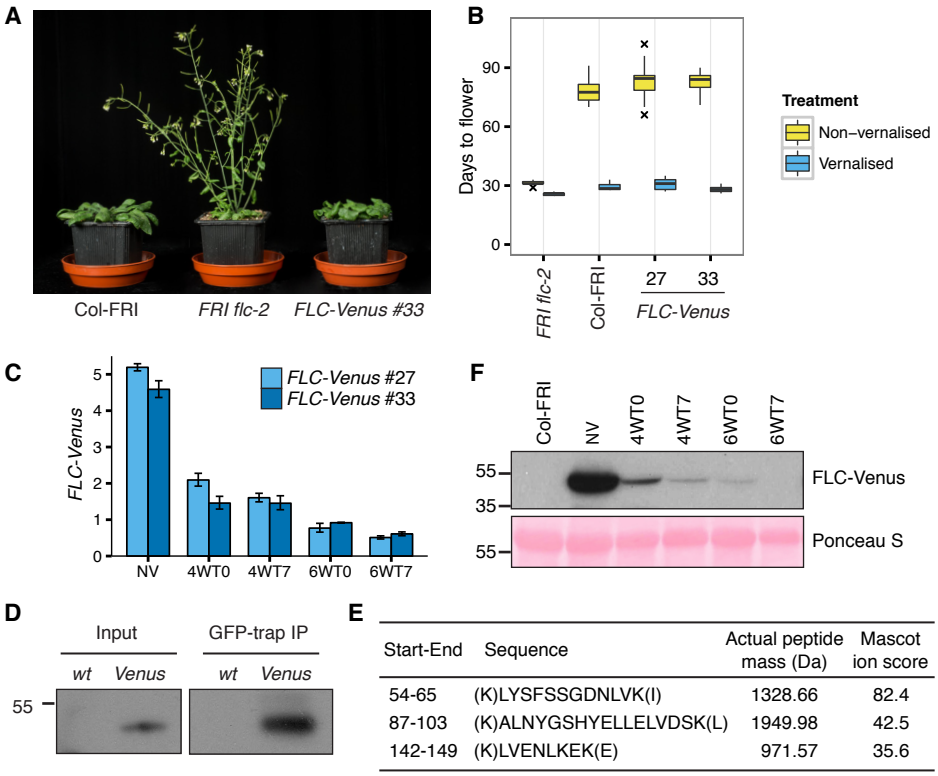
2.1.3 *Validation of FLC-Venus function and regulation*

Measurement of flowering time in controlled growth conditions was performed for the selected transgenic lines, after obtaining populations homozygous for *FLC-Venus*. Both lines showed similar flowering time to wild-type Col-FRI plants before and after cold treatment (Fig. 2.4A,B). To confirm that the *FLC-Venus* transgene was stably repressed in vernalised plants, *FLC-Venus* expression in homozygous T3 plants was analysed after a 4- or 6-week cold treatment. Stable repression following cold was observed, as was the quantitative dependence of expression level on cold duration (Fig. 2.4C).

Venus signal was observed in the nuclei of *FLC-Venus* plants by confocal microscopy (Fig. 2.2). Because free Venus protein is not normally nuclear-localised, this indicated that Venus was fused to FLC. To confirm that full-length FLC-Venus protein was expressed in these plants, immunoprecipitation (IP) was performed with GFP-trap beads and proteins were visualised by immunoblot using an anti-GFP antibody (Fig. 2.4E,F). A specific band at the expected size (48.8kDa) indicated that the fusion protein observed by microscopy was indeed full-length FLC-Venus. Like *FLC-Venus* RNA, immunoblots indicated that FLC-Venus protein showed stable epigenetic repression that depended on the duration of cold exposure (Fig. 2.4F).

GFP-trap pulldown samples were also analysed by mass spectrometry[f] and several peptides from the FLC N-terminus were identified. Among the interacting proteins identified by mass spectrometry was SHORT VEGETATIVE PHASE (SVP), a previously reported interactor of FLC [178]. Together with the rescue of the early-flowering *flc-2* mutant phenotype (Fig. 2.4A,B), these results suggest that FLC-Venus

[e] qPCR-based copy number assay for the BASTA-resistance gene, *bar*, performed by iDNA genetics
[f] Mass spectrometry performed by Gerhard Saalbach at the John Innes Centre proteomics facility.



Start-End	Sequence	Actual peptide mass (Da)	Mascot ion score
54-65	(K)LYSFSSGDNLVK(I)	1328.66	82.4
87-103	(K)ALNYGSHYELLELVDSK(L)	1949.98	42.5
142-149	(K)LVENLKEK(E)	971.57	35.6

Figure 2.4: Validation of selected *FLC-Venus* lines. (A) Photograph showing the early-flowering phenotype of non-vernalised (NV) parental *FRI flc-2* plants and the complementation of *flc-2* by *FLC-Venus*. (B) Flowering time for homozygous single-copy *FLC-Venus* plants compared to wild-type (Col-FRI) and parental *FRI flc-2* plants ($n = 20$) (Sec. 2.5.2). Vernalised plants were pre-grown for 1 week at 22°C and spent 4 weeks at 5°C before being returned to 22°C. Unlike Fig. 2.3A, these plants are homozygous single-copy *FLC-Venus* lines at the T3 generation. (C) *FLC-Venus* mRNA levels measured by RT-qPCR using primers specific for *FLC-Venus*. Data are normalized to the control gene *UBC* (Sec. 2.5.6). 4WT0 indicates that plants were harvested immediately after exposure to 4 weeks of cold, while 4WT7 indicates that plants were grown for a further 7 days in warm conditions. Data are represented as mean ± s.e.m. ($n = 3$). (D) Immunoblot of FLC-Venus GFP-trap immunoprecipitation (IP). *wt* indicates untransformed wild-type control sample (Col-FRI), whereas *Venus* indicates *FLC-Venus* transgenic sample. Input blot is from the same membrane as the IP sample except using a longer exposure time. Mass of protein size marker shown in kDa. The predicted mass of FLC-Venus is 48.8 kDa. (E) Table of peptides corresponding to FLC that were identified by mass spectrometry in FLC-Venus IP but not control IP sample. Start-End refers to the location of the peptide in the FLC protein sequence. (F) Immunoblot of FLC-Venus during and after 4 or 6 weeks cold exposure. Mass of protein size marker shown in kDa. Membrane stained with Ponceau S is shown as a loading control.

is able to assemble into the same repressor complex as endogenous FLC and, in doing so, repress floral activators such as *FT* and *SOC1* [178].

2.2 Vernalisation-induced *FLC* repression is cell-autonomous and heritable

As discussed in Sec. 1.2.1, a previous study using an *FLC-GUS* reporter indicated that *FLC* is not repressed in all cells after short cold exposures [83]. In roots, the pattern of cells that express *FLC-GUS* after cold differs between plants, indicating that cells stochastically switch to the *FLC*-OFF expression state during cold and remain in this state after cold. It was also previously observed that longer cold treatments are capable of switching *FLC* off in all cells [83]. This suggests that the number of *FLC*-OFF cells after cold depends quantitatively on the length of cold exposure [83]. The GUS staining system has several drawbacks: Quantification is difficult, samples are fixed, and not all parts of a tissue can be effectively assayed. Consequently, these analyses were limited to a small section of the upper root meristem, and conclusions regarding the stability of *FLC* expression states and their heritability through mitotic cell division were not possible.

To obtain a more complete picture of *FLC* expression at the level of individual cells, *FLC-Venus* plants were exposed to 2, 4, 6, 8, or 10 weeks of cold followed by 7 days further growth in warm conditions. In agreement with the FLC-GUS results, it was observed by confocal imaging of FLC-Venus in root meristems that the number of cells expressing *FLC* appeared to decrease with the length of cold exposure (Fig. 2.5, 2.6, 2.7). To quantitatively characterize the FLC-Venus expression status in individual cells after different durations of cold, an automated image analysis procedure was developed to calculate the mean FLC-Venus intensity inside each cell[g] (Sec. 2.5.10). As suggested by individual images, analysis of 63 roots (9870 reconstructed cells) revealed that the number of cells expressing FLC-Venus decreased quantitatively with the duration

[g] Image analysis algorithm developed by Matthew Hartley and Tjelvar Olsson.

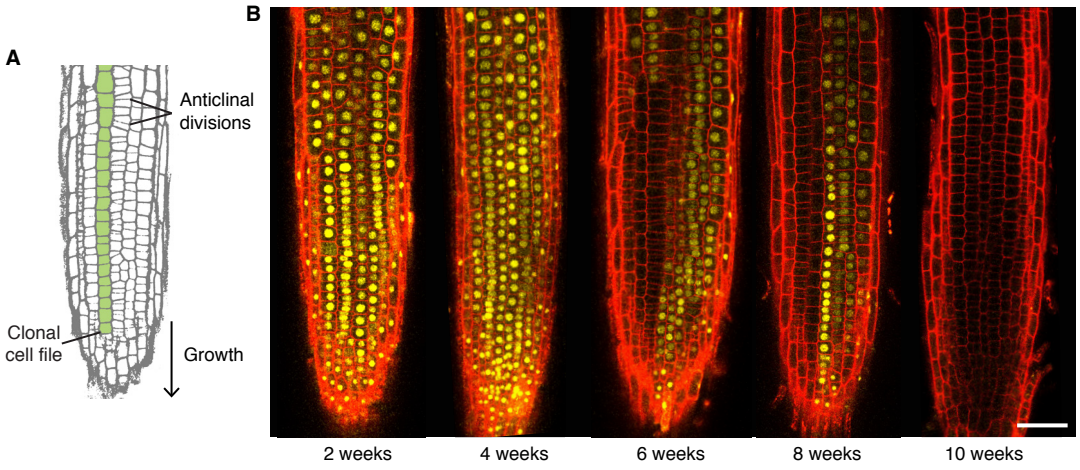


Figure 2.5: *FLC-Venus* expression states are mitotically inherited. (A) Schematic of an Arabidopsis root meristem showing how repeated anticlinal cell divisions give rise to clonal cell files along the axis of growth. (B) FLC-Venus (yellow) in root meristems of plants exposed to 2, 4, 6, 8, or 10 weeks cold followed by 7 days of warm conditions. Plants were imaged 7 days after return to warm. FLC-Venus channel is a maximum intensity projection over 2-3 z-planes. Propidium iodide (red) was used to stain the cell wall; a single z-plane corresponding to the centre of the projection was overlaid on the FLC-Venus image. Scale bars, 50 μm. Further imaging details in Sec. 2.5.10.

of cold exposure (Fig. 2.8). Importantly, FLC-Venus expression was bimodal after 6 or 8 weeks of cold, and almost all cells were silenced after 10 weeks of cold (Fig. 2.8).

Roots provide an excellent system for studying heritability of epigenetic states because cell lineages are visible as continuous files of cells that arise from repeated anticlinal divisions along the longitudinal axis of the growing root [164] (Fig. 2.5A). In warm conditions, stem cells at the root tip divide approximately once every two days, while for other meristematic cells this occurs approximately once per day [179]. Each cell undergoes several divisions before reaching the elongation zone whereupon cell division no longer takes place and cells begin endoreduplication[h]. Therefore, 7 days after plants are transferred from cold, a single cell in the stem cell niche will have given rise to a lineage that encompasses a long file of cells in the root meristem (Fig. 2.5A), thereby

[h] See Sec. 1.6.2 for an introduction to Arabidopsis root growth and cellular organisation.

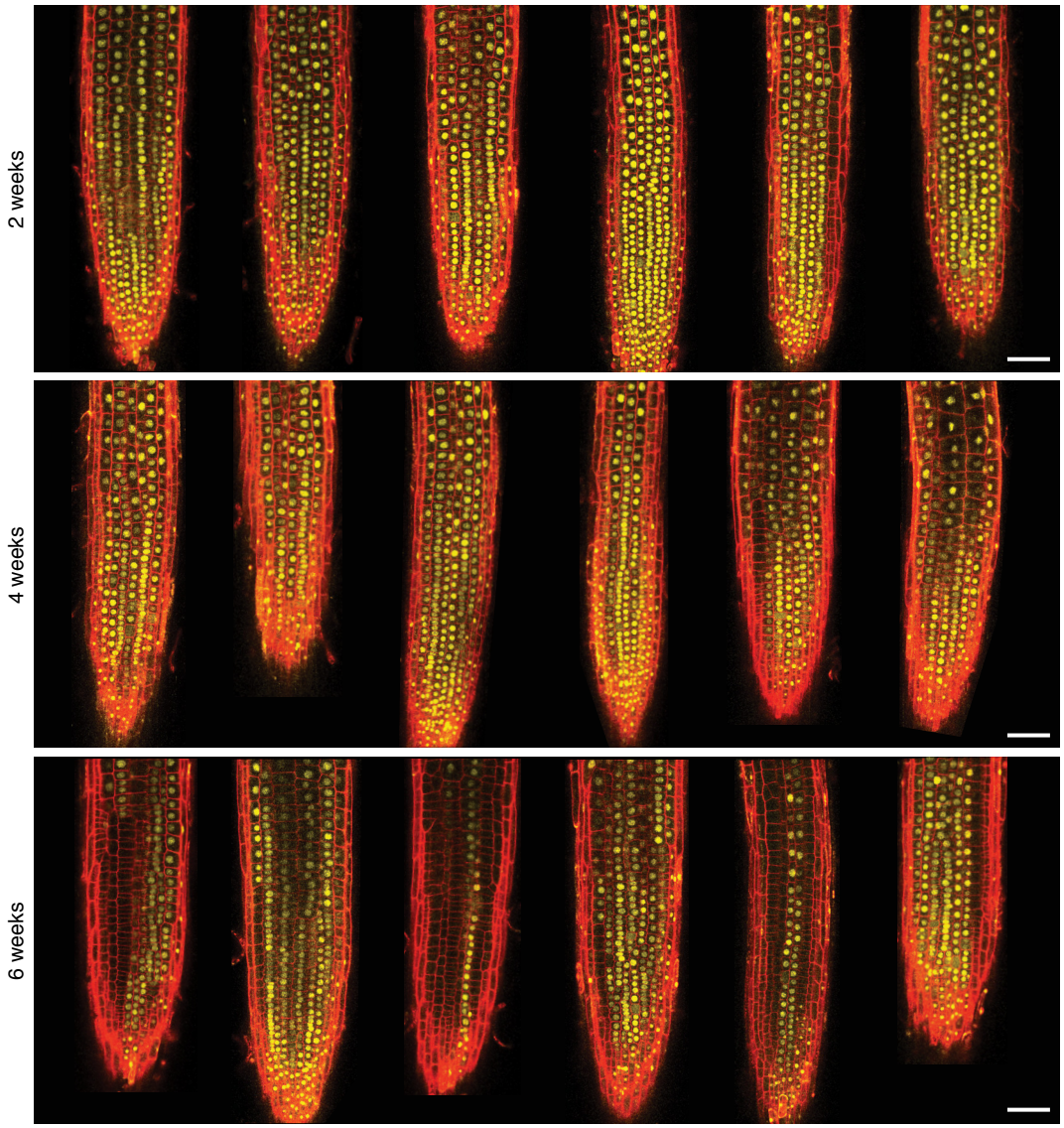


Figure 2.6: Confocal microscopy reveals long-term heritability of FLC-Venus expression states. FLC-Venus (yellow) in root meristems of plants exposed to 2, 4, or 6 weeks cold. (8 and 10 weeks shown in Fig. 2.7). For each treatment, 3 roots are shown for each of two independent *FLC-Venus* lines. Imaging details in Fig. 2.5 caption and Sec. 2.5.10. Scale bars, 50 μm.

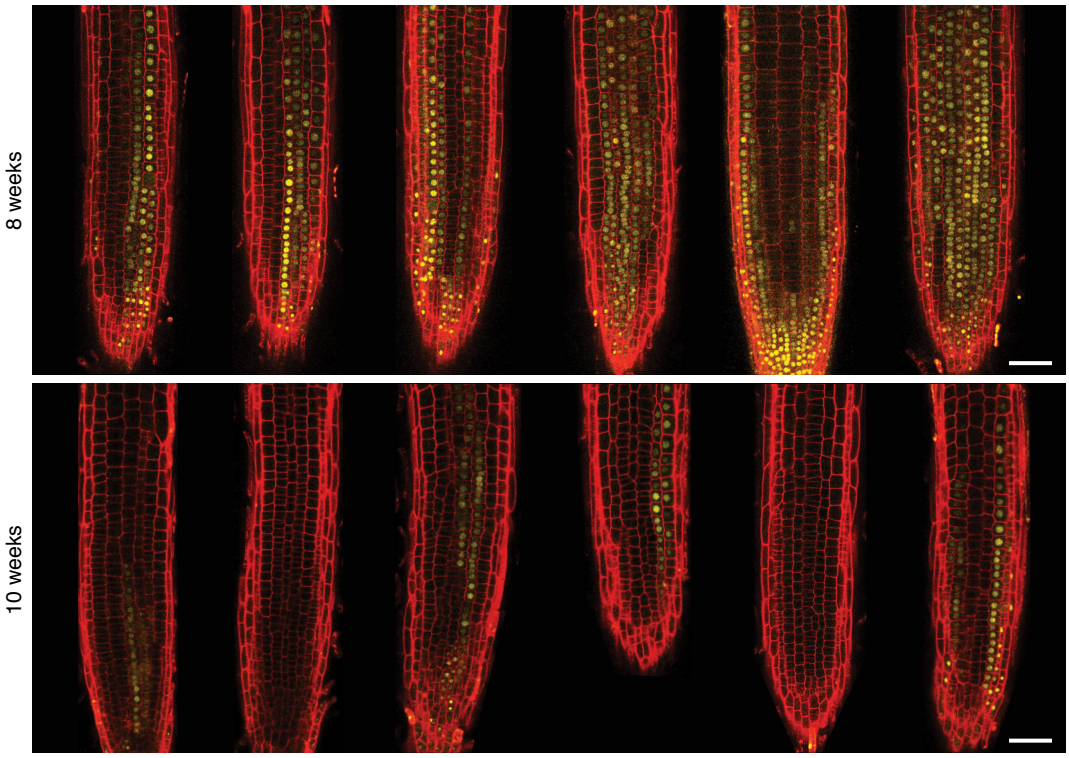


Figure 2.7: Confocal microscopy reveals long-term heritability of FLC-Venus expression states. FLC-Venus (yellow) in root meristems of plants exposed to 8, or 10 weeks cold. (2, 4 and 6 weeks shown in Fig 2.6). For each treatment, 3 roots are shown for each of two independent *FLC-Venus* lines. Imaging details in Fig. 2.5 caption and Sec. 2.5.10. Scale bars, 50 µm.

allowing a direct assay of the mitotic stability of the epigenetic state *in planta*. Furthermore, *FLC* expression in roots is quantitatively related to the length of prior cold exposure at the cell population level (Fig. 2.8B) and therefore provides a representative tissue to study *FLC* regulation after vernalisation. Strikingly, images of FLC-Venus taken 7 days after cold exposure showed long files of cells in the same expression state (Fig. 2.5B, 2.6, 2.7), demonstrating the long-term mitotic stability of the active and repressed transcriptional states. Such files make it implausible that the observed expression levels in single cells are the result of bursts of transcription. Instead, files indicate heritability of *FLC* expression states.

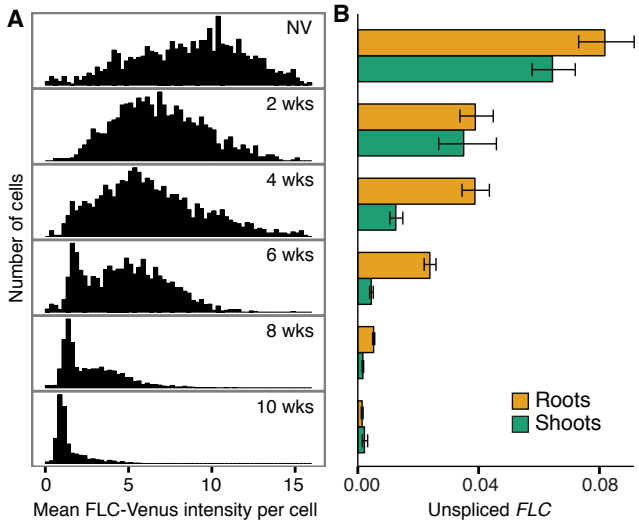


Figure 2.8: Quantification of FLC-Venus. (A) Histograms of mean FLC-Venus intensity in individual cells quantified from microscopy data for *FLC-Venus* transgenic lines 27 and 33. Each panel summarizes data from confocal z-stacks for 8-12 roots (1372-2067 cells). Quantitative image analysis is described in Sec. 2.5.10 (p. 80). (B) Unspliced *FLC-Venus* RNA in roots and shoots measured by qRT-PCR for non-vernalised (NV) plants or after 2, 4, 6, 8 or 10 weeks of cold followed by 7 days of warm. Data shown are mean ± s.e.m. for at least 2 biological replicates for each of 2 independent transgenic lines (n≥ 4).

2.3 Epigenetic memory is stored in cis

As discussed in Sec. 1.3, epigenetic memory can be stored in concentrations of diffusible factors (trans memory), or may also be stored in the local chromatin environment of the regulated gene, perhaps in patterns of histone modifications. Images of FLC-Venus expression patterns in partially-vernalised roots demonstrate that *FLC* expression states are bistable and heritable, but do not address whether epigenetic memory is stored in cis or trans. This section describes experiments to distinguish between these two possibilities.

Trans-acting factors identified from forward genetic screens for loss of *FLC* repression, such as *VRN1*, *VRN2* and *VRN5* show equal expression before and after cold exposure [45–47]. This suggests that none of these factors act in a bistable trans-regulatory network to regulate

FLC expression. However, genetic screens cannot exclude all possible factors because neither essential genes nor networks with multiple layers of redundancy can be accessed. It is therefore impossible to rule out trans memory using forward genetics. The requirement for PHD-PRC2 in epigenetic memory [45, 47, 49, 80, 81] is also not proof that chromatin actually stores epigenetic memory - a trans memory system may simply require PHD-PRC2 to cause gene repression (see 'responsive chromatin', Fig. 1.5, p. 32).

2.3.1 *Preliminary evidence for cis memory*

A common feature of trans memory systems is that the target gene product feeds back into the regulatory network, providing information about its expression state [12, 88]. For this reason it is interesting to consider whether *FLC* expression depends on FLC protein levels.

***FLC* is semi-dominant.** Early genetic studies indicated that the *FLC* gene acts in a semi-dominant manner to delay flowering [40, 177]. That is, plants that are heterozygous for *FLC* show flowering time intermediate to that of early-flowering *flc* mutants and late-flowering wild-type plants. It is also known that flowering time in non-vernalising conditions is quantitatively related to the level of *FLC* expression [40, 177, 180–184]. In the context of cis and trans memory, semi-dominance of *FLC* suggests that expression of *FLC* in plants containing one functional *FLC* copy is lower than in plants containing two copies. Therefore, a single functional *FLC* copy does not increase its expression to compensate for the loss of the other copy in heterozygous plants. This argues against a trans-acting feedback loop that maintains *FLC* expression in non-vernalising conditions.

FLC protein is not required for stable repression of *FLC* transcription. To further investigate whether *FLC* repression during vernalisation involves trans-regulation from *FLC* protein, the *flc-3* mutation was used. *flc-3* contains a 104 bp deletion in *FLC* exon 1, resulting in removal of

the annotated start codon [36]. *FRI flc-3* plants are very early flowering, consistent with absence of functional *FLC* protein [36]. *FLC* expression was analysed by RT-qPCR in *FRI flc-3* before and after a 6-week cold treatment. While *FLC* mRNA levels were ~10-fold lower in *flc-3* than in Col-FRI at all time points, unspliced *FLC* RNA levels were similar between the mutant and wild-type (Fig. 2.9A,B). Unspliced *FLC* RNA has previously been used to quantify *FLC* transcription [85, 121, 184–188]. These results indicate that levels of transcription at *FLC* are similar in Col-FRI and *FRI flc-3*, despite the partial exon 1 deletion in *flc-3*. Reduced mRNA stability for *FLC* RNA in *flc-3* could explain the failure to accumulate *FLC* mRNA. This may be due to mRNA surveillance mechanisms, such as nonsense-mediated decay, degrading the $FLC^{flc\text{-}3}$ transcript, which lacks an ATG. Despite the lack of functional *FLC* in *flc-3*, stable silencing of *FLC* transcription was observed for 7 days after a 6-week cold treatment (Fig. 2.9B).

To confirm that chromatin changes associated with vernalisation are unperturbed in the *flc-3* mutant, H3K27me3 levels were assayed by chromatin immunoprecipitation followed by qPCR (ChIP-qPCR). Consistent with expression analysis, *FLC* chromatin in *flc-3* showed nucleation of H3K27me3 during cold exposure and subsequent spreading to cover the locus, similar to wild-type (Fig. 2.9C).

H3K36me3 is commonly associated with actively transcribed genes in higher eukaryotes and shows a progressive decrease in the nucleation region during vernalisation in Col-FRI [123]. Consistent with unperturbed repression of *FLC* transcription in *flc-3*, it was also observed that H3K36me3 was progressively lost from *FLC* during and after vernalisation (Fig. 2.9C).

Together, these data indicate that epigenetic silencing of *FLC* does not require functional FLC protein. This argues against a trans regulatory feedback that involves FLC protein as part of the *FLC* repression mechanism after cold.

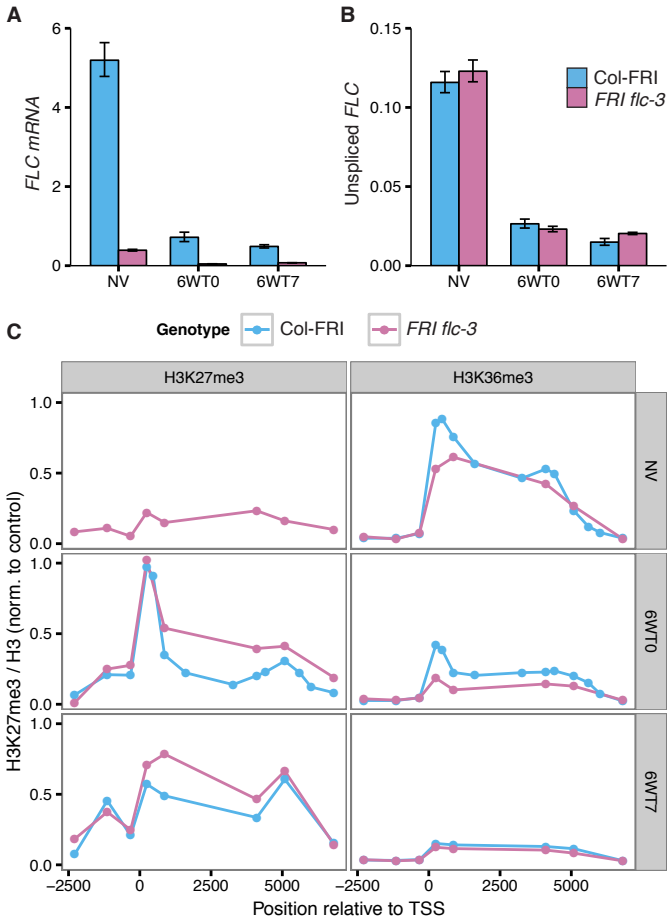


Figure 2.9: *FLC* expression and chromatin in *flc-3*. (A) Spliced *FLC* and (B) Unspliced *FLC* levels relative to *UBC* in *FRI flc-3* or wild-type Col-FRI plants, as measured by RT-qPCR. Plants were either non-vernalised (NV), harvested immediately after 6 weeks of cold treatment (6WT0), or grown for a further 7 days in warm conditions after cold (6WT7). Error bars represent standard error between biological samples ($n = 3$). (C) H3K27me3 and H3K36me3 ChIP during the same vernalisation experiment. H3K27me3 is normalised to the control gene *STM*, while H3K36me3 is normalised to the control gene *ACTIN* (Sec. 2.5.9). Note that H3K27me3 ChIP failed for the non-vernalised Col-FRI sample.

***FLC*$^{\text{Col-0}}$ but not *FLC*$^{\text{Lov-1}}$ alleles stably silence in Lov-1.** If *FLC* transcription states are defined by concentrations of diffusible factors (trans memory), then all copies of *FLC* within a cell should be in the same transcription state over long time scales. On the other hand, if *FLC* chromatin states instruct their own inheritance (cis memory), then it may be possible for different copies of *FLC* within a cell to exist in different heritable transcriptional states (Sec. 1.3). Indeed, observation of differential expression of two copies of a gene in the same cell is the primary evidence that memory is stored in cis for the case of genomic imprinting [6].

To search for evidence of this 'mixed' (*FLC*-ON / *FLC*-OFF) transcriptional state, the Arabidopsis accession Lov-1 was used. In Lov-1, *FLC* is repressed by cold exposure, however unlike the reference accession Col-0, Lov-1 shows slow, partial reactivation of *FLC* expression between 10 and 30 days after cold [189]. To determine if the Col-0 and Lov-1 *FLC* alleles would be differentially regulated in the same plants, a Col-0 *FLC-GUS* transgene (*FLC*$^{\text{Col}}$*-GUS*) was transformed into Lov-1 and Col-FRI[i].

As discussed above, a trans memory system common to Lov-1 and Col-0 *FLC* would result in expression of *FLC*$^{\text{Col}}$*-GUS* in all cells that express Lov-1 *FLC* because chromatin would be purely responsive to the concentration of the diffusible regulatory factors. In Col-FRI, it was observed that endogenous *FLC* and the *FLC*$^{\text{Col}}$*-GUS* transgene were both similarly repressed after a 10-week cold treatment. However, in the Lov-1 background, endogenous *FLC*$^{\text{Lov-1}}$ reactivated significantly faster than *FLC*$^{\text{Col}}$*-GUS* (Fig. 2.10, $p < 10^{-14}$)[j].

The reactivation of Lov-1 *FLC* likely reflects individual cells reverting from an *FLC*-OFF to an *FLC*-ON state after cold exposure. The difference in reactivation rates between the two alleles therefore suggests that *FLC*$^{\text{Col}}$*-GUS* can remain repressed in the same cells as the endoge-

[i] Transgenic plants generated by Amy Strange for a previous unpublished study.

[j] Rates of reactivation were obtained by linear regression of log-transformed expression data between 10 and 30 days after cold. These rates were compared using a two-tailed t-test with unpooled variance estimates.

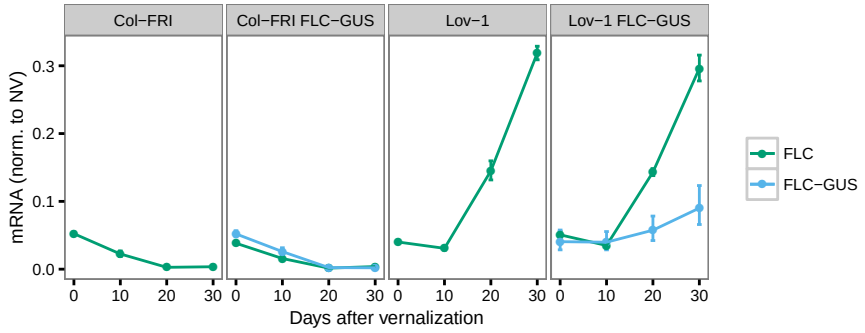


Figure 2.10: *FLC*^{Col}*-GUS* and endogenous *FLC* levels after vernalisation. Spliced mRNA levels of endogenous *FLC* and an *FLC*^{Col}*-GUS* transgene measured by RT-qPCR in transformed and untransformed Col-FRI and Lov-1 plants after a 10-week cold treatment. To normalize for differences in mRNA produced per locus and possible differences in mRNA stability between the endogenous locus and the *FLC*^{Col}*-GUS* transgene, expression data were normalized to non-vernalised levels. Data shown are mean ± s.e.m. for 3 biological replicates. For Lov-1 transformed with *FLC*^{Col}*-GUS*, data shown is further averaged over the 3 independent transgenic lines analysed (n = 9).

nous Lov-1 *FLC* allele is activated. That is, *FLC*^{Lov} in Lov-1 can switch from OFF to ON after cold exposure, while *FLC*^{Col}*-GUS* remains OFF in the same cell. This suggests that 'mixed' (*FLC*-ON / *FLC*-OFF) expression states can exist for cases in which there are polymorphisms between the two *FLC* alleles.

2.3.2 *Dual reporter assay to distinguish cis and trans memory*

Results presented so far suggest that FLC protein is not part of a trans-feedback mechanism that regulates its own expression. Furthermore, it was shown that one copy of *FLC* can remain in a repressed state while another copy in the same cell is expressed —when sequence polymorphisms exist between the two copies. However, at this stage it cannot be excluded that *FLC* repression is maintained by a bistable trans memory that acts independently on Col-0 and Lov-1 *FLC*, as these alleles differ in several sequence polymorphisms [122].

The ideal experiment would be to show that two identical copies of *FLC* in the same cell can be in different, heritable expression states over long periods of time. Confocal imaging of *FLC-Venus* lines after cold exposure (Sec. 2.2) allows visualisation of mitotically inherited *FLC* ex-

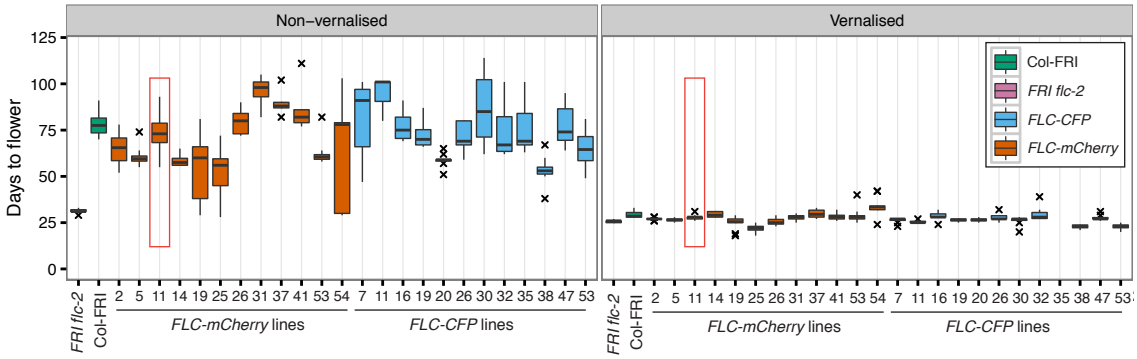


Figure 2.11: Characterisation of *FLC-mCherry* and *FLC-CFP* lines. Flowering time for T2 *FLC-mCherry* or *FLC-CFP* plants compared to Col-FRI and parental *FRI flc-2*, either non-vernalised (NV) or after a 5-week cold treatment (Sec. 2.5.2). Transgenic plants were isolated from T2 populations using a selectable herbicide-resistance marker on plates before transfer to soil. Vernalised plants were pre-grown for 1 week at 22°C and spent 5 weeks at 5°C before being returned to 22°C. Single-copy *FLC-mCherry* line 11 used for subsequent studies is highlighted with red box.

pression states at the single-cell level. To extend this assay to allow two copies of *FLC* in the same cells to be distinguished, a second fluorescent reporter for *FLC* expression was developed. The two reporters differ only in their fluorophore sequence and therefore this experiment approaches the ideal case.

Development of *FLC-mCherry* and *FLC-CFP*. *FLC-mCherry* and *FLC-CFP* lines were developed using an identical strategy to that used for *FLC-Venus*. Like *FLC-Venus*, both of these fusion proteins rescued the early-flowering phenotype of parental *FRI flc-2* plants (Fig. 2.11). FLC-CFP was not visible by confocal microscopy (data not shown) and no further experiments were performed with these lines. FLC-mCherry showed a similar pattern of expression to FLC-Venus both in non-vernalised and vernalised plants (Fig. 2.12A-C). The flowering time of a single-copy *FLC-mCherry* line agreed well with *FLC-Venus* and wild-type Col-FRI plants (Fig. 2.12D).

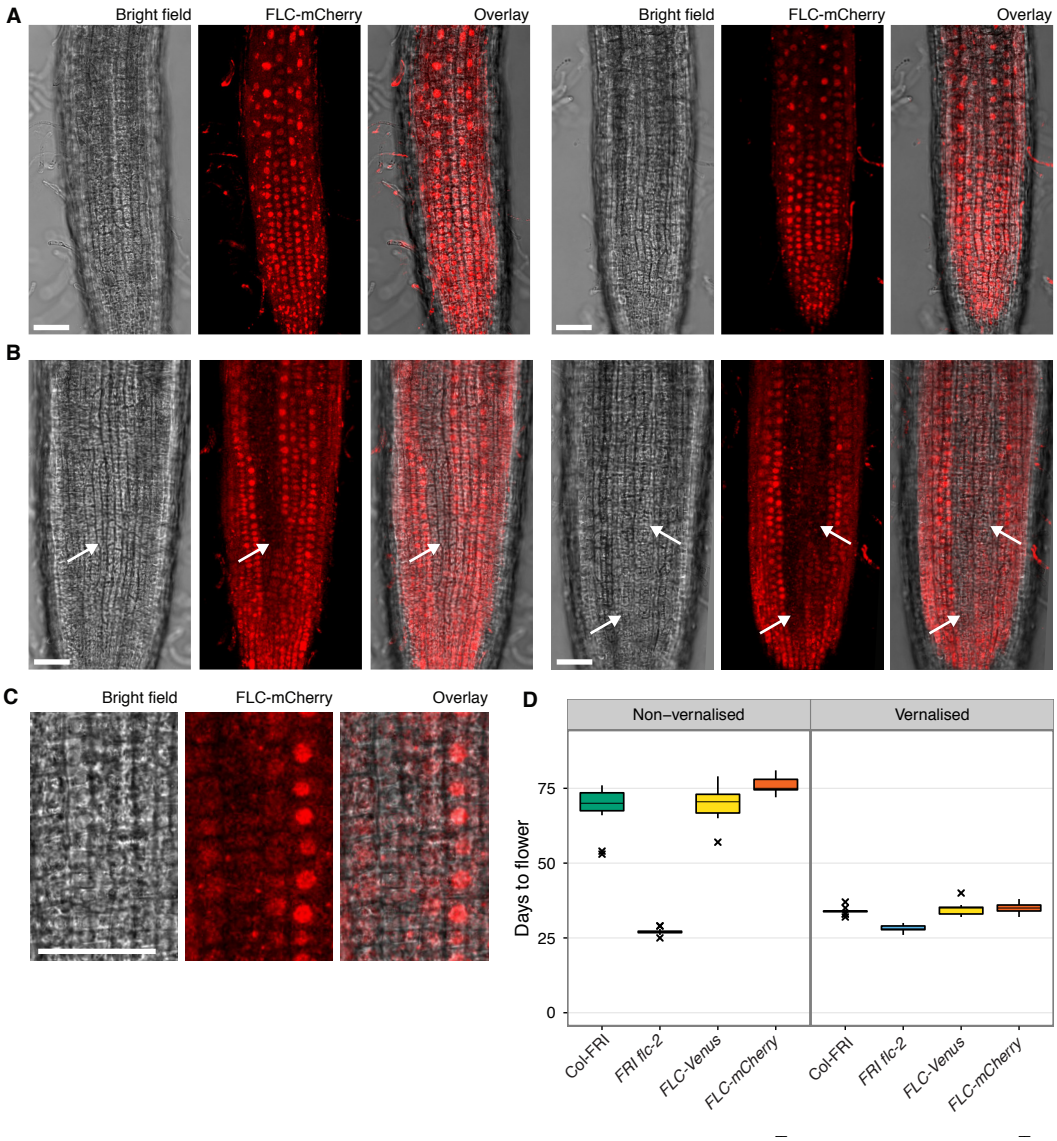


Figure 2.12: Validation of *FLC-mCherry*. (A) Confocal microscopy images of FLC-mCherry in roots of non-vernalised plants or (B) in plants grown for 6 weeks at 5°C, followed by 7 days at 22°C. Images are maximum intensity projections over 2-3 z-planes; a single z-plane corresponding to the centre of the projection was used for bright field images. Files of cells not expressing FLC-mCherry are indicated with white arrows. The cell wall stain propidium iodide is not compatible with the mCherry fluorophore due to overlap of the excitation and emission spectra, and was therefore omitted. (C) Higher magnification images showing adjacent files of ON and OFF cells for FLC-mCherry, 7 days after a 5-week cold exposure. Scale bars, 50 µm. (D) Flowering time for *FLC-mCherry* compared to parental *FRI flc-2*, wild-type Col-FRI and *FLC-Venus* (n = 12); vernalised plants were pre-grown for 1 week at 22°C and spent 4 weeks at 5°C before being returned to 22°C.

2.3.3 *Visualisation of 'mixed' expression states*

FLC-mCherry and *FLC-Venus* plants were crossed to generate F1 hybrids carrying a single copy of each transgene. In a trans-based memory, the only possible heritable expression states of FLC-Venus/FLC-mCherry are ON/ON and OFF/OFF, because epigenetic information is stored as a diffusible signal (Fig. 2.13: trans memory). In a cis-based memory, all four states ON/ON, ON/OFF, OFF/ON, and OFF/OFF are possible because the information is stored at the locus itself (Fig. 2.13: cis memory). As expected, non-vernalised (NV) roots showed uniform ON/ON expression of both transgenes in all cells (Figs. 2.14). To stochastically induce repression of the two *FLC* transgenes, F1 plants were exposed to 4 to 6 weeks of cold followed by 7 days growth in warm. Strikingly, this cold treatment generated long files of cells in which one *FLC* reporter was stably repressed, while the other remained stably activated (Figs. 2.14, 2.15). In fact, long-term mitotic stability of all four possible combinations: FLC-Venus/FLC-mCherry ON/ON, ON/OFF, OFF/ON, and OFF/OFF, was observed. Inheritance of the 'mixed' ON/OFF and OFF/ON states directly contradicts trans memory and instead provides direct evidence of a cis-encoded epigenetic state at *FLC*.

The *FLC-Venus* and *FLC-mCherry* transgenes in these double-hemizygous F1 plants are not in the same genomic location. However, like endogenous *FLC*, the reporters are actively expressed throughout development in warm conditions and epigenetically repressed specifically in response to prolonged cold. Epigenetic repression, therefore, depends on cold exposure and the *FLC*-specific sequences within the transgenes rather than the genomic location. Furthermore, since similar numbers of cells were observed in the FLC-Venus/FLC-mCherry ON/OFF and OFF/ON expression states, it is unlikely that differences in genomic location or fluorophore sequence between the transgenes causes one of the copies to be preferentially repressed over the other.

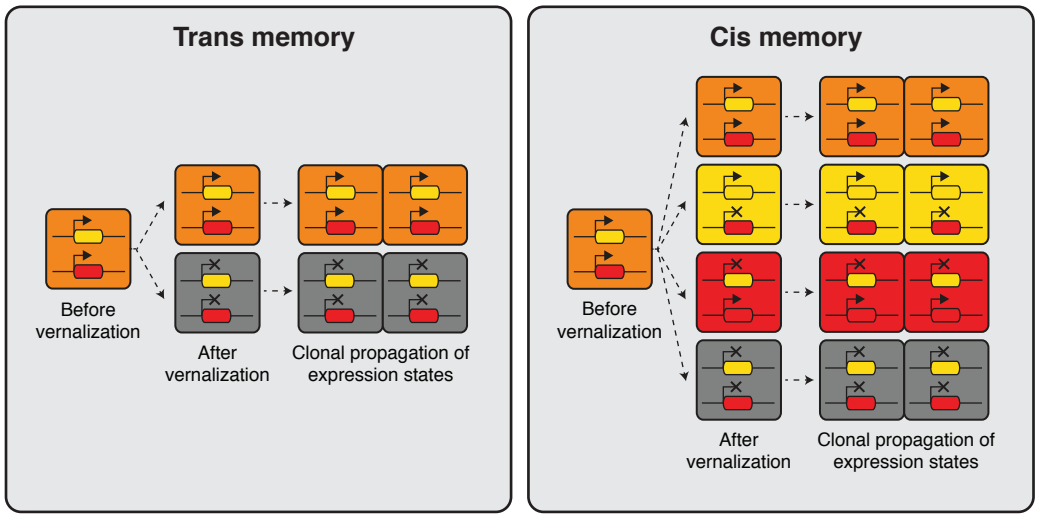


Figure 2.13: Distinguishing cis and trans memory. In a trans memory system, the two copies of *FLC* are coordinately regulated and only two mitotically heritable states are possible (FLC-Venus/FLC-mCherry ON/ON, OFF/OFF). In a cis memory system, the two copies of *FLC* can be maintained in alternative expression states, so four mitotically heritable states are possible (FLC-Venus/FLC-mCherry ON/ON, ON/OFF, OFF/ON, OFF/OFF).

2.4 Summary

Prolonged cold exposure induces a cell-autonomous switch in *FLC* expression from an ON to an OFF state. It was shown that these *FLC* expression states are then mitotically inherited through many cell divisions in growing plants (Sec. 2.2). The semi-dominant nature of *FLC* and the *flc-3* mutant, which disrupts FLC protein but not transcription, suggests that FLC does not feedback to regulate its own expression (Sec. 2.3.1). This argues against a trans-regulatory mechanism involving FLC protein. A Col-0 *FLC* allele can remain repressed in Lov-1 plants, while the endogenous Lov-1 *FLC* allele reactivates (Sec. 2.3.1). This suggests that the repressed chromatin state at *FLC* can be maintained at one but not both of these alleles in cis. Finally, an assay to visualise the expression state of two copies of *FLC in vivo* was developed. This allowed direct observation of cells containing an active *FLC* copy and a repressed *FLC* copy, without differences in regulatory sequence. These 'mixed' expression states were mitotically inherited (Sec. 2.3.3).

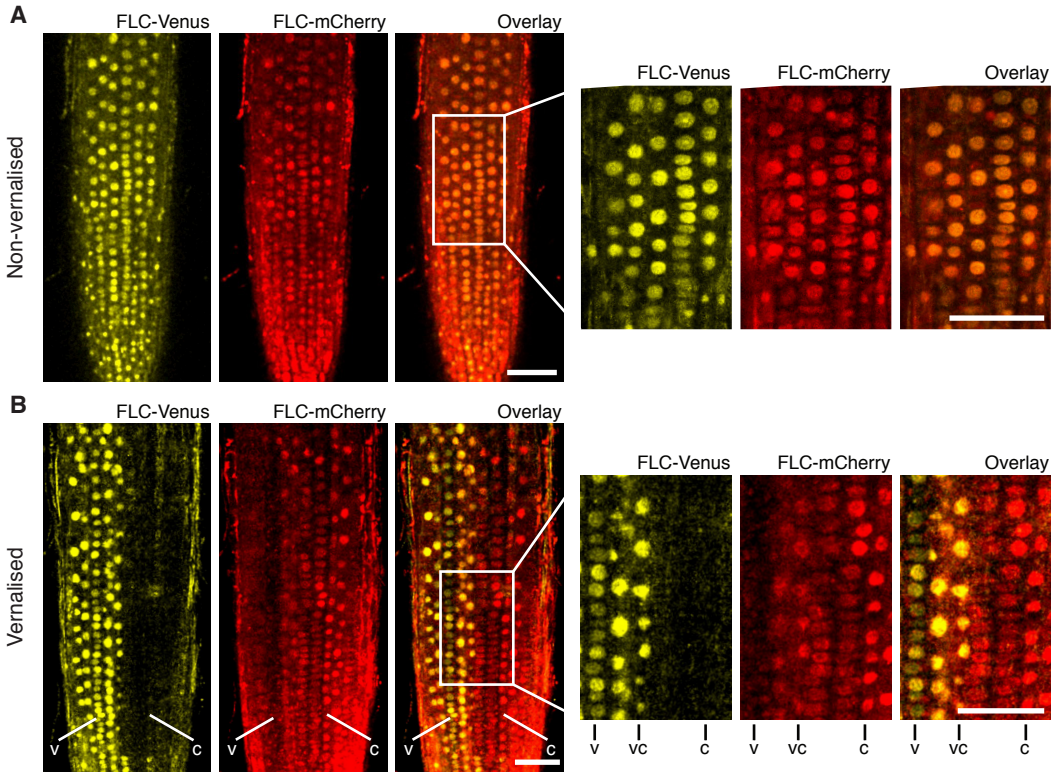


Figure 2.14: 'Mixed' transcriptional states are mitotically inherited. (A) Non-vernalised roots of *FLC-mCherry* × *FLC-Venus* F1 plants show uniform expression of FLC-Venus and FLC-mCherry in all nuclei. (B) After vernalisation, such plants can epigenetically repress a single gene copy while the other remains active. Images are maximum intensity projections over 2-4 z-planes, further details are provided in Sec. 2.5.10. The following notation is used to indicate files of cells in the various expression states: Both expressed, vc; FLC-Venus only, v; FLC-mCherry only, c. Scale bars, 50 µm.

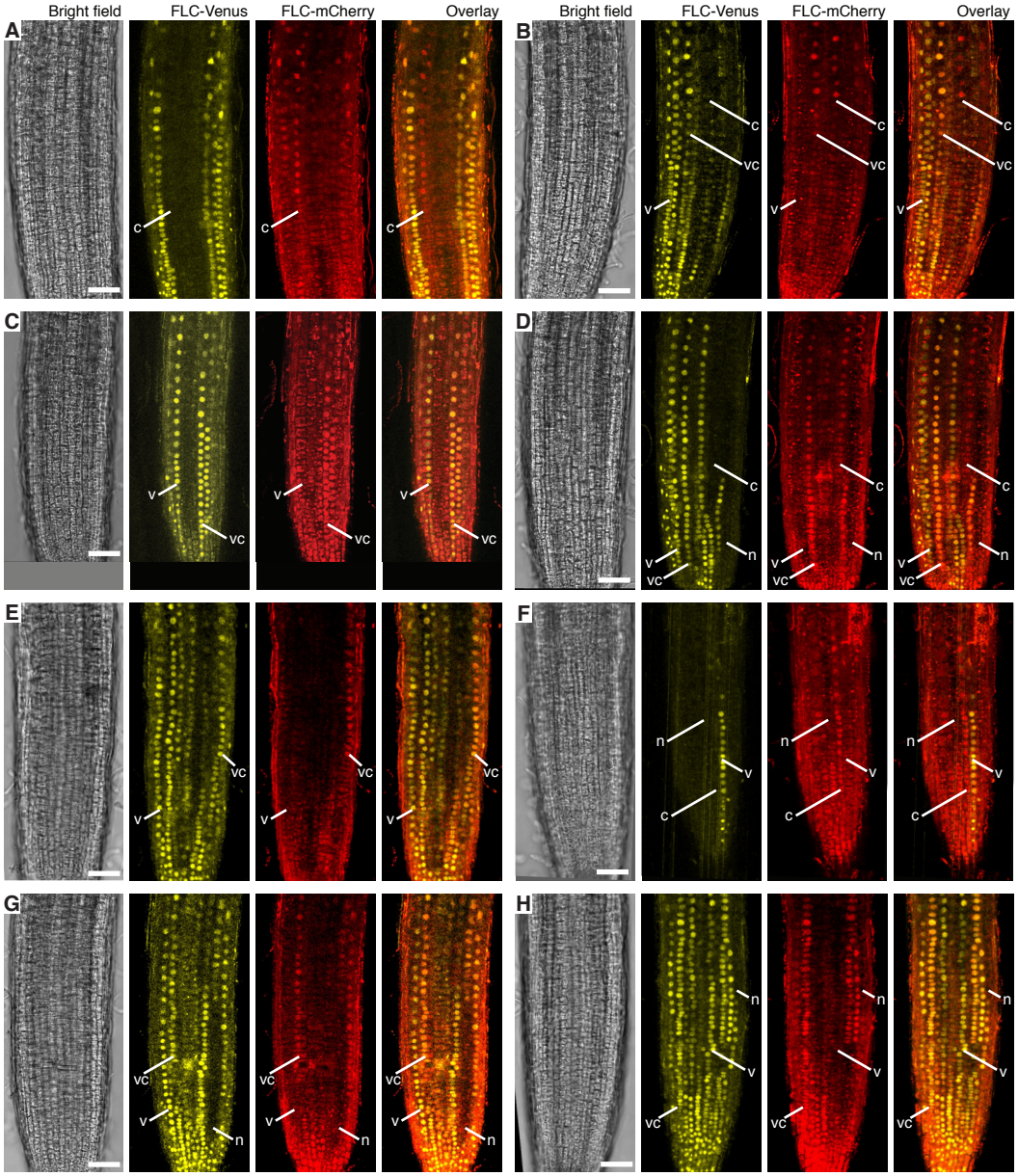


Figure 2.15: Confocal images of vernalised FLC-Venus × FLC-mCherry F1 plants. Confocal microscope images acquired simultaneously using dual excitation of FLC-Venus and FLC-mCherry. Images are maximum intensity projections over 2-4 z-planes, further details are provided in Sec. 2.5.10. A single bright-field image corresponding to the centre of the projection is also shown. Overlay contains the FLC-Venus (yellow) and FLC-mCherry (red) channels only. Using the notation 4WT10 to indicate that plants were treated with 4 weeks of cold followed by 10 days of growth in warm, vernalisation treatments are: (A) 4WT10, (B) 5WT10, (C) 5WT11, (D-H) 6WT7. Files of cells showing mitotically heritable expression of the different combinations of FLC-Venus/FLC-mCherry are identified by white lines. The following notation is used to indicate files of cells in the various expression states: Both expressed, vc; FLC-Venus only, v; FLC-mCherry only, c; both repressed, n. Scale bars, 50 µm.

The molecular changes to the chromatin environment of *FLC* induced by prolonged cold exposure are therefore sufficient to instruct epigenetic inheritance of the repressed transcriptional state. Together with the requirement for PRC2 in maintenance of the *FLC*-repressed state [45, 48, 49], this supports the hypothesis that H3K27me acts as a key component of this cis epigenetic memory. A discussion of the broader implications of this result for the field of epigenetics is left until Chapter 6.

Having demonstrated that cold-induced *FLC*-repression constitutes a Polycomb-dependent cis epigenetic memory, the following chapters focus on understanding the mechanistic basis of this memory.

2.5 Materials and methods

2.5.1 *Plant growth conditions*

For non-vernalising conditions, seeds were surface sterilized and sown on Murashige and Skoog (MS) media plates (no glucose), stratified at 4°C for 2 days, and grown for 10-14 days in long-day conditions (16 h light, 8 h darkness at 20°C). For vernalised plants, seeds were surface sterilized and sown on MS media plates (no glucose), pre-grown for 7 days, and then transferred to 5°C under short-day conditions (8 h light, 16 h darkness) before being returned to long-day conditions (16 h light, 8 h darkness at 20°C).

Plants harvested 0 - 10 days after vernalisation were grown on plates. Plants harvested > 10 days after vernalisation were transferred to soil 4-5 days after the end of 5°C treatment. For flowering time or extended non-vernalising growth conditions, plants were transferred to soil after 10 days on plates. Plants on soil were grown under long-day conditions (16 h light at 22°C, 8 h darkness at 20°C).

For confocal microscopy of roots, plants were grown almost vertically on MS media plates supplemented with 1% (w/v) sucrose and 0.5% (w/v) Phytagel (Sigma-Aldrich, P8169).

2.5.2 *Flowering time*

Flowering time was measured by counting days from sowing until bolting, defined as the time at which the primary floral meristem emerged from the rosette to a height of 1 cm. Unless indicated otherwise, flowering time does not include days spent in vernalising conditions.

2.5.3 *Plant materials*

Wild-type and *flc* mutants. Col-0 contains a non-functional allele of *FRIGIDA* and therefore shows low *FLC* expression and early flowering. The Columbia line *FRI*-Sf2 (Col-FRI) is Col-0 with an active *FRIGIDA* allele introgressed from the San Feliu-2 (Sf-2) accession [40]. The *flc-2* and *flc-3* mutants in Col-FRI were originally generated by fast-neutron mutagenesis [36]. *flc-3* is a 104-bp deletion in exon 1 which includes the start codon. *flc-2* is a larger chromosomal rearrangement, which resulted in the deletion of the entire 5′ half of the *FLC* locus.

***FLC-GUS* in Lov-1.** The Col-0 *FLC-GUS* translational fusion construct was generated during an earlier work [48]. A schematic of the transgene is shown in Fig. 2.1. *FLC-GUS* was transformed into Lov-1 for a previous study by Amy Strange. For the present study, 4 independent transgenic lines at the T3 generation were obtained from Amy's seed stocks. All T3 populations were found to be segregating for the transgene, so resistant plants were propagated to the T4 generation. Non-segregating T4 populations, each from a single parental plant, were used in the experiments described in the present study.

***FLC-Venus* and *FLC-mCherry*.** FLC-Venus and FLC-mCherry translational fusions were constructed for this study (Sec. 2.5.4, using a similar strategy to the *FLC-GUS* construct [48]. A schematic of the *FLC-Venus* transgene is shown in Fig. 2.1. These constructs were transformed into *FRI flc-2* plants using *Agrobacterium tumefaciens* (*A. tumefaciens*) (Sec. 2.5.4)). More than 50 lines were obtained for each construct.

2.5.4 *Cloning, transformation and selection of transgenic lines*

Plasmid minipreps and midipreps were performed using reagents and spin columns purchased from QIAGEN or Promega. Restriction enzymes were purchased from New England Biolabs (NEB) or Thermo Scientific.

To generate FLC-Venus, FLC-CFP and FLC-mCherry plasmids, either the *Venus* [175], *CFP* [190], or *mCherry* [191] coding sequence was amplified by PCR using primers with over-hanging NheI sites. PCR was performed with KOD Hot Start DNA Polymerase (Merck-Millipore). NheI-digested PCR products were cloned into similarly digested pBluescript FLC15[k] using T4 DNA ligase (NEB). pBluescript FLC-Venus, FLC-CFP and FLC-mCherry were confirmed by test digest and sequencing. 12.7 kb DNA fragments containing the genomic FLC fusions were then transferred into pSLJ-755I6 [192] using SacI/XhoI digestion, gel extraction, and ligation, to generate the plant binary vectors pSLJ FLC-Venus, pSLJ FLC-CFP and pSLJ FLC-mCherry.

Preparation of chemically competent *E. coli*. DH5α or TOP10 (Invitrogen) *E. coli* strains were used for cloning. The following protocol was adapted from [193]. 10 mL Lysogeny broth (LB) was inoculated with a single colony and incubated at 37°C for 6-8 hours, shaking at 200 rpm. This starter culture was then subcultured 1/50 into 250 mL LB and grown for 16-20 hours at 22°C, shaking at 200 rpm. When OD_{600} reached 0.55, the flask was chilled on ice, and cells were collected by centrifugation (2500 × *g*, 10 min, 4°C) using pre-cooled centrifuge and pots. Cells were then resuspended by swirling in 80 mL ice-cold Inoue buffer (10mM PIPES pH 6.7, 55 mM $MnCl_2$, 15 mM $CaCl_2$, 250 mM KCl). Cells were again collected by centrifugation and resuspended as before, this time in 20 mL buffer. After adding 1.5 mL DMSO and resting on ice for 10 min, cells were aliquoted (50 µL) into pre-chilled tubes, and flash-frozen on dry ice.

[k] Generated for a previous study by Peijin Li.

***E. coli* transformation (Heat-shock).** Competent cells (50 µL) stored at -80°C were thawed on ice for 30 min and 1-5 µL DNA was added. Tubes were incubated on ice for 30 min, mixing periodically by flicking. Tubes were then heat-shocked in a 42°C water bath for 30 seconds and returned to ice to recover for 2 min. Room temperature LB (1 mL) was added and tubes were transferred to 37°C for 1 hour, shaking at 200 rpm. Cells were collected by centrifugation (3000 × *g*, 5 min) and plated on LB media plates containing antibiotics for selection of transformants. Plates were incubated for 16-20 hours at 37°C and single colonies re-streaked before preparing glycerol stocks and plasmids.

***A. tumefaciens* transformation (Triparental mating).** To transfer the >25kB pSLJ755i6-based plasmids from *E. coli* to *A. tumefaciens*, triparental mating was used [194]. Tetracycline-resistant DH5α strains carrying the desired pSLJ-based constructs were cultured in liquid LB media. The Kanamycin-resistant HB101 *E. coli* strain carrying the helper plasmid pRK2013 [195] was also cultured in a separate tube. 0.2 mL log-phase culture of both of these strains was then added to 0.8 mL saturated culture of *A. tumefaciens* strain GV2260 (C58 background) [195]. Bacteria were then mated by growing this mixture at 30°C on LB media plates for 16 hours. *A. tumefaciens* carrying the pSLJ-based binary vector were selected on LB media plates containing 200 µg/mL Rifampicin and 1 µg/mL Tetracycline. Transformants were re-streaked on fresh plates and confirmed by colony PCR.

Arabidopsis transformation. The following transformation method was modified from [168]. Single colonies from Agrobacterium transformation or restreaked Agrobacterium glycerol stocks were used to inoculate 10 mL LB media containing 200 µg/mL Rifampicin and 1 µg/mL Tetracycline. These cultures were incubated for 36-48 hours at 30°C, shaking at 200 rpm. Starter cultures were then subcultured 1/40 into 200 mL LB with antibiotics and grown for a further 14-16 hours, shak-

ing at 200 rpm. Cells were collected by centrifugation (3000 × *g*, 4°C, 15 min) and pellets rinsed briefly with 5% sucrose. Cells were then resuspended in 150 mL buffer containing 5% sucrose, 0.02% SILWET L-77 (De Sangosse Ltd.) Suspended cultures were transferred to trigger spray bottles and sprayed directly onto Arabidopsis flowers. This process was typically repeated 3 times at 4 day intervals during flowering. Once dried, seeds were harvested and sown on soil for selection.

Selection of transgenic Arabidopsis. pSLJ755i6 [192] contains the bialaphos resistance gene, *bar* from *Streptomyces hygroscopicus* [196]. *bar* provides resistance to the herbicide Phosphinothricin (PPT), an inhibitor of glutamine synthetase in plants [197]. The trade name of the herbicide is Basta (Bayer). For selection of T1 transgenic plants, 0.25% (v/v) Basta was applied to T1 seedlings 3 times at 4 day intervals between 7-21 days after germination. For selection of T2 and T3 populations, PPT was included in MS media plates at a concentration of 10 µg/mL.

Identification of single-copy transgenic lines. Lines containing single copy transgene insertions were identified using a qPCR-based assay adapted for Arabidopsis from a previously described method [198]. This was performed by IDna Genetics. Primer and probe sequences are listed in Sec. 7.2.

2.5.5 *RNA extraction*

Total RNA was prepared as previously described [199], with modifications. Specifically, 0.1-0.3 g plant tissue was ground in liquid nitrogen. A mix of 250 µL phenol pH 4.3 (P4682, Sigma) and 500 µL homogenisation buffer (100 mM Tris pH 8.0, 5 mM EDTA, 100 mM NaCl, 0.5 % SDS) was pre-warmed to 60°C and added directly to ground plant tissue. Samples were then shaken at 60°C for 15 min before adding 250 µL chloroform and shaking at room temperature for a further 10 min. After centrifugation, the aqueous layer was isolated and again extracted with 500 µL phenol/chloroform/isoamyl alcohol pH 8.0 (25:24:1, Sigma,

P2069). Nucleic acids were then precipitated using isopropanol, and pellets resuspended in H_2O. To precipitate RNA away from genomic DNA, an equal volume of 4 M LiCl was added, and samples were incubated at 4°C for 16 hours. RNA pellets were obtained by centrifugation (20 000 × *g*, 4°C, 30 min) and were washed in 70% ethanol before resuspension in H_2O. RNA concentration was quantified using a Nanodrop 1000 (Thermo-Scientific) and quality was monitored by agarose gel electrophoresis.

Where required (e.g. for *FLC* unspliced measurements), contaminating genomic DNA was removed using TURBO DNA-free (Ambion, AM1907) following the manufacturer's guidelines, except that phenol-chloroform extraction and ethanol precipitation was used to further purify RNA after DNAse treatment.

2.5.6 *RT-qPCR*

Reverse transcription was performed using the SuperScript III First-strand synthesis system (Invitrogen, 18080-051), according to the manufacturer's protocol using either gene-specific primers or Oligo(dT)12-18 (Invitrogen, 18418-012). qPCR was performed using LightCycler 480 SYBR Green I Master (Roche, 04887352001) on the LightCycler 480 instrument (Roche). The reaction volume was 14 µL and thermal cycling conditions were 95°C, 5 min followed by 50 cycles of (95°C, 15 s; 60°C 25 s; 72°C, 30 s). Melting curves were calculated by heating from 40°C to 97°C at a rate of 0.11°C/s.

Threshold cycle (C_t) values were calculated using the 'Second Derivative Maximum method' in the LightCycler software. RNA levels relative to UBC (At5g25760) [200] were determined using the $\Delta\Delta C_t$ method [201].

2.5.7 *FLC-Venus pulldown and mass spectrometry*

FLC-Venus or untransformed control plants were ground in liquid nitrogen and suspended in extraction buffer (20 mM Tris-HCl pH 7.5, 150

mM NaCl, 2.5 mM MgCl2, 0.5% (w/v) Triton X-100, 10% (w/v) glycerol, cOmplete protease inhibitor EDTA-free (Roche, 04693159001)). After 10 min incubation with gentle rotation at 4°C, samples were cleared by repeated centrifugation at 20 000 × *g*, 4°C. Venus-tagged protein was precipitated by incubating soluble extract with GFP-Trap_M beads (Chromotek, gtm-20). Magnetic beads were washed three times with a mild wash buffer (20 mM Tris pH 8.0, 150 mM NaCl, 2 mM $MgCl_2$). Proteins were then eluted by denaturation using SDS, separated on polyacrylamide gels and either transferred to polyvinylidene difluoride membranes for analysis by immunoblotting or excised from gels for mass spectrometry. Liquid chromatography (LC)-MS/MS analysis was performed using a LTQ Orbitrap mass spectrometer (Thermo Electron) and a nanoflow-HPLC system (Surveyor, Thermo Electron), as previously described [49]. Mass spectrometry performed by Gerhard Saalbach at John Innes Centre proteomics. Data were analysed using Scaffold 4 (Proteome Software).

2.5.8 *Immunoblots*

Whole seedlings were ground to a fine powder in liquid nitrogen and incubated for 10 min in lysis buffer (50 mM Tris pH 8.0, 100 mM NaCl, 5 mM EDTA, 1% (w/v) SDS, 5 mM β-mercaptoethanol, cOmplete protease inhibitor EDTA-free (Roche, 04693159001)). After clearing by centrifugation at 20 000 × *g*, 4°C, proteins were separated on SDS polyacrylamide gels and transferred to nitrocellulose membranes (Hybond ECL, GE Healthcare). Pre-stained PageRuler PrecisionPlus (ThermoScientific) was used as a molecular weight marker. Venus-tagged protein was detected with either a rabbit polyclonal anti-GFP antibody (abcam, ab290) or a commercial mouse monoclonal anti-GFP antibody mixture (Roche, 11814460001). Signals were visualized by chemiluminescence (SuperSignal West Femto, Pierce) using secondary antibodies coupled to horseradish peroxidase (anti-mouse, Santa Cruz; anti-rabbit,

GE Healthcare). Membranes were reversibly stained using Ponceau S solution (Sigma-Aldrich, P7170).

2.5.9 *Chromatin Immunoprecipitation (ChIP)*

The following methods was adapted from [202], with modifications described in [203]. All buffers are described below the method.

Formaldehyde treatment and nuclei extraction. To stabilise protein DNA interactions, 2.0 g fresh plant tissue was vacuum infiltrated in PBS with 1% formaldehyde for 3 × 5 min [204, 205]. Formaldehyde was quenched by adding glycine to a final concentration of 125 mM and applying vacuum for a further 5 min. All samples from the same timepoint were incubated in separate nylon mesh bags in the same reaction vessel. Plants were then rinsed thoroughly with water, then frozen and ground in liquid nitrogen using a mortar and pestle. Samples were finally suspended in 25 mL Honda Buffer, filtered through two layers of Miracloth, and centrifuged at 3000 × *g* for 7 min. Nuclear pellets were washed three times with 1 mL of Honda buffer, with a 3 min spin at 3000 × *g* after each wash.

Chromatin fragmentation. Nuclear pellets were resuspended in nuclei lysis buffer and sonicated for 3 × 5 min on low-duty using a Bioruptor water bath sonicator (Diagenode). Lysates were cleared by centrifugation at 16 000 × *g* for 10 minutes and the supernatant was diluted 10-fold with ChIP dilution buffer, prior to immunoprecipitation to reduce the concentration of SDS.

Immunoprecipitation. All incubations and washes were performed at 4°C on a rotating mixer in 2 mL DNA LoBind tubes (Eppendorf). First, the relevant antibody (Sec. 7.1) was conjugated to pre-washed Protein A or Protein G magnetic Dynabeads, according to manufacturer's instructions, and antibody-bead complexes were washed for 5 min in low salt ChIP wash buffer. Samples were then incubated with antibody-bead

conjugates for 4-16 hours. To remove non-specific protein and DNA, beads were washed for 3 × 10 min in low salt ChIP wash buffer, 1 × 10 min high salt ChIP wash buffer and 1 × 10 min LiCl ChIP wash buffer. The final wash to remove residual salts was in TE containing 0.02% Triton X-100. Samples were then transferred to fresh tubes and all wash buffer was removed.

DNA recovery and clean-up. 100 µL of freshly-prepared 10% (w/v) Chelex resin (Bio-Rad) was added to the Dynabeads. Protein-DNA crosslinks were reversed by incubating on shaker at 95°C for 10 min. Samples were digested with 40 µg of proteinase K (Roche) for 30 min at 50°C followed by heat-inactivation at 95°C for 10 min. Input samples were extracted using phenol/chloroform/isoamyl alcohol pH 8.0 (25:24:1, Sigma, P2069) and ethanol precipitated using GlycoBlue (Ambion) as carrier. Immunoprecipitated samples were extracted with 15 µL StrataClean resin (Agilent), according to manufacturer's instructions.

qPCR. Performed as described in Sec. 2.5.6. ChIP data were normalised to either "Input" (total chromatin before immunoprecipitation) or H3 ChIP. Data were typically expressed relative to an internal control gene. For H3K27me3 the control gene was *SHOOT MERISTEMLESS* (*STM*, At1g62360). *STM* is expressed in only a few cells of the shoot meristem and accumulates H3K27me3 uniformly [206]. *STM* was previously used for H3K27me3 ChIP normalisation [83, 85, 123]. For H3K36me3 ChIP the control gene was *ACTIN2* (*ACT2*, At3g18780). *ACT2* is expressed uniformly in the tissues and developmental time-points studied here [200], and was previously used for H3K36me3 ChIP normalisation [123].

Honda buffer: 20 mM HEPES-KOH pH 7.4, 0.44 M sucrose, 1.25% (w/v) Ficoll, 2.5% (w/v) Dextran T40, 10 mM MgCl2, 0.5% (w/v) Triton X-100, 5 mM DTT, cOmplete protease inhibitor cocktail (Roche)

Nuclei lysis buffer: 50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1.0% SDS, cOmplete protease inhibitor cocktail (Roche)
ChIP dilution buffer: 1.1% (w/v) Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.0, 167 mM NaCl.
Low salt ChIP wash buffer: 20 mM Tris-HCl pH 8.0, 2 mM EDTA, 150 mM NaCl, 0.1% (w/v) SDS, 1.0% (w/v) Triton X-100.
High salt ChIP wash buffer: 20 mM Tris-HCl pH 8.0, 2 mM EDTA, 500 mM NaCl, 0.1% (w/v) SDS, 1.0% (w/v) Triton X-100.
LiCl ChIP wash buffer: 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 250mM LiCl, 1.0% (v/v) NP-40, 1.0% (w/v) sodium deoxycholate.

2.5.10 *Confocal microscopy*

Confocal microscopy data were obtained for homozygous single-copy FLC-Venus and FLC-mCherry lines at the T3 generation or F1 plants generated by crossing these lines. Imaging was performed using a 20× / 0.7 NA multi-immersion lens, with water as the immersion fluid on a Leica TCS SP5 confocal microscope equipped with Leica HyD Hybrid detectors. For z-stacks, the step size was 3 µm, which meant that each nucleus was typically observed in 2-3 consecutive confocal z-slices. For single-fluorophore experiments with FLC-Venus lines, roots were immersed in 2 µg/mL propidium iodide (Sigma-Aldrich, P4864) to label the cell wall. The emission spectrum of propidium iodide overlaps with that of mCherry and labels DNA in permeable cells, so could not be used in FLC-mCherry or double-fluorophore experiments. The following wavelengths were used for fluorescence detection: FLC-Venus excitation 514 nm and detection 511-555 nm (with 514 nm notch filter), propidium iodide excitation 514 nm and detection 626-697 nm, FLC-mCherry excitation 561 nm and detection 570-620 nm. To allow comparison between treatments, the same laser power and detector settings were used for all FLC-Venus images and all FLC-mCherry images, respectively. For double-fluorophore experiments, Venus and mCherry

fluorophores were simultaneously excited at 514 nm and 561 nm, respectively.

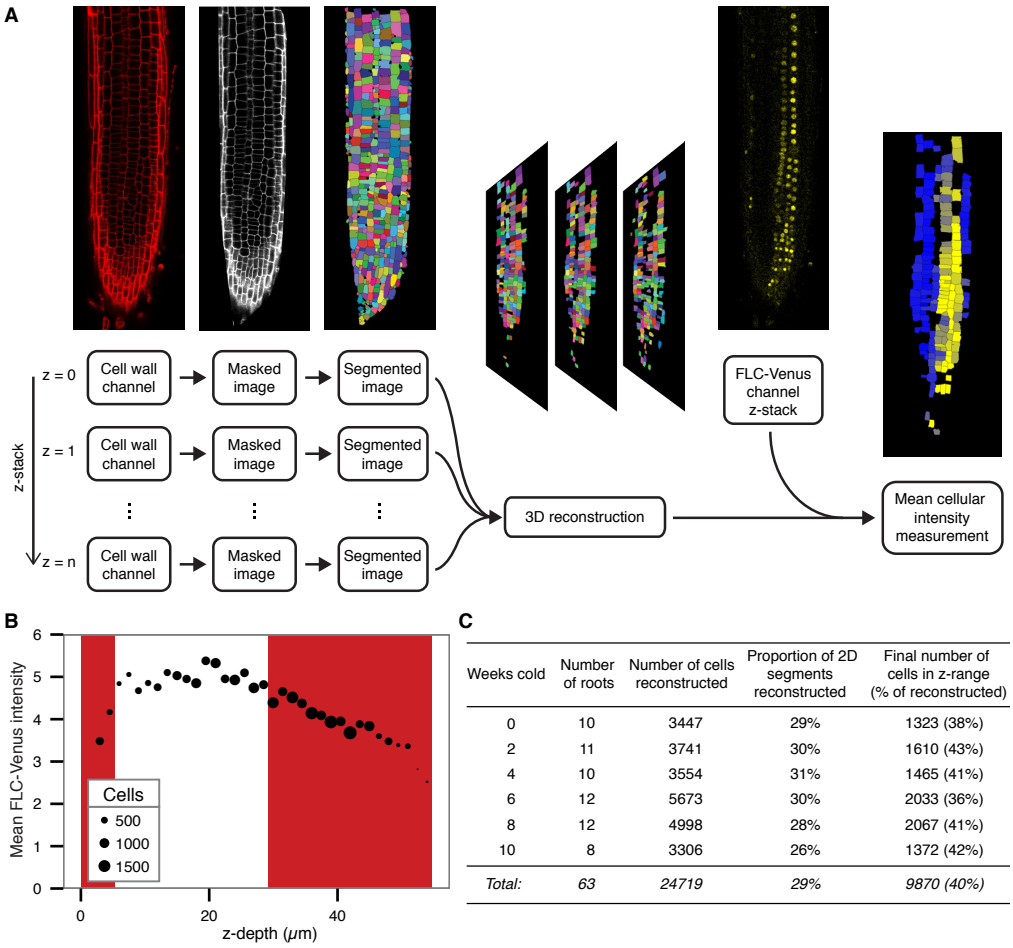
The following steps were used to prepare images for presentation: raw confocal z-stacks were aligned using the MultiStackReg plugin in Fiji [207]. To reduce detector noise, a Gaussian blur with a 1.5 pixel radius was then applied to images measuring 2048 × 1024 pixels (510 × 255 µm), before taking maximum intensity projections over 2-4 z-planes (6-12 µm). Finally, the intensity was linearly adjusted separately for each channel. For visual comparison of nuclear intensity between different treatments in FLC-Venus images, the same linear adjustment was used.

Quantitative image analysis. Confocal z-stack images were analysed using a custom image-processing pipeline to reconstruct cellular volumes and calculate the mean FLC-Venus fluorescence intensity per cell. The algorithm[1] used for quantification of FLC-Venus from confocal z-stacks was developed by Matthew Hartley and Tjelvar Olsson at the John Innes Centre, Department of Computational and Systems Biology [208]. The approach is described below and summarized in Figure 2.16.

Since the propidium iodide cell wall stain is not compatible for imaging with mCherry, cell segmentation could not be performed for FLC-mCherry or double fluorophore experiments. Image analysis was therefore undertaken for FLC-Venus only.

Pixels not corresponding to root tissue were masked using a series of morphological transforms of the propidium iodide (cell wall) images from each stack. To prepare the masked cell wall data for segmentation into individual cells, a Gaussian filter (using a standard deviation of 2 pixels) followed by median-based local thresholding (using Fiji's Auto Local Threshold plugin and a radius of 40 pixels [207]) was applied to each plane of the stack. Each image was skeletonised using

[1] Source code is available at https://github.com/JIC-CSB/root-image-analysis



Weeks cold	Number of roots	Number of cells reconstructed	Proportion of 2D segments reconstructed	Final number of cells in z-range (% of reconstructed)
0	10	3447	29%	1323 (38%)
2	11	3741	30%	1610 (43%)
4	10	3554	31%	1465 (41%)
6	12	5673	30%	2033 (36%)
8	12	4998	28%	2067 (41%)
10	8	3306	26%	1372 (42%)
Total:	*63*	*24719*	*29%*	*9870 (40%)*

Figure 2.16: Quantitative image analysis. (A) Flowchart showing the processing steps used to calculate mean FLC-Venus intensity per cell. Cell wall (propidium iodide) images from each layer of the z-stack were masked and segmented. Each of these 2D segmentations was compared to those in neighbouring layers of the z-stack to assign *bona fide* cells and obtain a 3D reconstruction. Mean FLC-Venus intensity within each reconstructed cell was then computed from the corresponding region in the FLC-Venus channel. Quantification can be visualized as a heat map showing mean FLC-Venus intensity within each successfully reconstructed cell. Cells in black were not successfully reconstructed. (B) Mean FLC-Venus intensity extracted from all cells at the same depth plotted against depth in the confocal z-stack. Depth is defined as the mean distance of a cell from the top focal plane. Point area represents the number of cells at each depth. z-stacks were taken from top-to-bottom and mean fluorescence intensity per cell was found to generally decrease during image acquisition. The red shaded areas represent data that was excluded from the analysis. (C) Statistics from the automated image analysis procedure corresponding to data presented in Figure 2.8. These data contain approximately equal numbers of roots for each treatment from two independent transgenic lines: *FLC-Venus* lines 27 and 33.

the Fiji Skeletonize plugin [207] and then segmented using the Watershed plugin [207]. Together, these steps generated an individual 2D segmentation for each cell wall image in the stack. The structure of the root was then reconstructed in 3D by comparing cells in segmented 2D images with those in neighbouring planes. Briefly, cells in 2D planes were considered part of the same 3D cell if the following two criteria were satisfied: first, their centroids were within a distance of 20 pixels from one another, and second, their relative areas did not vary by more than 50%. In addition to these criteria, the maximum extent of a single cell in the z-direction was limited to 18 μm (6 z-planes). This algorithm was implemented in Python (http://www.python.org) using the Scikit Image library (http://scikit-image.org). Finally, reconstructed 3D volumes were applied to the images from the FLC-Venus fluorescent channel to calculate mean intensity across the reconstructed volume by summing FLC-Venus intensity inside the reconstructed cell (from multiple z-planes) and dividing by the total volume of the reconstructed cell (summed area from multiple z-planes).

To validate the method, the mean FLC-Venus intensity per cell was estimated manually for a random selection of 50 cells from 22 different roots. Comparison of these results with those generated by the automated procedure for the same cells indicated that the mean cellular FLC-Venus intensities were accurate in approximately 80% of cells. The remaining cells in this test set were incorrectly segmented by the algorithm.

Consecutive z-stack images were separated by 3 μm and each root typically contained 14-18 images, which encompassed approximately the top third of the root in the meristematic and elongation zones. It was observed that FLC-Venus intensity decreased with depth in the image stack (Fig. 2.16B). This effect may have arisen due to photobleaching as a greater depth also corresponded to a later image acquisition time. To reduce this effect, the analysis was restricted to those planes where the

intensity was approximately constant. The number of roots analysed and other statistics are shown in Fig. 2.16C.

3 MOLECULAR MECHANISMS OF CIS EPIGENETIC MEMORY

How epigenetic memory could be stored in patterns of histone modifications is a subject of much experimental and theoretical work. As outlined in Sec. 1.4 (Fig. 1.1C), current models propose that modified histones inherited by daughter chromosomes through DNA replication act in cis as inherited epigenetic memory elements. This inheritance, together with cis-acting positive feedbacks, could reinforce and maintain the histone modification status of a chromatin domain (reviewed in [23, 24, 30]). Theoretically, this mechanism can only generate stable epigenetic memory for sufficiently large systems [28]. That is, systems with a sufficient number of histones.

In this chapter, a previously-developed mathematical model of *FLC* chromatin [83] is introduced and predictions of the model are compared with new experimental results. Two major inconsistencies between the model and experimental data are highlighted. First, the model hypothesized the existence of an opposing histone modification to H3K27me3 with a well-defined spatial profile across the *FLC* locus. Experiments so far have not found a histone modification able to satisfy the requirements of the opposing mark in the model, but instead led to the identification of an antagonism between H3K27me3 and H3K36me3 that is confined to the *FLC* nucleation region only[a] [123]. However, this region is not of sufficient size for storing epigenetic memory in the original model. Second, the model predicted that spreading of H3K27me3 should occur rapidly after the end of cold exposure [83]. Quantitative H3K27me3 ChIP data obtained as part of the present study shows that spreading actually occurs much more slowly, taking weeks rather than the hours or days required in the original model. Another piece of new

[a] Experiments performed by Hongchun Yang.

data relevant to the issue of 'spreading' that challenges the model is the observation that spreading does not take place at all in the *lhp1* mutant[b], despite wild-type nucleation of H3K27me3 during cold.

The two major problems for the original model are therefore the apparent lack of an opposing histone mark in the *FLC* gene body, and the slow time-scale of spreading of the nucleation peak to the gene body.

To further investigate these issues, this chapter focuses on mathematical models of the *FLC* nucleation region, where H3K27me3 has been shown to have an opposing mark, and where epigenetic memory seems to be maintained without spreading in the *lhp1* mutant. After introducing the original model and presenting the relevant experimental data, two qualitatively different models are proposed to explain the slow spreading of H3K27me3 in wild-type plants, and maintenance of the nucleation peak in *lhp1* mutants. These distinct models generate different testable predictions.

The problem of what opposes H3K27me3 accumulation in the *FLC* gene body is left unresolved in this chapter however a model in which transcription acts as the opposing state to H3K27me3, without an opposing histone modification, is introduced in Chapter 4. How these models could be combined for the case of *FLC* is discussed in Chapter 6 (Sec. 6.3).

3.1 Previous mathematical modelling of *FLC*

Based on the M-U-A model introduced in Sec. 1.4 [28], a mathematical model was previously developed for chromatin-based *FLC* regulation throughout vernalisation [83]. In the model, 'M' was identified as H3K27me3 and the identity of 'A' was not specified but was assumed to be mutually exclusive to H3K27me3. The cis-acting positive feedback for H3K27me3 was motivated by the binding and allosteric activation of PRC2 by H3K27me3 (Sec. 1.2.1, p. 25) [60]. The model was able to

[b] LIKE HETEROCHROMATIN PROTEIN 1 (LHP1) is a H3K27me3-binding protein required for epigenetic memory of *FLC* repression after cold. LHP1 was introduced in Sec. 1.5.2 (p. 1.5.2) and is studied in detail in Chapter 5.

quantitatively fit H3K27me3 ChIP data for the *FLC* nucleation region and gene body.

The *FLC* model represented the first attempt to incorporate experimentally measured parameters such as histone turnover rates and cell cycle duration, and to quantitatively compare a model of histone-based epigenetics with experimental data. In the model, the alternative high-A and high-H3K27me3 states are both self-perpetuating due to their ability to recruit marks of the same type[c]. During the cold, the nucleation region is stochastically switched from 'non-nucleated' to 'nucleated'. A 'nucleated' locus loses 'A' and acquires H3K27me3 at the nucleation region. This nucleation of H3K27me3 simulates the effect of localisation of PHD-PRC2 to the *FLC* nucleation region, which occurs during cold (Sec. 1.2.1) [49]. This PHD-PRC2 complex contains the PHD protein VIN3, which is expressed specifically during the cold and accumulates quantitatively during prolonged cold exposure [81]. VIN3 is required for epigenetic silencing of *FLC* [80] and is therefore thought to act as a trans-acting thermosensor that targets the *FLC* nucleation region. However, *VIN3* expression decreases immediately when plants are returned to warm [81]. This VIN3-mediated nucleation must therefore be transferred to a different epigenetic memory element at the end of the cold. In the model, the nucleation peak —through the internal histone-modification dynamics, can 'flip' the state of the gene body to the H3K27me3-state, causing spreading of H3K27me3 across the gene body. Once 'spread' across the gene body, the high H3K27me3 state in repressed cells is stable through cell division, allowing memory of the *FLC*-repressed state. Since H3K27me3 spreads from the nucleation region to the gene body with high probability, the number of cells in which *FLC* is repressed after cold is encoded in the number of nucleated cells at the end of cold.

The first prediction of this model was that H3K27me3 levels at the *FLC* locus should be all-or-nothing (i.e. 'digital'). This was experimen-

[c]This is similar to the M-U-A model depicted in Fig. 1.1.

tally verified at the level of *FLC* expression, using an *FLC-GUS* reporter [83]. In the present work, the mitotic heritability of the *FLC*-ON and *FLC*-OFF states was shown by confocal microscopy using *FLC-Venus* (Sec. 2.2) [208].

The second prediction of the model was that there should be an A-mark, mutually exclusive to H3K27me3 and showing opposite spatial profiles across *FLC*. This A-mark should also have its own positive feedback mechanism, and should be capable of recruiting a H3K27-demethylase. Several putative A-marks have been tested, leading to the discovery of an antagonistic relationship between H3K36me3 and H3-K27me3 at the nucleation region [123]. These results are discussed in detail in Section 3.2.1.

The third prediction of the model was that spreading of H3K27me3 from the nucleation region to the gene body should occur rapidly after cold, before the first DNA replication. This was needed theoretically because the H3K27me3 in the nucleation region is very unstable after cold exposure, after the disappearance of VIN3. Initial ChIP data taken 1, 2, 3 or 7 days after cold suggested that rapid spreading was indeed experimentally observed [83]: very little difference in gene body H3-K27me3 was observed between days 3 and 7, suggesting that 'spreading' was saturated. Subsequent experiments presented in the present work challenge this view, which has stong implications for the model. This is discussed in Section 3.2.2.

3.2 Testing original model predictions

The model predictions relating to the 'A'-mark and the time-scales of spreading are now compared with new experimental data. Inconsistencies with the *FLC* model are highlighted. The following section (Sec. 3.3, p. 99) contains a consideration of how the *FLC* model could be modified to incorporate these discrepancies.

3.2.1 *The search for the A-mark*

Based on a literature survey, histone modifications that are found in association with actively transcribed genes or have reported roles in antagonism of Polycomb silencing (H3K4me2, H3K4me3, H3K36me3, H2Bub1) were selected for quantitative ChIP analysis[d]. This list includes those marks associated with proteins of the Trithorax group (Ash1 and Trx in Drosophila [24]). In agreement with previously published data [82, 83], H3K27me3 was again found to be at low levels before cold, to accumulate gradually during the cold, and to spread over the entire locus after cold (Fig. 3.1) [123]. The level of H3K27me3 in the gene body after cold depended on the duration of cold exposure, due to an increasing number of cells having switched to the high-H3K27me3 state [83, 208].

H3K36me3 and H3K4me3 displayed similar spatial profiles across the *FLC* locus, with a peak close to the transcription start site (TSS) (Fig. 3.1). However, H3K36me3 also accumulated at low levels in the gene body, while H3K4me3 showed a second peak at the 3′ end of *FLC*. This 3′-peak increased during cold exposure, at the same time as transcription of the antisense RNA *COOLAIR* increases [84, 85], and therefore likely represents *COOLAIR* promoter activity. H2Bub1 shows a flat spatial profile across *FLC* and levels of this mark decreased for longer cold exposures.

These data indicate that nucleation of H3K27me3 during cold correlates with a loss of 'active' histone marks H3K4me3, H3K36me3 and H2Bub1. However, do any of these marks fit the profile expected of the hypothetical A-mark in the model? In the model, histones are almost always in either the A or M state. Therefore, an idea of the expected profile of the A-mark can be obtained by subtracting the spatial H3K27me3 profile from the maximum H3K27me3 level. This is shown in Fig. 3.1B. The hypothetical A-mark shows high levels across *FLC* before

[d] ChIP-qPCR experiments studying *FLC* chromatin during and after a 2, 4, 6, or 8 week cold treatment were performed by Hongchun Yang. Results were published [123] and are presented here with Hongchun's permission.

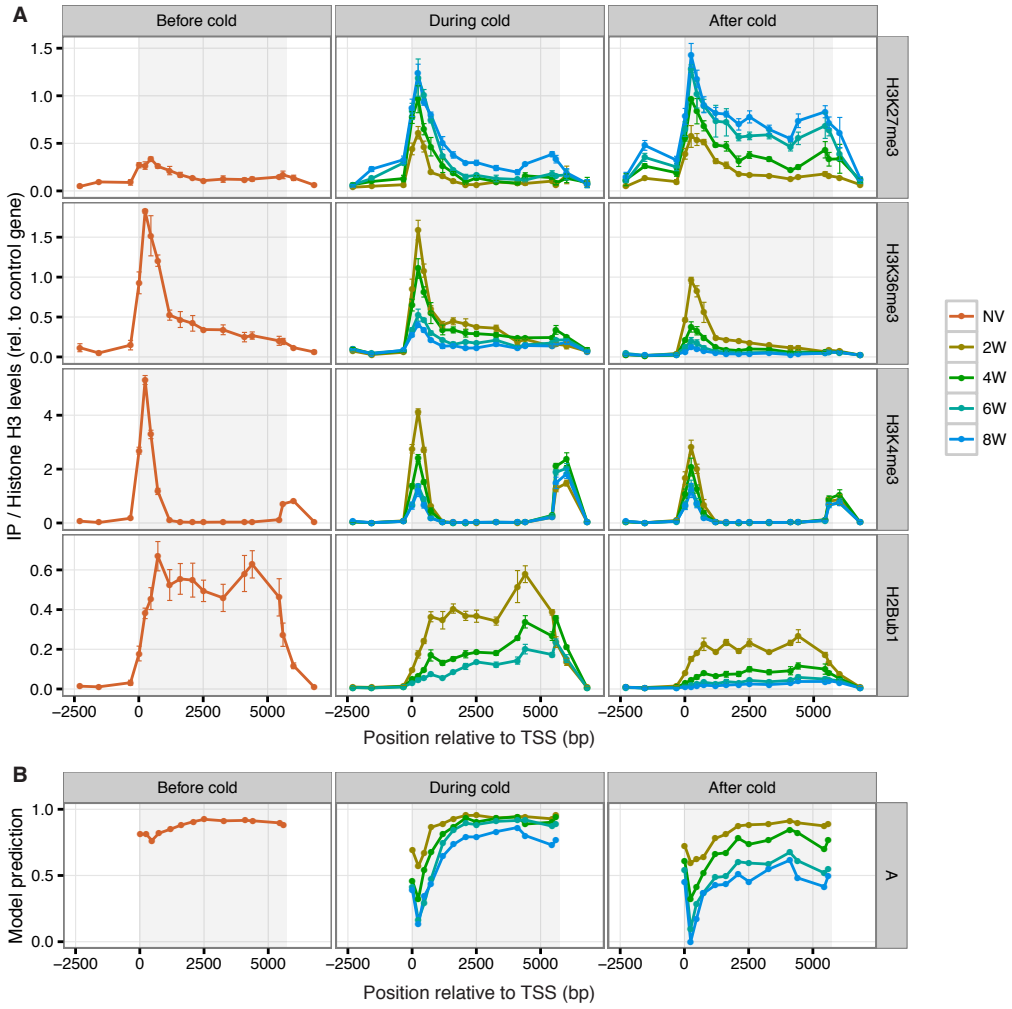


Figure 3.1: *FLC* chromatin through vernalisation. (A) ChIP for H3K27me3, H3K36me3, H3K4me3, H2Bub1 at *FLC* before, during and (7 days) after cold exposure. NV = Non-vernalised. 2W, 4W, 6W, 8W = 2, 4, 6, 8 weeks of cold-exposure, respectively. Error bars represent ± s.e.m (n = 3) between biological replicates. Grey shaded region represents the *FLC* gene. Data courtesy of Hongchun Yang [123]. (B) Prediction for the hypothetical 'A'-mark, obtained by subtracting the H3K27me3 ChIP data from the maximum H3K27me3 level

cold and this is maintained in the gene body during cold. In the model, gene body 'A' is important to stop H3K27me3 from 'invading' the gene and causing repression. Therefore, H3K36me3 and H3K4me3 alone cannot satisfy all of the requirements of the predicted A-mark. However, levels of H3K4me3 and H3K36me3 do show opposing dynamics to H3K27me3 in the nucleation region, suggesting that they are mutually exclusive in this region. Indeed, when average levels of these marks in the nucleation region are plotted against levels of H3K27me3 in the same region, an excellent negative correlation is observed (Fig. 3.2A,B). H2Bub1, on the other hand is enriched over the entire *FLC* locus before cold, suggesting that it could play a role in preventing H3K27me3 from accumulating over the gene body. To examine this more closely, H3K27me3 levels in the nucleation region (Fig. 3.2C) and gene body (Fig. 3.2D) were plotted against H2Bub1 in the gene body. It can be seen that nucleation region H3K27me3 anti-correlates with H2Bub1 in the gene body (Fig. 3.2C). However, in the gene body, large changes in H2Bub1 levels are not accompanied by corresponding changes in H3K27me3 (Fig. 3.2D). Thus H2Bub1 cannot be responsible for preventing the spread of H3K27me3 to the gene body during the cold.

To summarise, this data suggests that the active *FLC* chromatin state corresponds to H3K36me3 and H3K4me3 in the nucleation region, and H2Bub1 in the gene body. During the cold, H3K27me3 gradually replaces H3K4me3 and H3K36me3 in the nucleation region and is accompanied by loss of H2Bub1 from the gene body. After cold, H3K27me3 spreads out to cover the locus, without dramatic changes in the profiles of any of the other histone marks measured.

H3K36me3 antagonises H3K27me3 at *FLC*. If either H3K4me3 or H3K36me3 correspond to 'A' in the nucleation region, then constitutive H3K27me3 accumulation should be observed if the methyltransferases responsible for these marks are removed. It was observed that this is indeed the case for *SDG8* (*EFS*), a H3K36-methyltransferase, as

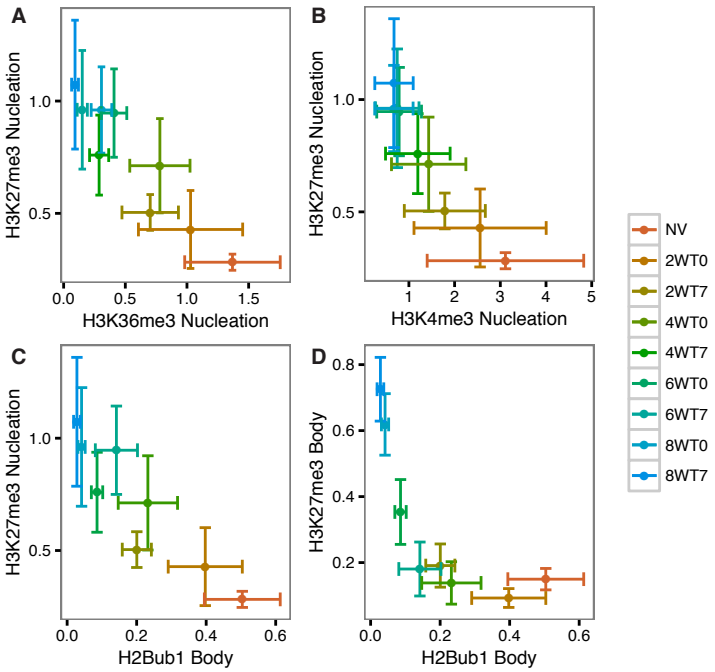


Figure 3.2: Correlations between histone modifications at *FLC*. Nucleation region H3K27me3 versus (A) Nucleation region H3K36me3, (B) Nucleation region H3K4me3, and (C) Gene body H2Bub1. (D) Gene body H3K27me3 versus H2Bub1. 'Nucleation' level is averaged over primers centred at +12, +245, +470 and +740 bp relative to TSS, while 'Body' level is averaged over primers centred at +1200, +1612, +2093, +3275, +4103, +4405, +5449, +5598 bp relative to TSS. Error bars represent standard deviation between all data points for these primers pairs, from 3 biological replicates.

FRI sdg8 plants show constitutive H3K27me3 at *FLC* in non-vernalising conditions [123]. Additional pieces of evidence that H3K36me3 acts as an 'A'-mark are that H3K36me3 is not observed on histone tails carrying H3K27me3 [123, 209], and that H3K36me3 can antagonise PRC2 activity *in vitro* [61, 62]. These data suggest that H3K36me3 could function as an A-mark in the nucleation region. Other predictions of the model that are still to be validated are that H3K36me3 is capable of recruiting a H3K36-methyltransferase (positive feedback, either direct or indirect) and also that H3K36me3 interacts with a H3K27-demethylase.

Therefore, H3K36me3 fulfils some of the requirements for an A-mark in the nucleation region. However, the nucleation region contains at most only 3 nucleosomes (6 histone tails) (Fig. 3.3), which could be easily lost during DNA replication. In fact, assuming that these 3 nucleosomes are shared at random between daughter chromosomes, 1 in 8 (2^3) DNA replications would result in loss of all 3 nucleation-region histones and therefore loss of epigenetic memory. The current model would therefore predict that this region is too small to store the epigenetic memory of the active and repressed states. Models developed later in this chapter (Sec. 3.3.3) consider how epigenetic memory could nevertheless be stored in such a small region. Before proceeding with the modelling results, however, it is useful to introduce some further experimental results.

The apparent lack of an A-mark in the gene body may simply be because not all marks have been tested. It is difficult to exclude this hypothesis without testing all marks and all combinations of marks. Alternatively, a two-state model M-U model could be implemented for the gene body —similar to models proposed for the *S. cerevisiae* mating-type locus [94–98]. In this case nonlinearity would have to be added explicitly to generate bistability, and a mechanistic hypothesis underlying this nonlinearity would need to be proposed.

In Chapter 4, another alternative is considered: that the process of transcription itself acts as an opposing state to Polycomb silencing. The

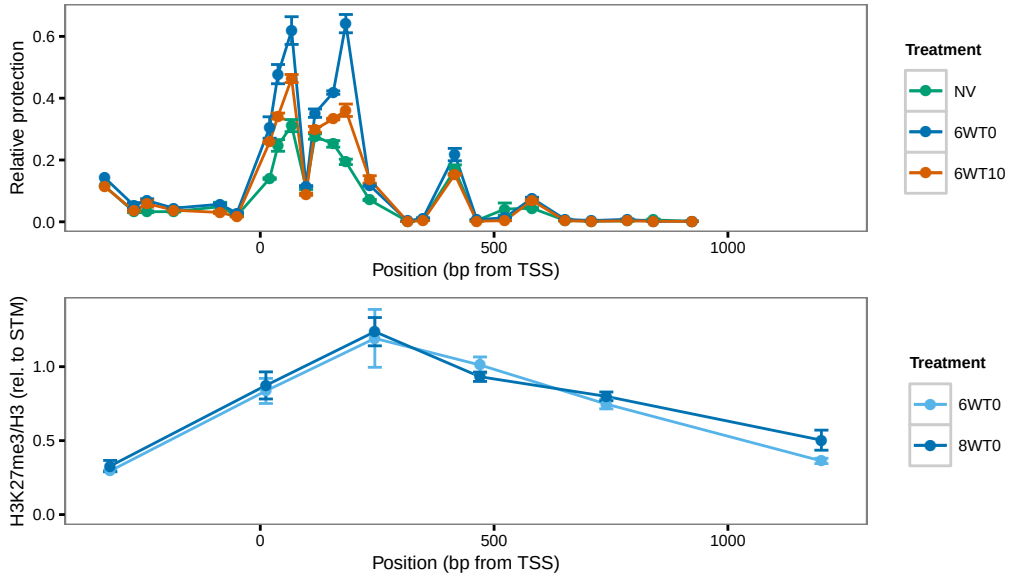


Figure 3.3: The nucleation region contains approximately 3 nucleosomes. Upper panel: Nucleosome scanning assay (NuSA) performed by Danling Zhu according to the protocol described in [210]. Nucleosomes that are well-positioned protect the DNA from digestion with micrococcal nuclease and appear as peaks in 'relative protection'. Data is represented as mean ± s.e.m. for 3 biological replicates. Lower panel: H3K27me3 ChIP data over the nucleation region, obtained by Hongchun Yang. Data is represented as mean ± s.e.m. for 3 biological replicates. The resolution of ChIP is limited by the DNA fragment size used in the assay, which is typically 500-1000 bp. NV = non-vernalised seedlings, whereas *c*WT*x* indicated that plants were harvested *x* days after a cold treatment of *c* weeks.

model developed in Chapter 4 is capable of generating bistability without adding explicit nonlinearity and without an A-mark. How this model could be integrated with nucleation region models developed in this chapter is considered in Chapter 6.

3.2.2 *Time scales of spreading*

Nucleation of H3K27me3 in the original *FLC* chromatin model was mechanistically motivated by accumulation of a PRC2 complex containing VIN3/VRN5 in the nucleation region [49, 81, 83]. For this reason, the nucleation region does not rely only on the H3K27me3/PRC2-based feedback mechanism for its maintenance during the cold. Instead, the increased concentration of VIN3 in the cold acts to drive targeting of PHD-PRC2 to *FLC*. However, VIN3 is thought to disappear within 3

days of warm conditions after cold [81]. By this time, the model requires that the nucleation peak has caused H3K27me3 to spread to the gene body. This spreading allows the H3K27me3/PRC2-based feedback to drive self-perpetuation of the chromatin state in cis because the number of histone modifications in the gene body is sufficient to buffer the perturbation of DNA replication. In the original *FLC* model, if spreading has not occurred by the time VIN3 disappears, then the nucleation peak is extremely unstable and is removed with high probability via recruitment of H3K27-demethylases by 'A' marks in the gene body, and also nucleosome turnover and DNA replication.

To examine in more detail how quickly H3K27me3 actually spreads to the gene body after cold, ChIP experiments were performed in Col-FRI at different time-points following a 6-week cold treatment. As shown in Fig. 3.4, spreading of the nucleation peak takes many days after cold, with H3K27me3 in the gene body still showing a marked increase between 10 and 18 days after cold exposure. This contrasts with the model prediction that spreading should occur, in most cases, before the first DNA replication after cold exposure. This is a major problem for the original model because in the period after cold exposure, epigenetic memory cannot be stored in VIN3 concentration and cannot be stored in the nucleation region (only 3 nucleosomes).

Underlying this 'slowness', it is assumed that spreading of H3K27me3 from the nucleation region to the gene body occurs rapidly in individual cells but with low probability in the population (i.e. spreading is all-or-nothing). Indeed, 7-10 days after cold exposure, *FLC* expression is ON or OFF in root meristem cells (Sec. 2.2) [208]. In this case, slow spreading could be explained if the nucleation peak was actually stable for a long time in the absence of spreading, and resulted in switching of the gene body to the high-H3K27me3 state reliably but infrequently in the weeks after cold exposure. That is, stochastic transitions from the 'nucleated' to the 'spread' state could be driven by H3K27me3 nucleation and the positive feedbacks in the model, provided the H3K27-

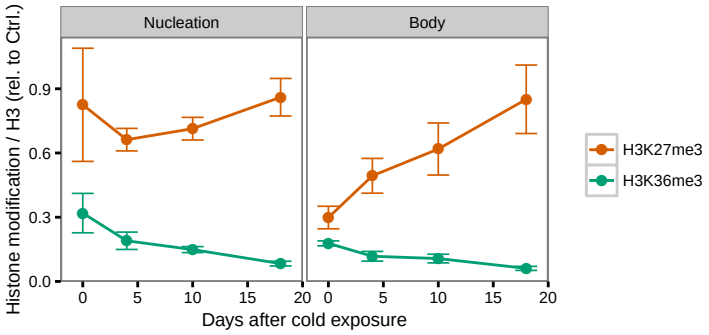


Figure 3.4: Spreading of the H3K27me3 nucleation peak. Average H3K27me3 levels in the nucleation peak and gene body in Col-FRI after a 6-week cold treatment. Nucleation region levels are averaged over primers centred at +245, +470, +868 bp from TSS, and gene body levels are averaged over primers centred at +1612, +2093, +3275, +4103, +4405, +5089, +5598 bp from TSS. Data is plotted as a function of time after cold exposure. Error bars represent the standard deviation of all data points from the respective regions.

me3-nucleation peak has sufficient stability to be maintained over this time-scale in the absence of spreading.

3.2.3 *LHP1 is required for spreading*

Other experimental data interesting in relation to this problem are the histone modification levels at *FLC* in *lhp1* mutant plants. LIKE HETEROCHROMATIN PROTEIN 1 (LHP1) binds to H3K27me3 and is required for maintenance of the repressed *FLC* state[e] (Fig. 3.5) [148, 172]. Similar loss of *FLC* repression is commonly observed in *vrn* mutants (*vrn1*, *vrn2*, *vrn5* and *vin3*). However, in contrast to these other mutants [48, 81], ChIP assays for H3K27me3 and H3K36me3 in *FRI lhp1-3* show that gain of H3K27me3 and loss of H3K36me3 at the nucleation region are unperturbed during cold (Fig. 3.6A). Differences between the wild-type and mutant at *FLC* chromatin are observed when plants are returned to warm conditions: in *lhp1*, H3K27me3 fails to spread to the gene body and H3K36me3 slowly begins to recover (Fig. 3.6 A,B). Eventually H3K27me3 levels in the nucleation region also decrease. This contrasts with the conclusion of previous chromosome-

[e] Introduced in Sec. 1.5.2

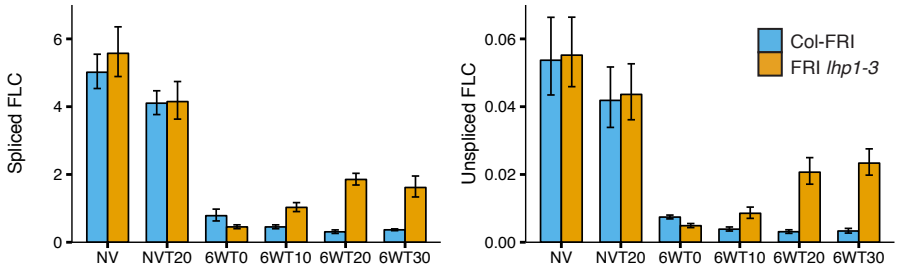


Figure 3.5: LHP1 is required for maintenance of *FLC* repression. (A) Spliced and (B) unspliced *FLC* levels in wild-type Col-FRI and *FRI lhp1-3* plants, measured by RT-qPCR. Expression levels are relative to control gene, *UBC*. Error bars represent s.e.m. between biological replicates (n = 3). NV = non-vernalised plants, whereas NVT20 indicates plants harvested after 7 days pre-growth followed by 20 days growth on soil in non-vernalising conditions. 6WTx, indicates plants harvested x days after a 6-week cold treatment (5°C).

wide studies [151], which found that H3K27me3 levels were not perturbed in *lhp1* mutants and suggested that LHP1 acts downstream of H3K27me3 in mediating gene repression. Instead, the results presented here show that H3K27me3 in the *FLC* gene body is de-stabilised in the *lhp1* mutant, despite wild-type nucleation during cold. Therefore, LHP1 is required for establishment of H3K27me3 in the *FLC* gene body.

In the context of nucleation and spreading of H3K27me3, this result is interesting for two reasons: First, it shows that nucleation peak H3K27me3 can be stable in the absence of spreading for 2-3 weeks after cold exposure. As discussed above, previous modelling had assumed that nucleation was maintained by high levels of the trans-factor VIN3, and H3K27me3 spread quickly after cold so that the ~30 nucleosome region of H3K27me3 was stable through cell division. However, the prolonged stability of the nucleation peak in warm conditions in *lhp1* suggests that the maintenance of the nucleation peak after cold is independent of VIN3. Second, since LHP1 is required for long-term maintenance of *FLC* repression, this supports the hypothesis that spreading of H3K27me3 to the gene body is required for long-term maintenance of the repressed state.

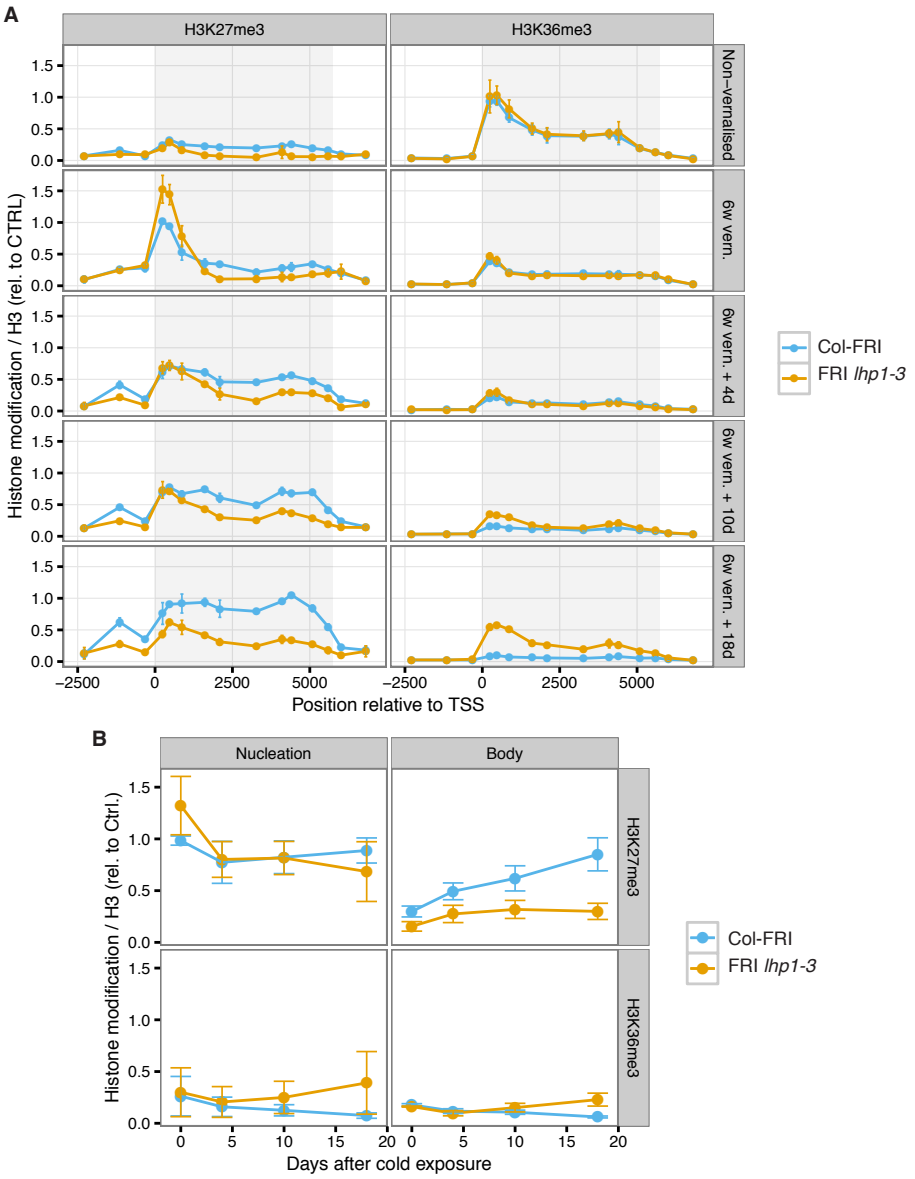


Figure 3.6: H3K27me3 spreads only weakly in *lhp1*. (A). ChIP-qPCR for H3K27me3 and H3K36me3 at *FLC* in Col-FRI and *FRI lhp1-3* during and after a 6-week cold treatment (5°C). Data is normalised to internal control gene (*STM* for H3K27me3, *ACTIN* for H3K36me3). Error bars represent s.e.m. between biological replicates ($n = 3$).(B) Average H3K27me3 levels in the nucleation region (+245, +470, +868 bp from TSS) and gene body (+1612, +2093, +3275, +4103, +4405, +5089, +5598 bp from TSS). Data is plotted as a function of time after cold exposure. Error bars represent the standard deviation of all data points from the respective regions.

In summary, spreading of H3K27me3 in wild-type Col-FRI plants is much slower than predicted by the original model at the whole plant level. This is consistent with the nucleation peak being stable for many days and slowly causing a switch to high H3K27me3 in the gene body. In *lhp1* mutants, a similar time-scale is observed for the stability of the nucleation peak in the absence of spreading.

3.3 Modelling the *FLC* nucleation region

In this section, two different explanations are considered to explain the experimental results presented in Sec. 3.2.

First, models of cis memory in small regions of chromatin are considered. This is motivated by the anti-correlation of H3K27me3 and H3K36me3 specifically in this region (Sec. 3.2.1) [123], and also the observation that the histone modification status of this region can be maintained for many days after cold in the absence of spreading (Fig. 3.6). Because there are too few histones to robustly store memory in the nucleation region alone, a hybrid protein-histone modification model is developed. This model is able to reproduce much longer lifetimes of nucleation peak H3K27me3, similar to those observed experimentally.

Second, an alternative explanation for this data is developed based on the hypothesis that non-replicating and replicating chromatin show qualitatively different H3K27me3 behaviour at *FLC*. The whole-plant samples used for expression analysis and ChIP are made up of a mixture of different cell types, the majority of which no longer undergo DNA replication, as described in Section 1.6. Therefore, the stability of the H3K27me3 nucleation peak after cold in the population-average data, does not necessarily imply that the H3K27me3-peak is actually maintained through DNA replication. Alternatively, the nucleation peak may be stable only in non-replicating chromatin, which may persist as a large portion of the total chromatin in whole-seedling ChIP samples for weeks after cold. Chromatin that is replicated after cold would initially make up only a small proportion of the total plant chromatin. As the

plant grows after cold, however, this proportion would increase gradually and eventually make up the majority of chromatin. If H3K27me3 was able to spread to the gene body quickly in replicating chromatin but not at all in non-replicating chromatin, this 'two-population effect' could underlie the apparent slow-spreading on the whole-population level.

Toward the end of this section, these two models are compared and experiments are proposed to differentiate between these hypotheses.

3.3.1 *Introduction to the M-U-A model*

Before proceeding with the *FLC* modelling, it is necessary to describe more formally the theoretical modelling framework introduced in previous work. The abstract M-U-A model (Sec. 1.4) introduced in [28] is conceptually the simplest model of histone-modification-based memory and provides a useful starting point. The model is schematically illustrated in Figure 3.7A.

Consider a region of N histones. Histones can exist in one of three states: M, U, A. From some initial configuration, Monte Carlo simulations (Sec. 3.5.4) are used to study how the chromatin state evolves over time. At each time step, a histone n_1 is selected at random from the N histones and either a 'recruited' transition is performed (with probability α) or else a 'noisy' transition is performed. In a recruited transition, another randomly selected histone n_2 is chosen from the same region. If n_2 is in either the M- or A-state, then n_1 is converted one step toward the state of n_2. e.g. M can convert A to U or U to M. In a noisy transition, n_1 is converted one-step in either direction with a probability of $1/3$, as shown in Figure 3.7A. One 'sweep' represents, on average, one attempted reaction per histone [28].

The recruited conversion simulates the stabilisation or activation of a histone modifying complex by a histone modification of the same type within the chromatin region. Implicit in this algorithm are the 'long-range interactions' described in Sec. 1.4. Bistability emerges when the

ratio of recruited to noisy transition is sufficiently high. This is conveniently expressed by the feedback, $F = \alpha/(1-\alpha)$, the ratio of signal to noise.

3.3.2 *The effect of system size on epigenetic stability*

To determine how epigenetic stability in histone-based memory depends on the size of the system, the simple M-U-A model was simulated according to the algorithm described above. It should be noted that the models simulated here are based on two modification sites per nucleosome (i.e. two H3 tails), whereas the original model [28] considered only one. Figure 3.7B shows the proportion of M and A histones over time obtained from Monte Carlo simulations of the M-U-A model. It can be seen that once initiated in the high-M state, the internal chromatin dynamics stabilise the high-M state, and the system can exhibit robust bistability. However, the system becomes increasingly noisy as the number of components is reduced (Fig. 3.7B). Fig. 3.8 shows the average lifetime of the high-A or high-M state as a function of the number of histone modifications in that state (in the absence of DNA replication). While only large systems are very stable at high noise levels (low F), it can be seen that at reasonably high noise levels ($F = 10$, $\approx$10% noise), even systems as small as 3 nucleosomes (6 histones) can generate somewhat stable epigenetic memory in the absence of DNA replication.

The major problem for storing epigenetic memory in small systems comes when DNA replication is included (Figs. 3.7, 3.8). This suggests that regardless of the underlying mechanism, in order to generate stability through DNA replication in a cis-memory model of this type (where memory elements are shared randomly between daughter DNA strands), a sufficient number of components must be present so that complete loss of the memory carriers is very unlikely[f]. The exact numbers of components depends on the noise level and structure of the model, but a minimum number of cis memory elements may be expected to be on

[f] This is analogous to the situation in trans memory where one or more of the diffusible regulatory factors is present at a very low copy number in cells.

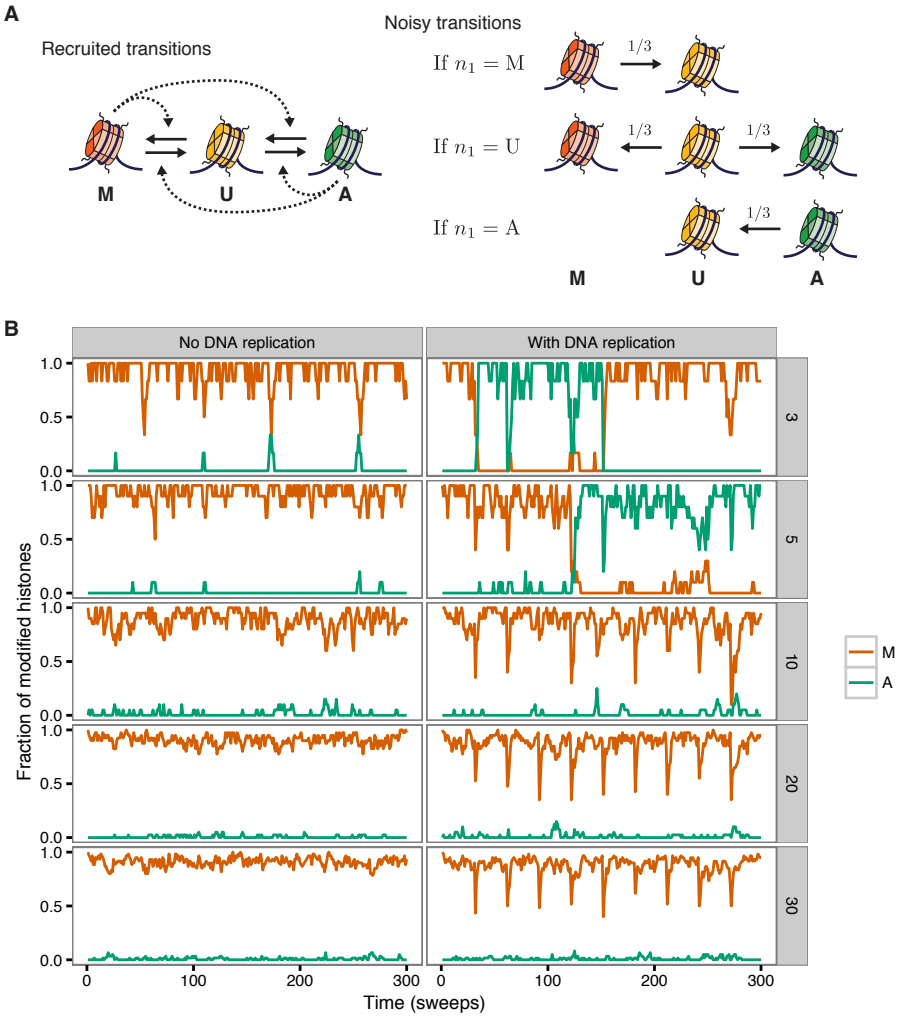


Figure 3.7: Stochastic Monte Carlo simulations of the M-U-A model (A) Schematic illustration of the M-U-A model [28]. Recruited transitions are performed with probability α at each time-step and consist of an A-mark recruiting the transition of an M- or U-mark one step toward the A-state, or vice versa for M-marks, as shown. Noisy transitions, performed with probability $(1 - \alpha)$, inter-convert between the M-, U- and A-marks, as shown. (B) The fraction of histones with an M or A modification plotted as a function of time for systems of 3-30 nucleosomes (6-60 histone tails). All systems initialised in the uniform M-state. Feedback $F = 4.0$, either without DNA replication or with DNA replication occurring every 30 sweeps. DNA replication results in replacement of half of the nucleosomes with U/U nucleosomes, and can be seen in the trajectories on the right as a halving in the number of M and A marks every 30 cycles.

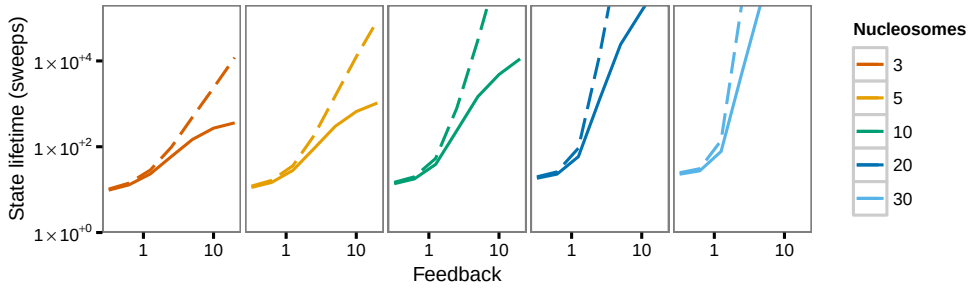


Figure 3.8: DNA replication makes small systems unstable. The lifetime of an epigenetic state in an M-U-A model containing 3-30 nucleosomes (6-60 histones) for various values of *F*, the ratio of signal to noise. Dashed lines indicate state lifetimes measured in sweeps observed with no DNA replication, whereas solid lines indicate those obtained with DNA replication every 30 sweeps [28]. The high-M state is defined as M > A, and the high-A state as $n_A > n_M$. Transition to the high-M state is scored when $n_M > 1.5n_A$ and to the high-A state when $n_A > 1.5n_M$, where n_x is the number of x histones. The lifetime is defined as the number of times a state changes between high-A and high-M divided by the total simulation time.

the order of 10. In this case, only 1 in $2^{10} = 1024$ DNA replications results in loss of all cis memory elements from the chromatin.

3.3.3 *Alternative models for the FLC nucleation region*

Having introduced the M-U-A model, this section proceeds to describe two alternative models of the *FLC* nucleation region. The first model examines the hypothesis that epigenetic memory is stored through DNA replication by a cis-memory mechanism confined to the nucleation region, and is entitled "Memory through DNA replication in the nucleation region" (p. 104). In contrast, the second model assumes only limited stability through DNA replication of nucleation-region cis memory and instead attempts to explain the slow spreading of H3K27me3 in Col-FRI as a two-population effect. That is, with replicating and non-replicating chromatin showing qualitatively different H3K27me3 behaviours at *FLC* after cold. This section is entitled "Two-populations model" (p. 114).

Memory through DNA replication in the nucleation region

Since there are not more than 3 nucleosomes in the nucleation region (6 histones), it is not possible for a purely histone modification-based system to convey stable epigenetic memory through many cell divisions, regardless of the strength of the feedback (Fig. 3.8). What other carriers of cis epigenetic memory could be located at the nucleation region?

Returning to the genetics, it is known that a group of VEL-domain proteins (VRN5, VIN3, VEL1) are necessary for *FLC* epigenetic silencing and *VIN3* and *VRN5* are both required to generate a H3K27me3 nucleation peak [48, 81]. Furthermore, it is known that VIN3 and VRN5 can heterodimerise through their VEL domains, and that VRN5 can also homodimerise [47]. This is reminiscent of the Sir silencing complex in *S. cerevisiae*, where Sir-Sir interactions are essential for the assembly of silent heterochromatin [211].

Interactions among proteins to stabilise their association on chromatin could act in a similar manner to how histone modifications are proposed to positively feed back to recruit more modifications of the same type. That is, proteins could stabilise the binding of other proteins of the same type. Including proteins in addition to histone modifications in the cis-memory model may allow more 'memory molecules' to be physically located at the nucleation region to increase stability of the epigenetic state through cell division after cold.

Two Coupled M-U-A models. To investigate the ability of proteins and histones to synergise and generate extra stability, two M-U-A models were coupled and the lifetimes of the two subsystems, and the combined system, were considered. The time-scale of protein binding and unbinding is likely to be much faster than the time-scale over which histone modifications are added and removed. Crudely, a protein-based cis memory can be modelled as a more 'noisy' M-U-A model.

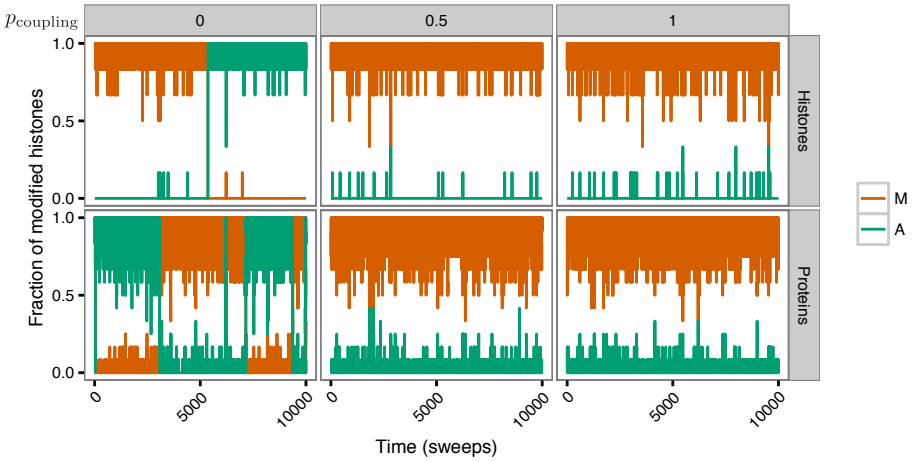


Figure 3.9: Two coupled M-U-A models. (A) Example trajectories for two M-U-A models of size 3 nucleosomes (6 histones) and 12 proteins, respectively. $p_{\text{coupling}} = 0$, represents uncoupled models, whereas $p_{\text{coupling}} = 0.5$, $p_{\text{coupling}} = 1$ represent partly and fully coupled models respectively. DNA replication is not included in these trajectories. Signal-to-noise ratios for the two subsystems are $F_{\text{H}} = 20.0$ and $F_{\text{P}} = 4.0$, respectively.

To couple two M-U-A models, the algorithm presented in Sec. 3.3.1 was modified to include the coupling parameter, p_{coupling}. In the limit $p_{\text{coupling}} = 0$, two independent M-U-A models are simulated. In the limit $p_{\text{coupling}} = 1$, recruited transitions always occur from the combined system. This corresponds to single M-U-A model of size equal to that of the combined system.

The following algorithm was used for simulations:

1. Select a histone or protein n_1 randomly from the combined system.
2. With probability p_{coupling}, select a recruiter component n_2 randomly from the combined system, or else (with probability $1 - p_{\text{coupling}}$) select a recruiter n_2 randomly from the same subsystem.
3. With probability α_{P} (for n_1 in protein subsystem) or α_{H} (for n_1 in histone subsystem), perform a recruited transition of n_1 based on the state of n_2.
4. Otherwise perform a noisy transition for n_1.

Consider a system of 3 nucleosomes (6 histone tails) and 12 proteins (assume 2 binding sites per histone tail). Before considering coupled systems with different noise characteristics, it was verified that the algorithm was able to reproduce the same state lifetimes as earlier calculated in the limits $p_{\text{coupling}} = 0$ and $p_{\text{coupling}} = 1$ (with $\alpha_{\text{P}} = \alpha_{\text{H}}$). Figure 3.9 shows example trajectories and Figure 3.10 shows the state lifetimes for the histone and protein subsystem when $F_{\text{H}} = 4.0$, $F_{\text{P}} = 4.0$ for p_{coupling} ranging from 0 to 1. State lifetimes of both subsystems initially decrease as a small amount of coupling is added. This is because the subsystems are not communicating often enough to remain in the same state and changes in one subsystem therefore act as noise for the other subsystem. However, as coupling is increased further, both subsystems eventually become more stable through their mutual interaction.

Interestingly, when the histone system has reduced noise ($F_{\text{H}} = 20.0$, i.e. 5% noise) the uncoupled state lifetime is longer than that of the protein subsystem despite having half the number of components. This is because state lifetimes depend on the *relative* fluctuations in the number of M and A marks, which can be lower for systems with fewer components, provided the ratio of recruited to noisy transitions is increased to compensate. The small, low-noise histone subsystem can therefore act as a 'long-term memory', which buffers fluctuations in the protein subsystem. Indeed, the stability of the protein subsystem increases dramatically as it becomes increasingly coupled to the histones (Fig. 3.10).

This analysis demonstrates that two systems with different noise characteristics can interact and that if the interaction strength is appropriately chosen, the state lifetimes can be increased beyond the individual lifetimes of the two systems individually.

When DNA replication is included every 30 sweeps [28], the same trends are observed. In this case the benefit of interaction is now even more pronounced, and occurs at relatively lower coupling strength. This is due to the added benefit of increased system size when DNA replication is present. In this case, due to its small size, the histone subsystem

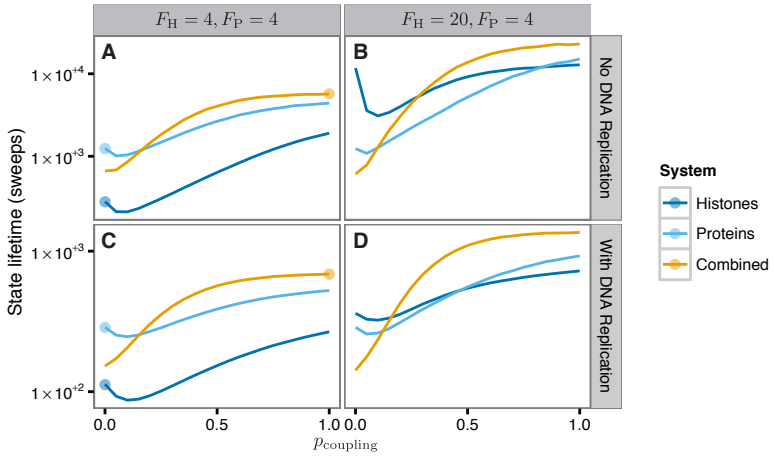


Figure 3.10: Two M-U-A systems can synergise to increase overall state lifetime. State lifetimes (in sweeps) on a logarithmic scale for M-U-A models of 3 nucleosomes (6 histone tails) or 12 proteins as the coupling strength between the two systems is increased from 0 (uncoupled) to 1 (completely coupled). Lifetimes are also shown for the combined system (A) $F_H = F_P = 4.0$. (B) $F_H = 20.0$, $F_P = 4.0$. (C,D) same as (A,B) respectively, except including DNA replication. Points in A and C show the equivalent results from simulation of a single M-U-A model.

has very low stability. However, by coupling to the protein subsystem, the stability can be dramatically increased. This is true with both equal ($F_H = 4$) and non-equal noise levels ($F_H = 20$).

Developing a nucleation-region model based on M-U-A. The results of coupling two M-U-A models suggest that protein-protein interactions could contribute to the stability of epigenetic memory stored in a region containing only 3 nucleosomes. To investigate this further, a model closer to the biological details of the *FLC* nucleation region was developed (Fig. 3.11). The remainder of simulations performed in this thesis use Gillespie's stochastic simulation algorithm [212]. A description of the advantages and validation of this approach are provided in Sec. 3.5.4.

This model is an extension of the M-U-A model with M identified as H3K27me3, and A as H3K36me3. However the model has the following important modifications:

Parameter	Description	Value
γ_{off}	Noisy protein off-rate (site^{-1}s^{-1})	0.05
γ_{on}	Noisy protein on-rate (site^{-1}s^{-1})	$\gamma_{\text{off}}/5$
α_{p}	Protein recruitment rate by proteins (site^{-1}s^{-1})	$50\gamma_{\text{on}}$
α_{h}	Protein recruitment rate by histones (site^{-1}s^{-1})	$50\gamma_{\text{on}}$
β_{h}	Protein removal rate by histones (site^{-1}s^{-1})	$50\gamma_{\text{off}}$
k	Act/Rep enzymatic activity (histone^{-1}s^{-1})	$\gamma_{\text{off}}/20$
θ	Nucleosome turnover rate (histone^{-1}s^{-1})	3×10^{-5}
μ	Bias towards H3K36me3	1.2
N_{h}	Number of histones	6
N_{p}	Number of proteins	12

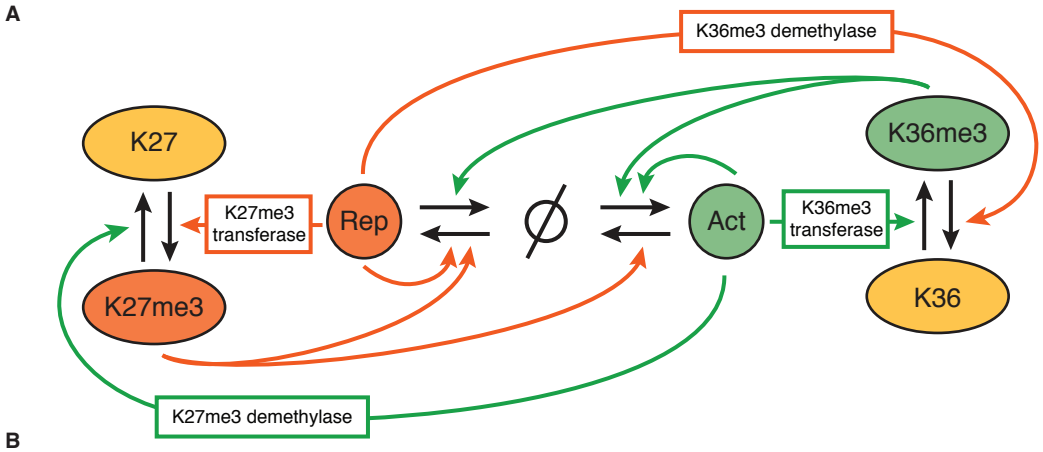
Table 3.1: Hybrid protein-histone modification model parameters. Parameter description, units and values used for simulations in Fig. 3.12.

1. H3K27me3 and H3K36me3 are not mutually exclusive. Rather, they are opposing states that destabilise one another.

2. Rather than the histone modifications implicitly recruiting proteins to modify nearby histones, the proteins are explicitly tracked, labelled as 'Act' and 'Rep' for activator and repressor, respectively. Rep represents a group of proteins, including PHD-PRC2 and any sequence-specific DNA and RNA binding proteins that are associated with the nucleation region in the repressed state. Likewise, Act represents the H3K36me3-methyltransferase EFS as well as the H3K27me3 demethylase and other 'activating' proteins.

3. Act and Rep can also recruit more proteins of the same type via direct protein-protein interactions.

4. All noise is captured by the dynamic protein binding and unbinding processes, rather than being implicitly incorporated into the histone modification transition rates.



B

Proteins

$$\text{Rep} \underset{\gamma_{\text{on}} + \alpha_{\text{p}} P_{\text{Rep}} + \alpha_{\text{h}} P_{\text{K27me3}}}{\overset{\mu\,(\gamma_{\text{off}} + \beta_{\text{h}} P_{\text{K36me3}})}{\rightleftharpoons}} \varnothing \underset{\gamma_{\text{off}} + \beta_{\text{h}} P_{\text{K27me3}}}{\overset{\mu\,(\gamma_{\text{on}} + \alpha_{\text{p}} P_{\text{Act}} + \alpha_{\text{h}} P_{\text{K36me3}})}{\rightleftharpoons}} \text{Act}$$

$$P_x = \begin{cases} \frac{1}{N_{\text{p}}} \sum_{i=1}^{N_{\text{p}}} \delta_{x,S_i} & \text{for } x \in \{\text{Act}, \text{Rep}\} \text{ (proteins)} \\ \frac{1}{N_{\text{h}}} \sum_{i=1}^{N_{\text{h}}} \delta_{x,S_i} & \text{for } x \in \{\text{K36me3}, \text{K27me3}\} \text{ (histones)} \end{cases}$$

where $S_{\text{i}} \in \{\phi, \text{Act}, \text{Rep}\}$ (proteins) or $S_{\text{i}} \in \{\text{K27}, \text{K36}, \text{K27me3}, \text{K36me3}\}$ (histones)

Histones

$$\text{K27} \underset{\mu k P_{\text{Act}}}{\overset{k P_{\text{Rep}}}{\rightleftharpoons}} \text{K27me3} \qquad \text{K36} \underset{k P_{\text{Rep}}}{\overset{\mu k P_{\text{Act}}}{\rightleftharpoons}} \text{K36me3}$$

$$\text{With probability } \theta: \begin{cases} S_j S_{j+1} \to \text{K27}_j \text{K27}_{j+1}, \\ S_j S_{j+1} \to \text{K36}_j \text{K36}_{j+1}, \end{cases} \quad \text{for } j = 1, 3, 5, \ldots, N_{\text{h}} - 1$$

DNA replication

$$\text{With probability } \frac{1}{2}: \quad S_j \to \phi, \text{ for } j \in [1, N_{\text{p}}] \text{ (proteins)}$$

$$\text{With probability } \frac{1}{2}: \begin{cases} S_j S_{j+1} \to \text{K27}_j \text{K27}_{j+1}, \\ S_j S_{j+1} \to \text{K36}_j \text{K36}_{j+1}, \end{cases} \quad \text{for } j = 1, 3, 5, \ldots, N_{\text{h}} - 1 \text{ (histones)}$$

Figure 3.11: Hybrid protein-histone modification model for the nucleation region. (A) Model Schematic. (B) Mathematical description. Parameters are defined in Table 3.1. DNA replication implemented once per day. P_x represents the proportion of proteins or histones in the state x. δ_{x,S_i} is the Kronecker delta, equal to one if $x = S_i$ and zero otherwise.

In the model, there are N_h histones ($N_h/2$ nucleosomes) and N_p binding sites for Act or Rep proteins. Rep proteins cause H3K27-methylation and H3K36-demethylation. Rep proteins are also able recruit more Rep proteins to free sites in the nucleation region. In addition, H3K27me3 can recruit more Rep proteins. For simplicity, the model is constructed to be symmetric under exchange of Act/H3K36me3 and Rep/H3K27-me3, except for the bias μ towards the Act/H3K36me3 state. As mentioned in Sec. 3.2.1 there is currently no direct experimental evidence for the positive feedback of H3K36me3 to recruit more Act proteins, though there is a theoretical requirement for multiple feedbacks to generate the nonlinearity required for bistability [213].

Nucleosomes are exchanged independently of replication with a rate $3 \times 10^{-5}s^{-1}$, which gives a mean histone lifetime of approximately 10 hours.

Like the M-U-A model [28] and the original *FLC* model [83], all interactions are 'long-ranged'. However, this is conceptually less problematic than in previous models, due to the relatively compact nature of the region being modelled (3 nucleosomes).

In agreement with earlier results for the coupled M-U-A models (Sec. 3.3.3), this model is able to generate moderate levels of bistability. State lifetimes depend on the relative rates of recruited to noisy transitions and the time-scales of protein binding and unbinding. Figure 3.12A shows the population-averaged H3K27me3 and H3K36me3 levels when the model is initialised in the repressed state, with parameter values given in Table 3.1.

It is necessary to bias the model towards the active (H3K36me3) state to prevent spontaneous re-nucleation of H3K27me3 at loci which have lost H3K27me3 and are in the active state after cold (Fig. 3.12B). Although the time-scales of protein binding and unbinding are quite fast (γ_{on} = 0.05 site^{-1}s^{-1}, γ_{off} = 0.25 site^{-1}s^{-1}), the positive feedbacks between the proteins and histones, and the slower dynamics of the histone mod-

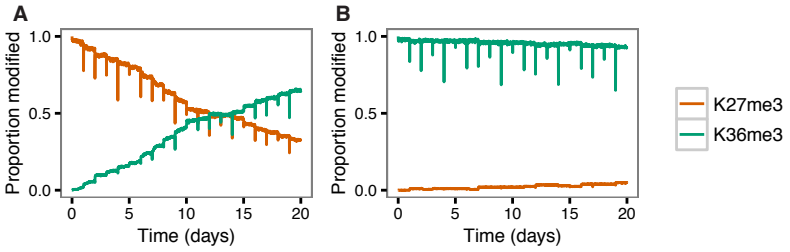


Figure 3.12: Modelling slow decay of the nucleation peak post-cold. Simulations of the hybrid protein/histone modification model shown in Fig. 3.11. Population-average H3K27me3 and H3K36me3 levels as a function of time for an initial population of 200 loci, initialised in either (A) the uniform H3K27me3-state, or (B) the uniform H3K36me3-state. Parameter values given in Table 3.1. DNA replication is implemented once per day.

ifications allow the system to only decay slowly back from the H3K27me3-nucleated to the H3K36me3-active state (Fig. 3.12A).

The time-scale over which the H3K27me3 nucleation peak is lost can be calculated directly from the simulations as the mean first passage time from an initial high-K27me3 state,

$$t_{\mathrm{FP}} = \min\{\, t \mid P_{\mathrm{K27me3}} = 0\}, \tag{3.1}$$

where P_{K27me3} is the proportion of histones carrying H3K27me3, as defined in Fig. 3.11. In these simulations, the nucleation region is almost always either completely covered in H3K27me3 or H3K36me3. The switch from majority H3K27me3 to majority H3K36me3 occurs very shortly before $t = t_{\mathrm{FP}}$, because only 6 histones are present in the region. Due to the bias towards the active state, reversals from the H3K36me3-state to the H3K27me3-state are relatively rare (Fig. 3.12B). Therefore, when the system is initialised in the H3K27me3-nucleated state, the population-average H3K27me3 level is initially well described by an exponential decay process with mean lifetime t_{FP},

$$P_{\mathrm{K27me3}}(t) = e^{-t/t_{\mathrm{FP}}}. \tag{3.2}$$

This is shown in Figure 3.13B, in which Eq. 3.2 (with t_{FP} calculated from the simulations) is plotted together with the population-average H3K27me3 data. The half-life of the nucleation peak is related to the mean first passage time by $t_{1/2} = ln(2)t_{\mathrm{FP}}$.

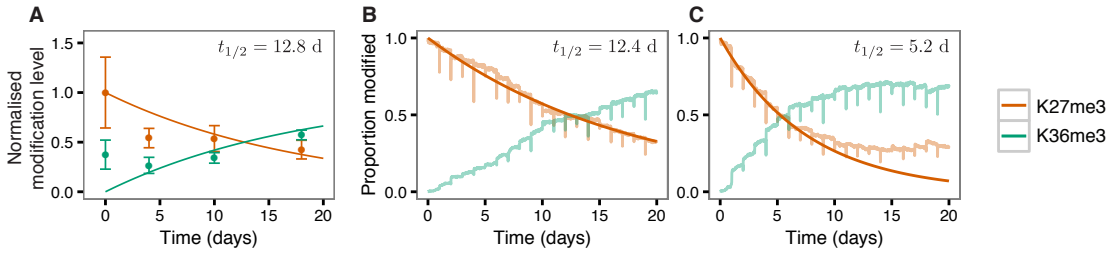


Figure 3.13: Lifetime of the H3K27me3 nucleation peak. (A) Nucleation region H3K27me3 and H3K36me3 levels in the *lhp1* mutant after a 6-week cold treatment. ChIP data are normalised and plotted as in Fig. 3.6B. Solid line shows the fit of a model where reversal of H3K27me3 to H3K36me3 occurs at a constant rate (exponentially-decaying H3K27me3). (B) Population-average H3K27me3 and H3K36me3 levels in the nucleation region as a function of time from an initial repressed state from simulations of the hybrid protein/histone modification model shown in Fig. 3.11. Parameter values are given in Table 3.1. Over-plotted exponential decay (Solid dark line, Eq. 3.2) of the nucleation peak H3K27me3 with the mean t_{FP} calculated according to Eq. 3.1, directly from simulations. (C) Same as (B), except with $\alpha_p = 0$. In C, deviation of Eq. 3.2 from the simulation data is caused by reversal of H3K36me3 loci in the population to H3K27me3.

To calculate $t_{1/2}$ from the experimental ChIP data in *lhp1*, H3K27me3 nucleation peak data were normalised to the mean level at $t = 0$ (the end of cold exposure), and H3K36me3 data were normalised to the mean non-vernalised level. Eq. 3.2 was then fit to all data for $t \geq 0$, with the requirement that $P_{\mathrm{K36me3}}(t) = 1 - P_{\mathrm{K27me3}}(t)$. The least-squares estimate for $t_{\mathrm{FP}} = 18.4$ days, corresponding to $t_{1/2} = 12.8$ days. The resulting fit is shown in Fig. 3.13A.

From model simulations, the nucleation peak half-life was calculated for various parameter values, as shown in Table 3.2. A nucleation peak half-life of the same order of magnitude to that observed in the experimental data requires very strong feedback, with recruited transitions approximately 50 times more likely than noisy transitions (Table 3.2). Such long lifetimes in this model also required the enzymatic rate, k of the proteins bound at the nucleation region to be carefully chosen. If k is too fast, occurring on a similar time-scale to protein unbinding (γ_{off}), the ability of the histone modifications to act as a long-term memory for the protein subsystem is diminished because noisy protein binding is rapidly converted into altered histone modification

status. However if k is too slow, enzymatic histone modifying processes fail to compete with nucleosome turnover, which opposes accumulation of histone modifications. Maximal stability was obtained when k was 10-50 times slower that γ_{off}.

To determine the role of the protein-protein interactions in contributing to stability in this model, the protein recruitment rate by proteins, α_{p}, was set to zero. Without changing any other parameters, this caused $t_{1/2}$ to reduce from 12.4 to 5.2 days (Table 3.2). The population-average simulation results are shown in Figure 3.12C, together with the fit of Eq. 3.2. By removing the ability of proteins to recruit more proteins, the proteins can no longer act as heritable epigenetic memory elements because they typically bind and unbind many times before resulting in enzymatic modification of a histone. As such, the model then becomes more limited in its ability to maintain epigenetic memory through DNA replication because of the small number of nucleosomes.

One further test of the model was to determine how many proteins need to be able to bind at the nucleation region to provide an increase in lifetime of the nucleation peak. Initially this was set to 2 proteins per histone (Table 3.2). Reducing the number of binding sites per histone from 2 to 1, (i.e. total 6 proteins and 6 histone tails) only reduced the lifetime of the nucleation region by 6% (Table 3.2). This suggests that, provided proteins can recruit more proteins, a doubling of the number of memory elements in the nucleation region is sufficient to greatly increase the lifetime of the nucleation region H3K27me3 in this model.

Therefore the key prediction of this model is that in order to contribute to epigenetic memory, proteins must be able to stabilise the binding of other proteins of the same type. That is, protein-protein interactions could extend the cis-memory capability of the small nucleation region and thereby increase the lifetime of the H3K27me3 nucleation peak after cold to that seen in the experimental data for the *lhp1* mutant. Further discussion of this model is left until the end of this Chapter (Sec. 3.3.4, p. 124)

γ_{off}	α_p/γ_{on}	α_h/γ_{on}	β_h/γ_{off}	N_p/N_h	μ	k/γ_{off}	t_{FP} (days)	$t_{1/2}$ (days)
0.05	50	50	50	2	1.2	5	0.7	0.5
0.05	50	50	50	2	1.2	0.5	6.8	4.7
0.05	50	50	50	2	1.2	0.05	17.9	12.4
0.05	50	50	50	2	1.2	0.005	3.6	2.5
0.05	50	50	50	2	1.2	0.0005	0.5	0.4
0.05	40	40	40	2	1.2	0.05	13.1	9.1
0.05	30	30	30	2	1.2	0.05	9.0	6.2
0.05	20	20	20	2	1.2	0.05	4.3	3.0
0.05	10	10	10	2	1.2	0.05	0.9	0.6
0.05	0	50	50	2	1.2	0.05	7.5	5.2
0.05	50	50	50	1	1.2	0.05	16.9	11.7
					Experiment (*lhp1*)		18.4	12.8

Table 3.2: Nucleation-peak lifetime. t_{FP} and $t_{1/2}$ calculated from simulations with the parameters shown, or calculated by fitting experimental H3K27me3/K36me3 ChIP data in the *lhp1* mutant (Fig. 3.13).

Two-populations model

The hybrid protein-histone modification model introduced in the previous section was developed to explain how heritable memory could be maintained in a region of chromatin the size of the *FLC* nucleation region. However, as mentioned in the introduction to Section 3.3 (p. 99), it is not clear that the nucleation region H3K27me3 actually needs to be maintained through DNA replication. An alternative explanation could be that nucleation peak H3K27me3 exists in almost all cells at the end of a 6-8 week cold exposure, but only spreads to the gene body if DNA is subsequently replicated. Chromatin that is not replicated after nucleation would remain in a K27-'nucleated' state. As shown in Figure 3.8, small cis memory models can be stable over long time-scales if there is no requirement for stability through DNA replication. In this 'two-populations' model, spreading of H3K27me3 to the gene body would

occur with the first replication after cold exposure and would then be stable through subsequent cell divisions due to the usual positive feedback in histone modifications.

Up to now, DNA replication has been regarded as an obstacle to be overcome by histone-modification-based epigenetic memory. It is therefore somewhat paradoxical to propose that DNA replication could be required for generating the H3K27me3-spread state. This is discussed further towards the end of this section (p. 122). For now, we proceed with an investigation into whether this is able to quantitatively explain the H3K27me3 time-course ChIP data.

The root meristem[g] provides a useful paradigm to think about the contributions of DNA-replicating and non DNA-replicating cells to a population over time. In root tissue, meristematic initial cells give rise to a variety of cell types (epidermal, endodermal, cortex etc.), which typically divide only a few times before commencing cycles of endoreduplication —DNA replication without cell division. After a few such endocycles, these cells then exit the cell cycle. The majority of DNA in a mature root therefore comes from quiescent, endoreduplicated cells. The initial cells of the meristem, however, continue through mitotic cell cycles, gradually repopulating the meristem and displacing the older cells. This causes differentiated cells to become more distant from the growing root tip.

In this analysis, we are interested in the relative contributions to whole-plant chromatin of DNA that *has* been replicated after cold exposure, and DNA which *has not*. Therefore, non-replicated DNA (N) is distinguished from replicated DNA (R), where 'replicated' means having undergone at least one DNA replication since cold exposure. The majority of DNA in the mature plant samples considered here is from endoreduplicated cells that have exited the cell cycle. Therefore, the DNA from DNA-replicating cells themselves is initially neglected. This approximation simplifies the analysis.

[g] Introduced in Sec. 1.6.2 (p. 42), and analysed by confocal microscopy in Sec. 2.2 (Fig. 2.5, p. 55)

At the end of cold exposure, $R(t) = 0$ and $N(t) = N(0) = N_0$, the total number of copies of DNA in cell-cycle-arrested cells. If it is assumed that the total number of DNA-replicating cells in the plant remains constant throughout development[h], then the average rate, r, at which newly synthesised DNA is added to the plant is also constant in time. Therefore, the amount of R DNA in cell-cycle-arrested cells at time t days after cold is,

$$R(t) = rt. \tag{3.3}$$

In contrast to R DNA, the absolute amount of N DNA does not increase[i], $N(t) = N_0$. Therefore, as a proportion of the total DNA,

$$\frac{R(t)}{\text{Total}} = \frac{R(t)}{R(t) + N(t)} = \frac{rt}{rt + N_0}. \tag{3.4}$$

With the following additional assumptions:

1. H3K27me3 spreading of a nucleated locus requires DNA replication,
2. H3K27me3 spreading occurs with probability 1 at a nucleated locus that undergoes DNA replication,
3. H3K27me3 spreading results in stable maintenance of H3K27me3 across the locus,
4. Once spread, the H3K27me3 level in the gene body is equal to the that of the nucleation region,

the levels of H3K27me3 in the gene body as a proportion of the total chromatin are given by,

$$\text{K27me3}_{\text{Body}}(t) = \text{K27me3}_{\text{Body}}(0) + \frac{vrt}{rt + N_0}, \tag{3.5}$$

where v is the proportion of all DNA with a H3K27me3 nucleation peak at $t = 0$ and $\text{K27me3}_{\text{Body}}(0)$ is the level of H3K27me3 in the gene body at the end of cold. In contrast, the level of H3K27me3 in the nucleation

[h] This approximation is clearly not valid for early development but is assumed to be reasonable for the time-points considered here, i.e. mature plants after a 6-week cold treatment.

[i] The amount of DNA in non-DNA-replicating cells *does* increase, however N DNA is DNA that has not divided after cold exposure.

region is,

$$\text{K27me3}_{\text{Nucl.}} = v\frac{rt + N_0 e^{-t/\tau_N}}{rt + N_0}, \tag{3.6}$$

where τ_N is the time-scale of loss of H3K27me3 from the nucleation region at *FLC* loci that do not undergo DNA replication.

Fitting H3K27me3 ChIP data. Eq. 3.5 can now be fitted to the H3K27me3 ChIP data presented in Fig. 3.4. After a 6-week cold treatment, the nucleation peak is almost fully saturated [83, 123]. It is therefore assumed that $v = 1$. The value $\text{K27me3}_{\text{Body}}(0) = 0.3$ is taken from the H3K27me3 ChIP data at 6WT0 (Fig. 3.4). There remains a single free parameter, r/N_0, to fit the slow rise of H3K27me3 in the gene body. With $r/N_0 = 0.06\ \text{day}^{-1}$, Eq. 3.5 fits the experimental data well (Fig. 3.14). The value $r/N_0 = 0.06\ \text{day}^{-1}$ indicates that for every 100 copies of non-replicating DNA in the plant, an additional 6 copies arise each day through replication. Another way of considering this rate, is to determine the time taken for the DNA synthesised after cold exposure to make up the majority of DNA in the plant. As shown in Fig. 3.14B, the amount of R and N DNA is equal after $N_0/r = 16.7$ days.

Eq. 3.5 was derived under the assumption that the contribution of the DNA-replicating cells themselves to the total DNA, could be neglected. After the first post-cold DNA replication, these previously-neglected replicating cells contain R DNA, and therefore the amount of R DNA in the plant is underestimated by neglecting these cells. Eq. 3.5 therefore represents a lower bound for the predicted levels of H3K27me3 in the gene body. A model including this DNA would therefore estimate a lower value of r/N_0 when fit to the data. Consequently, the value $r/N_0 = 0.06\ \text{day}^{-1}$ extracted from fitting Eq. 3.5 is an upper bound, and $N_0/r = 16.7$ days is a lower bound.

Is the $r/N_0 = 0.06\ \text{day}^{-1}$ value reasonable? It is difficult to know what proportion of DNA exists in replicating and non-replicating cells at a given stage of development, because a large amount of tissue growth occurs through cell growth rather than cell division [161]. For roots,

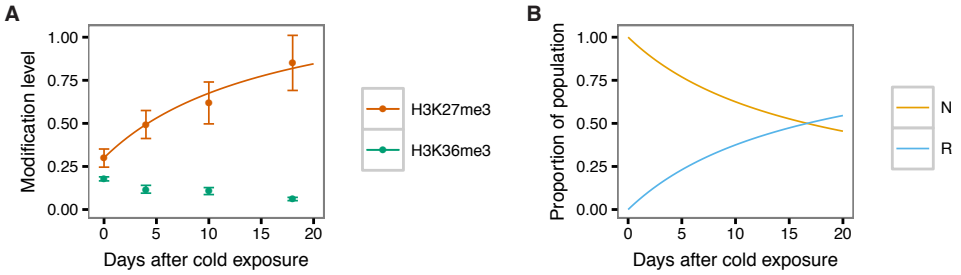


Figure 3.14: Fitting the two populations model to the slow spread of H3K27me3. (A) Data points represent average gene body H3K27me3 in Col-FRI after vernalisation. Data are analysed and presented as in Fig. 3.4. Solid line represents model prediction (Eq. 3.5) with $r/N_0 = 0.06\ \text{day}^{-1}$. (B) Proportion of the population made up of *R* DNA and *N* DNA as a function of days after cold, again with $r/N_0 = 0.06\ \text{day}^{-1}$.

rates of cell division have been estimated at 17 hours and 30 hours per cell in the meristematic zone and elongation zones, respectively [166] (Fig. 1.7, p. 41). For the shoot meristem, the estimates are similar, with the majority of cells dividing every 12-36 hours [214]. Therefore, the fastest DNA replications *in vivo* occur with a frequency of 1-2 days.

If we consider meristems as the source of all *R* DNA after cold, and assume that all meristematic cells replicate DNA on average once per day, then in a single day a meristem contributes a number of copies of *R* DNA to the plant which is equal to the total number of copies of DNA in the meristem. For the example of the root, the value $r/N_0 = 0.06\ \text{day}^{-1}$ then implies that the number of copies of DNA in the meristem equals 6% of the total DNA content of the root. Since non-replicating (endoreduplicated) cells can contain many copies of the genome, and the meristem makes up only a fraction of the total length of the root, this value seems reasonable for the mature plants considered here, which have been exposed to 6 weeks of cold.

In Sec. 3.2.3, it was observed that H3K27me3 in the nucleation region in the *lhp1* mutant is lost over a similar time-scale as H3K27me3 spreads to the gene body in Col-FRI. For *lhp1*, if it is assumed that H3K27me3 in the nucleation region is lost with a time-scale of τ_R on repli-

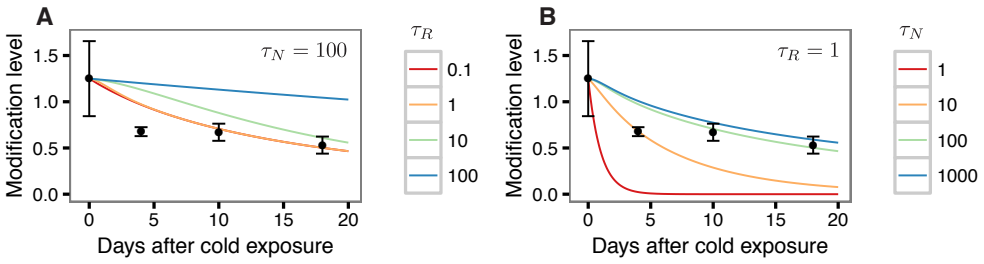


Figure 3.15: Two populations model fit to the nucleation peak H3K27me3 levels in *lhp1*. H3K27-me3 ChIP data represented as points with error bars representing standard deviation. Data are analysed and presented as in Fig. 3.6B. (A) Lines show model prediction with variable τ_R and fixed τ_N. (B) Variable τ_N and fixed τ_R.

cating chromatin. Eq. 3.6 then becomes,

$$\text{K27me3}^{lhp1}_{\text{Nucl.}} = v\frac{rte^{-t/\tau_R} + N_0 e^{-t/\tau_N}}{rt + N_0}. \tag{3.7}$$

Simulations of the M-U-A model (Sec. 3.3.1) showed that memory is more stable in non-replicating than replicating cells (Fig. 3.8, p. 103). That is, $\tau_R < \tau_N$. With τ_R on the order of several days and $r/N_0 = 0.06$ day^{-1} taken from fitting the gene body H3K27me3 data in Col-FRI, the fit of Eq. 3.7 to the ChIP data in *lhp1* is dominated by the value of τ_N (Fig. 3.15). Eq. 3.7 fits the data reasonably well for $\tau_N = 100$ days, and $\tau_R < 5$ days. In this case, the stability of the nucleation peak observed in the *lhp1* mutant is mostly due to the maintenance of a nucleation peak in cells which do not divide after cold. These cells make up a large proportion of the total DNA content, even 2-3 weeks after cold (Fig. 3.14B).

In Sec. 3.3.2, it was observed that in the absence of DNA replication, a small M-U-A model can have an arbitrarily long state lifetime that scales with the ratio of recruited to noisy transitions (Fig. 3.8, p. 103). However, this model neglected explicit nucleosome turnover and is not related to chronological time, rather the number of 'sweeps' in a Monte Carlo simulation.

To determine if a more realistic 3-nucleosome M-U-A model is capable of generating such a stable nucleation peak in non-dividing cells, simulations of the model illustrated in Fig. 3.16A were performed. DNA

Parameter	Description	Value
γ	Noisy methylation/demethylation rate (histone^{-1}s^{-1})	Various
α	Recruited methylation/demethylation rate (histone^{-1}s^{-1})	Various
θ	Nucleosome turnover rate (histone^{-1}s^{-1})	$\gamma/5$
μ	Bias towards H3K36me3	1.35
N_h	Number of histones	6

Table 3.3: Parameters for the simple K27me3/K36me3 model. Description of parameters and values used for simulation results shown in Fig. 3.16.

replication was not included. It was found that nucleation peak lifetimes are possible with a reasonable ratio of feedback to noise ($\alpha/\gamma \approx 20$). However, the time-scales of both noisy demethylation and nucleosome turnover must be extremely low, on the order of 10 days (Fig. 3.16). Histone turnover rates are known to be correlated with transcription, so if 'nucleated' *FLC* in non-replicating cells is not expressed, then nucleosome turnover may be an extremely rare event[j]. Therefore, such low noise may not be completely unreasonable.

An alternative explanation for stability in non-replicating cells, could be related to the physical clustering of *FLC* gene copies that occurs in nuclei of endoreduplicated cells after vernalisation [215]. This physical association is correlated with H3K27me3-nucleation during cold, and clusters remain stable after cold [215]. The clustering of many copies of *FLC* in non-DNA-replicating cells could allow feedback between the histone modifications at different *FLC* copies, so that PRC2 could be recruited to one *FLC* copy by a H3K27me3 mark at another copy. This could effectively increase the number of epigenetic memory molecules, and lead to increased nucleation peak lifetimes even in the presence of higher rates of nucleosome turnover.

Further evidence that nucleation of H3K27me3 is linked with this physical clustering in endoreduplicated cells is that clustering, like H3K27me3-nucleation, is unperturbed in the *lhp1* mutant, but is reduced in the *vrn2* and *vrn5* mutants [215].

[j] A detailed discussion of nucleosome turnover can be found in section 4.2.1

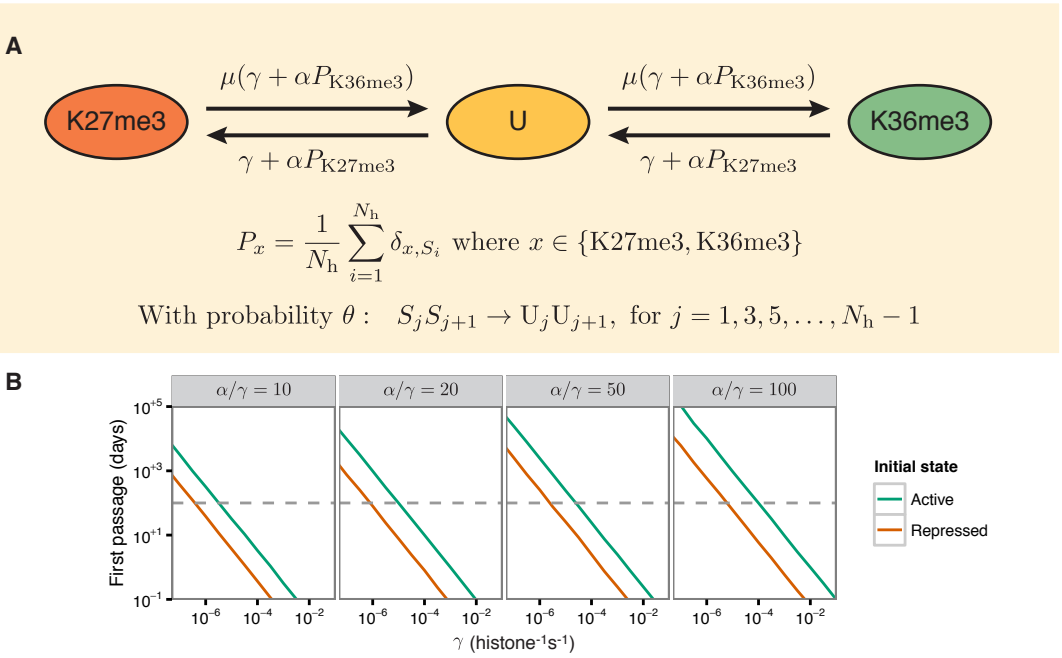


Figure 3.16: Simple histone-modification based model for the nucleation region. (A) Mathematical description of a simple model with opposing K27me3 and K36me3 histone H3 states. DNA replication is not included in the model. Each of K27me3 and K36me3 histone H3 states have positive feedback and are also mutually antagonistic. P_x is the proportion of x histones. (B) First passage time from initial active (K36me3) or repressed (K27me3) states, measured from stochastic simulations of this model using the Gillespie algorithm. The parameter α/γ expresses the ratio of recruited to noisy transitions. Dashed line indicates the first passage time τ_N = 100.

The role of DNA replication in spreading. The 'two populations' model is able to capture the slow spreading of H3K27me3 in Col-FRI and the long lifetime of H3K27me3 nucleation peak in *lhp1*. This required the assumption that DNA replication is necessary for spreading of H3K27me3 from the nucleation region to the gene body after cold exposure. Before trying to understand the mechanistic basis of this requirement in the context of the model, it is useful to introduce some published experimental data.

Cell division has historically been strongly implicated in vernalisation in a variety of plant species [216–218]. Classic experiments performed from the 1930s until the 1960s typically focused on locating the part of a plant capable of perceiving vernalising cold. In many cases it was found that growing tips, rather than mature tissues, needed to be chilled in order to invoke an acceleration in flowering [216]. This was later explained as a requirement for cell division in cold perception [218]. More recently, studies in Arabidopsis have reported that *FLC* loci in mature leaves acquire H3K27me3 at the nucleation region during cold, but that this is lost 2 weeks after cold [82]. These experiments suggest a link between DNA replication during cold, and maintenance of H3K27me3 after cold. However, for the the two-populations model to fit the H3K27me3 ChIP data presented here, spreading of H3K27me3 after cold must occur only in replicating cells, while chromatin that is H3K27me3-nucleated but does not replicate should maintain a H3K27me3 nucleation peak for several weeks. Differences in experimental design, including the age of plants when exposed to cold, mean that the results presented here cannot be directly compared with those in [82]. Therefore, previous work suggests a link between DNA replication and spreading but does not provide a direct test of the 'two populations model' hypothesis.

How could a requirement for DNA replication in H3K27me3 spreading be implemented mechanistically in the model? A trivial but plausible explanation would be that one or more of the factors required to

allow feedback of H3K27me3 in the gene body is not expressed in non-replicating cells.

Another explanation could be related to histone variants. The canonical histone H3.1 is synthesised specifically in S-phase, while the replacement histone H3.3 is synthesised independently of the cell cycle [219, 220]. It could be the case that S-phase specific H3.1 needs to be incorporated at *FLC* to allow initial spreading of H3K27me3 to the gene body. The gene body of *FLC* loci that do not divide after H3K27me3-nucleation would be primarily made up of histone H3.3. There is a precedent for this type of mechanism in the case of the H3K27me1-methyltransferase ATXR5, which selectively methylates histone H3.1 [221]. Histone-variant-dependent activity has not yet been reported for PRC2.

Yet another possibility is that there is indeed an opposing A-state in the *FLC* gene body as predicted by the original *FLC* model, that is yet to be identified. DNA replication may be required to disrupt this A-state to allow H3K27me3 spreading.

Another possibility is that DNA replication may be required to reorganise a higher-order chromatin structure which brings the H3K27me3-rich nucleation region into a conformation in which PRC2 can then be recruited to the gene body.

There are therefore many possibilities for how a single DNA replication event could act as a switch causing a nucleated *FLC* locus to spread H3K27me3 to the gene body. Before following up these possibilities, it is necessary to first show that DNA replication is indeed required for spreading.

The two models presented so far to explain current experimental data can now be compared, and experiments proposed to distinguish between these models.

3.3.4 *Comparison of the two alternative models*

The first model that was considered assumed that the long lifetime of the nucleation peak H3K27me3 in *lhp1* was due to a hybrid protein/histone-modification cis-memory. This allowed the stability of the nucleation peak through DNA replication to be increased beyond that possible in a purely histone-modification-based system. The second model considered assumed that the slow spreading of H3K27me3 to the gene body in Col-FRI was actually a population effect, with H3K27me3 only spreading to the gene body on chromatin that replicates after nucleation. The plant growth parameters extracted by fitting the time-scale of H3K27me3 spreading in Col-FRI were consistent with the slow loss of the nucleation peak H3K27me3 seen in *lhp1* - provided the nucleation peak was quite stable in the absence of DNA replication. These two models are not mutually exclusive but each makes qualitatively different predictions. The hybrid protein-DNA model suggests that protein-protein as well as histone-protein interactions are crucial for ensuring maintenance of the nucleation peak H3K27me3 through several DNA replications (Fig. 3.13B,C). In contrast, the 'two populations' model does not require such stability of the nucleation peak and instead predicts that spreading of H3K27me3 occurs only on replicating chromatin with a pre-existing nucleation peak. This failure to spread in *lhp1* mutants would cause rapid loss of the H3K27me3-nucleation peak in DNA replicating cells.

Both models require quite low noise to quantitatively reproduce the H3K27me3 ChIP data. On theoretical grounds, however, the two-populations model is favoured. This is because in the hybrid protein/histone modification model requires such high levels of feedback to be functional in maintaining epigenetic memory through cell division. At this feedback strength, all Act/Rep binding sites in the nucleation region are almost always continuously occupied with proteins. In contrast, the two-populations model does not require epigenetic memory to be

maintained in the nucleation region through cell division and is therefore more appealing. The low noise requirement for the M-U-A nucleation model is less problematic since it seems plausible that nucleosome turnover rates are extremely low at a poorly transcribed locus in a non-DNA-replicating cell.

Nonetheless, it is important to test the predictions of both models, particularly as these models are not mutually exclusive.

3.3.5 *Proposed experiments to test model predictions*

Hybrid protein/histone modification model. ChIP experiments are currently under way to precisely map the protein binding sites at *FLC* of PHD-PRC2 complex components VRN2, SWN, CLF, VRN5, and VIN3 (Hongchun Yang), and VEL1 (Danling Zhu). In addition, Yusheng Zhao is undertaking mutagenesis of the VEL domains of VRN5, VEL1 and VIN3 to identify key residues required for homo- and heterodimerisation, using a yeast two-hybrid assay. *In planta* studies with these mutants will be used to elucidate the role of protein-protein interactions in generating and maintaining H3K27me3-nucleation at *FLC*.

Two-populations model. The key prediction of this model is that H3K27me3-spreading will occur primarily at H3K27me3-nucleated *FLC* loci that have been replicated since cold exposure. That is, in tissues that contain DNA that is recently synthesised, the H3K27me3 ChIP profile across *FLC* several days after cold exposure should show similar levels of H3K27me3 in the nucleation region and gene body. The model predicts that this will contrast sharply with the whole-plant average H3K27me3 ChIP (high nucleation but relatively low spreading) and ChIP performed on non-replicating tissues (high nucleation, no spreading). To this end, experiments are under way to perform H3K27me3 ChIP on root meristems 4 days after cold, at which time most cells in this tissue will have undergone at least one DNA replication. While H3K27me3 ChIP is a routine assay in the lab, collecting sufficient 'newly-replicated'

tissue will be the main technical challenge. These experiments are being performed by Hongchun Yang.

FLC is re-expressed in vernalised *lhp1* plants only after 10 days of subsequent growth in warm conditions (Fig. 3.5, p. 97). This suggests that *FLC* re-expression is co-incident with the loss of H3K27me3 in the nucleation region (Fig. 3.6, p. 98). Therefore, another prediction of the two-populations model is that *lhp1* mutants that fail to spread H3K27me3 to the gene body, will rapidly re-activate *FLC* expression in dividing cells but not in non-dividing cells. To this end *FLC-Venus* has been crossed to *lhp1-3* and homozygous *FRI FLC-Venus lhp1-3* plants have been generated. Confocal microscopy of *FLC-Venus* in meristematic and fully differentiated (non-replicating) cells after cold exposure in these plants will allow this prediction to be tested.

3.4 Summary

The two major outstanding problems with the original model of *FLC* chromatin through vernalisation identified here are the identity of epigenetic memory elements of the active state, and the slow time-scale of H3K27me3 spreading to the gene body. This chapter focused on the role of the nucleation region in maintenance of epigenetic memory. Mathematical modelling was used to propose two possible explanations for how spreading of H3K27me3 to the gene body can be longer than the time-scale of VIN3 decay after cold, and still provide robust cis-epigenetic memory. Both of these models also help to explain the maintenance of the nucleation peak H3K27me3 in *lhp1* mutant plants. These hypotheses await experimental tests.

The identity of the epigenetic memory elements corresponding to the active *FLC* expression state seems to be partially satisfied by the H3K36me3 and possibly H3K4me3 marks in the nucleation region. However, an opposing mark in the gene body to prevent H3K27me3 accumulation has not yet been found. In Chapter 4, a model of Polycomb silencing is proposed in which the process of transcription itself acts as

an 'opposing state'. A discussion of how the nucleation peak models proposed here could be integrated with this transcription-as-the-opposing-state model is left until Chapter 6.

3.5 Materials and Methods

See the following sections for plant growth conditions (Sec. 2.5.1), plant materials (Sec. 2.5.3), RNA extraction (Sec. 2.5.5), RT-qPCR (Sec. 2.5.6), and ChIP (Sec. 2.5.9). The nucleosome scanning assay (NuSA), performed using a published protocol [210] by Danling Zhu (Fig. 3.3), is described in the figure legend. ChIP data presented in Figures 3.1 and 3.2 were obtained by Hongchun Yang, using the protocol described in Sec. 2.5.9.

3.5.1 *lhp1-3*

lhp1-3 (*tfl2-1*) is a point mutation in the LHP1 coding sequence, resulting in a stop codon at Q280. The mutant was originally isolated in [222]. The stock used in this study was obtained by former postdoctoral researcher in the Dean lab, Joshua Mylne. In the present work, *lhp1-3* was crossed to Col-FRI (Sec. 2.5.3). Plants homozygous for *FRI* and *lhp1-3* were obtained using PCR-based genotyping, as described in [223]. Primer sequences and instructions for use are provided in Sec. 7.2.

3.5.2 *Computational hardware*

Development and initial simulations were performed on a Macbook Pro running Mac OS 10.10.4. This was equipped with a dual-core Intel Core i5 processor, running at 2.7 GHz, with 16 GB system memory. More computationally intensive simulations were run on the Howard group cluster, which comprises 4 compute nodes, each equipped with 16-core Xeon E5-2650 processors, running at 2.6 Ghz, with 16 GB of system memory. The cluster runs the CentOS 6.6 distribution of the Linux operating system. Code was parallelised for the 16-core processors by

executing multiple instances of stochastic simulations on different cores in parallel.

3.5.3 *Programming languages and libraries*

All simulations were written in C and compiled using GCC (version 4.4.7). Pseudo-random numbers were generated in the GNU scientific library (GSL, version 1.13) random number environment using the Mersenne Twister 19937 algorithm [224]. The seed was either specified manually (for code development and simulating specific trajectories) or set based on the system clock using the time function of the C standard library.

R was used for all data analysis and plotting, both for results of simulations and experiments [225]. Specifically, the Rstudio environment, ggplot2 [226], reshape2 [227], plyr [228] and scales [229] packages were commonly used.

3.5.4 *Stochastic simulation algorithms*

Previous published simulations of models in chromatin-based epigenetics have employed constant time-step Monte Carlo simulations [28, 95, 96, 98, 99, 213, 230], which have also been presented in this work (Secs. 3.3.1, 3.3.3). In this algorithm, time is incremented in a series of discrete steps of constant duration, Δt. At each new time $t_{i+1} = t_i + \Delta t$, random numbers are used to determine whether or not a system update occurs. The M-U-A model is a special case in which either a recruited or a noisy transition is performed at most time steps. Typically, for more complex models, Δt must be chosen to be much smaller than the most frequently occurring system update, to ensure that the sum of all possible transition probabilities per time step Δt is much less than 1.

Running such simulations for complex systems (with a combination of short and long time-scales) becomes computationally inefficient because many time steps are simulated during which events do not actually take place. A more efficient Monte Carlo algorithm is Gillespie's

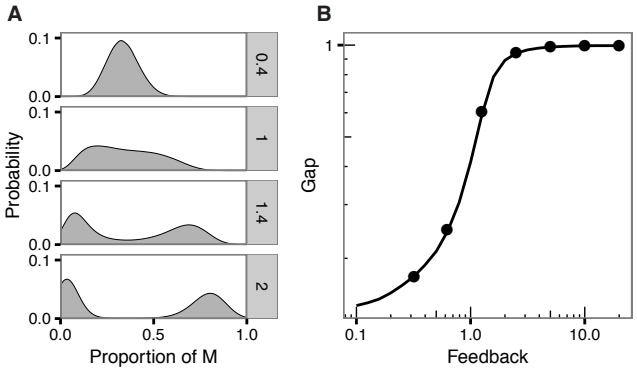


Figure 3.17: Validation of the Gillespie algorithm simulation for the M-U-A model. (A) Probability distributions for the proportion of M histones in the M-U-A model obtained using long-time simulations of the Gillespie algorithm with Feedback, $F = 0.4, 1, 1.4, 2.0$. (B) Quantitative comparison of the Gap, G calculated using the Gillespie algorithm implementation (solid line) or the Monte Carlo algorithm implementation (points).

stochastic simulation algorithm [212]. This technique is used increasingly in computational biology [231], and has already been applied in several theoretical studies of chromatin-based epigenetics [102–104, 232–234]. The Gillespie algorithm approach incorporates a sum over the probabilities of all possible system updates at any given time to calculate a time increment at which the next system update will take place [212, 235]. Simulations are therefore not performed with a constant time increment, ensuring that each system update requires a fixed number of random numbers. The Gillespie algorithm has the added advantage that reaction rates can be easily specified in units of real time and varied over arbitrary ranges without changes to the underlying computer program. This makes it ideally suited for computational biology, where searches of parameter spaces spanning many orders of magnitude are common.

To validate the implementation of the Gillespie algorithm method implemented in this work, the M-U-A model was simulated using both techniques and quantitatively compared. The Gillespie algorithm simulation reproduced the expected bistability properties of the M-U-A model (Fig. 3.17A). The 'Gap' parameter, a measure of bistability, is

defined as,

$$G = \left| \frac{n_{\mathrm{M}} - n_{\mathrm{A}}}{n_{\mathrm{M}} + n_{\mathrm{A}}} \right|, \tag{3.8}$$

where n_M and n_A are the number of M and A histones, respectively. The dependence of G on Feedback, F, agreed quantitatively between the two simulation techniques (Fig. 3.17B). These figures can be compared with the results of published simulations of this model (Fig. 2 in [28]).

4 THE ROLE OF TRANSCRIPTION IN ANTAGONISM OF POLYCOMB SILENCING

Models of chromatin-based epigenetic memory are based on the hypothesis that chromatin states determine gene expression. In the case of PRC2-dependent gene repression, there is considerable support for this hypothesis. H3K27 has been shown to be required for Polycomb-mediated repression [64] and can be passed on to daughter chromosomes [26, 236]. In Chapter 2, it was shown that two copies of a PRC2 target gene can exist in alternative, heritable expression states in the same cell, indicating that the memory of gene expression can be stored in cis [208]. Tethering of PRC2 to chromatin has also been shown to be capable of initiating transcriptional repression in both human cells [107, 237] and Drosophila [238]. Together, these findings suggest that deposition of H3K27me can cause heritable gene repression.

In contrast to this, there is also evidence that H3K27me can accumulate 'by default' at Polycomb target genes that are transcriptionally down-regulated by means other than recruitment of PRC2. It has been known for a long time from studies in Drosophila that PRC2 is often dispensable for initial transcriptional repression *in vivo* and is instead required for maintenance of the repressed transcriptional state [108–110]. More recently, similar results have been obtained in mouse cells [111]. In the latter case, it was observed that global repression of transcription through depletion or inhibition of RNA polymerase II (Pol II) is sufficient to induce ectopic recruitment of PRC2 to its target genes [111]. The same study also showed that PRC2 was dispensable for initial transcriptional shut-down of many genes that are switched off during *in vitro* differentiation of embryonic stem cells. Moreover, studies in which transcription across PRC2 target genes can be exogenously regulated have shown that transcription can counteract Polycomb repression [121,

139, 239, 240]. Considering these data, it is less clear that PRC2 and H3-K27me act as determinants of the gene expression state and may instead 'respond' to transcription (c.f. instructive versus responsive chromatin, Fig. 1.5, p. 32).

In certain cases, it therefore seems that PRC2 can determine the expression state of its targets, while in other cases it seems that PRC2 plays a more passive role, with H3K27me accumulating in response to transcriptional shutdown. In this chapter, the interplay between transcription and PRC2/H3K27me in determining and maintaining gene expression states is further explored. To this end, a mathematical model in which transcription acts as an 'opposing state' to Polycomb silencing is developed. Integration of transcription into a model of chromatin-based epigenetics allows consideration of how both cis and trans signals can regulate gene expression. The model presented here suggests that PRC2-mediated gene repression in cis is functionally significant in conveying epigenetic memory only when trans-factor determinants of expression are relatively balanced. This suggests the existence of a class of PRC2-repressed genes at which H3K27me is present but does not act as the decisive factor determining the maintenance and inheritance of gene repression.

The chapter begins by considering a two-state model in which PRC2 silencing is antagonised directly by transcription. This is shown to be insufficient for generating cis-epigenetic memory. However, by extending this model to consider different methylation states of H3K27me (me1/me2/me3), a model capable of heritable bistability is found. This model eliminates some undesirable features of previous models and allows co-ordination of the chromatin and transcription state of a gene. This model is then further extended to allow external control of the transcription rate. This permits consideration of the role of chromatin in buffering noisy input signals and also the time-scales over which chromatin states can be reliably switched. In the model, H3K27me3 can accumulate by default at transcriptionally repressed genes but can also

be recruited to induce transcriptional repression. This allows a synthesis of 'instructive' and 'responsive' models of PRC2 repression within a single quantitative framework. The model shows that cis memory is able to determine gene expression states only over a narrow range of external 'gene activation' levels, suggesting that the difference between instructive and responsive chromatin is quantitative rather than categorical.

Generality of the modelling approach. This chapter is focused on Polycomb target-gene regulation generally, rather than specifically treating the case of *FLC*. Therefore, the modelling is developed in a more abstract sense without specific reference to *FLC*. Parameters are taken mostly from studies of human or mouse PRC2. However, the model shows how 'transcription as the opposing state' could alleviate the need for an A-mark that everywhere opposes H3K27me3. In this sense, the model was developed with the *FLC* gene body in mind. A discussion of how this model could be integrated with the previous nucleation region modelling (Chapter 3) for the case of *FLC* is left until Chapter 6 (Sec. 6.3).

4.1 Two-state model

In the abstract models considered so far (Chapter 3), both the M and A histone states are able to recruit enzymatic complexes that convert other nearby histones to the same type (e.g. Sec. 3.3.1). M and A are thus opposing states and the model is symmetric. To break this symmetry and develop a model in which transcription acts to directly oppose H3K27me3 addition, the two-state model in Figure 4.1 was considered. In this model, a nucleosome contains two H3 tails, which can be either unmodified (me0) or carry the repressive K27me3 mark (me3) (Fig. 4.1A). The positive feedback for me3 is based on the allosteric activation of Ezh2 by EED-mediated binding to H3K27me3 [60].

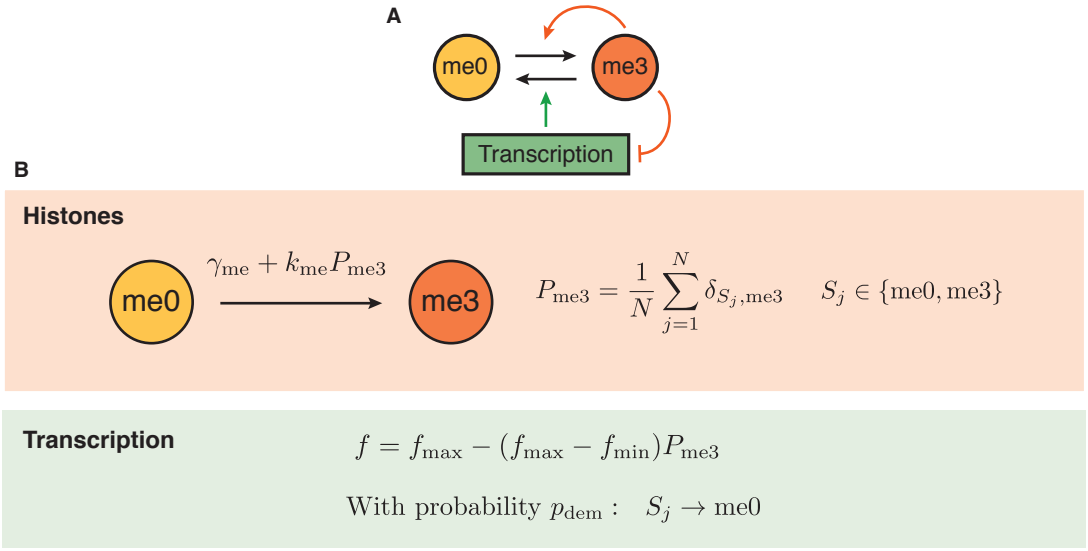


Figure 4.1: Two-state model. (A) Schematic. Black arrows represent state transitions, while coloured arrows represent feedback interactions. (B) Mathematical description. Parameter values defined in Table 4.1.

Parameter	Description	Value
k_{me}	Directed methylation rate (histone^{-1}s^{-1})	Free
f_{max}	Maximum firing rate (s^{-1})	Free
p_{dem}	Transcription-induced demethylation rate (histone^{-1}transcription^{-1})	Free
γ_{me}	Noisy methylation rate (histone^{-1}s^{-1})	$\gamma_{me} = \frac{1}{20} k_{me}$
f_{min}	Minimum firing rate (s^{-1})	$f_{min} = 10^{-4}$
N	Number of histones	60

Table 4.1: Parameters in the two-state model. Parameter definitions and values.

For the process of transcription to directly antagonise silencing by PRC2, it must be capable of removing H3K27me3. In the model it is hypothesized that this occurs in two ways: active demethylation and histone turnover. First, H3K27 demethylases have been shown to localise to both promoters and coding regions of PRC2 target genes and associate with the elongating form of Pol II [112–114]. Second, histone turnover rates are correlated with the transcriptional activity of genes, with histones resident for longer times at repressed genes [115–118]. Histone turnover is commonly interpreted as a source of noise for histone-

modification based memory [28, 83, 94, 97–103, 106, 213, 232, 241], but may also represent a mechanism by which certain histone modifications are prevented from accumulating. Histone-DNA contacts must be at least transiently disrupted to allow the passage of Pol II [242–245], however, the mechanistic basis of the transcription-dependence of histone turnover is unknown (reviewed in [246]). This effect may be due to a more compact chromatin structure and lower levels of histone acetylation at repressed genes [19, 247], which tends to promote retention of histones. Alternatively, recycling of histones in the wake of Pol II may be imperfect, resulting in loss of histones with low probability at each transcription event. These two effects (demethylation of H3K27 and increased histone turnover) provide a possible mechanism for the process of transcription to directly remove H3K27me3 and thereby antagonise silencing by PRC2.

Based on this hypothesis, the two-state model shown in Fig. 4.1A was developed. Stochastic simulations were implemented with the Gillespie algorithm. A mathematical description is shown in Fig. 4.1B and details of the simulation algorithm are provided in Sec. 4.6. H3K27me3 is added through either noisy methylation, with rate γ_{me}, or recruited methylation, with a rate k_{me}, scaled in proportion to the number of me3 modifications within the simulated region. Transcription is modelled as a discrete event which can cause H3K27 demethylation at each histone with probability p_{dem}. This single demethylation rate incorporates both histone turnover and active demethylation[a]. In this formulation, noisy demethylation is not modelled explicitly but occurs through the low level of noisy transcription in the repressed state. As discussed in Sec. 3.3.2, bistability in models of histone-modification-based epigenetic memory requires a regulatory chromatin domain on the order of tens of nucleosomes. In all simulations, a region of chromatin con-

[a] Histone turnover is similar to H3K27 demethylation in this model because both result in conversion of histones from me3 to me0. The only difference between the two processes is that histone turnover results in loss of two neighbouring histones because a H3/H4 tetramer is exchanged. To keep the model as simple as possible, this difference is neglected at this stage. In subsequent models, histone turnover is incorporated explicitly (Sec. 4.2.1).

taining 30 nucleosomes (60 H3 tails) is considered, which corresponds to a gene length of ~5kb [248].

The molecular mechanism by which PRC2 and H3K27me lead to gene repression is not well understood, but genes enriched for H3K27-me show reduced rates of productive initiation of transcription, deacetylated histones and a compact chromatin structure [63]. Acetylation of positively charged residues on histones can reduce the strength of the histone-DNA contact [249] and is believed to play both specific roles in gene activation, and more general roles in loosening chromatin structure (reviewed in [19]) Acetylation is associated with highly transcribed regions [250] and is notably depleted at repressed Polycomb target genes. In addition, H3K27Ac mediated by p300/CBP increases at PRC2 targets in loss-of-function Su(z)12 cells[b] [251]. It is currently unknown whether acetylation is a cause or consequence of gene activation. Polycomb repressive complexes have been shown to be capable of compacting chromatin templates *in vitro*, sometimes independently of their catalytic activity [69, 252, 253].

To incorporate the repressive effect of PRC2/H3K27me3 on transcription, it is necessary for the rate of RNA production to depend on the H3K27me3 levels at the locus. In the model, it is assumed that the transcriptional firing rate f is linearly dependent on the proportion of H3K27me3 marks at the gene P_{me3},

$$f = f_{\max} - (f_{\max} - f_{\min})P_{\mathrm{me3}}, \tag{4.1}$$

where $f_{\min}$ and $f_{\max}$ are the minimum and maximum transcription firing rates. It has been previously shown that adding explicit nonlinearity to two-state models can give rise to bistability [28, 94]. In this case, explicit nonlinearity is avoided by using a linear function relating the number of histone marks to the firing rate.

After fixing the rate of noisy methylation, γ_{me}, relative to the rate of recruited methylation, k_{me}, and specifying a minimum firing rate, $f_{\min}$,

[b] Su(z)12 is a core PRC2 subunit. Fig. 1.3, p. 26

there remains only three free parameters: f_{max}, the maximum firing rate; k_{me}, the directed methylation rate and p_{dem}, the probability of demethylation per firing event. The small number of parameters and computational efficiency of the algorithm allows simulations to be performed over large regions of parameter space at high resolution. For each set of parameters, 100 loci were initialised in each of the uniform me0 or me3 states and simulated according to the Gillespie algorithm [212] for 1100 hours (equivalent to 50 x 22 hour cell cycles). The aim of these initial simulations was to determine whether this model had the capacity for generating two stable gene expression states (bistability). Since DNA replication is a perturbation that weakens the H3K27me3 state, it was not included in initial simulations of the model.

4.1.1 *Lack of bistability*

To maintain a gene expression state, the system must remain in either the low expression state (high me3) or high expression state (low me3) over time. To quantify this property from the simulated time-courses, the probabilities that the gene is in the ON or OFF state (P_{ON} and P_{OFF}) were calculated. The model is intrinsically asymmetric, so large regions of parameter space represent systems biased to either the me0 or me3 states. A quantity used to measure balanced bistability is $B = 4P_{OFF}P_{ON}$ [213]. P_{ON}, P_{OFF} and B are defined precisely in Sec. 4.6.2. Values of B close to zero indicate either that P_{OFF} or P_{ON} (or both) are close to zero and represent poor ability of the model to maintain either the active and/or the repressed gene expression states. Values of B close to one, however, occur when $P_{ON} = P_{OFF} = 1/2$, and represent bimodality in the distribution of me3 histones per gene.

The ratio $F = f_{max}/f_{min}$ is equal to the fold-change in transcription firing rate between the fully active and fully repressed states. This ratio also expresses the maximum change in demethylation activity between the active and repressed states. Because demethylation in the repressed state can be considered as 'noisy', this firing rate change can

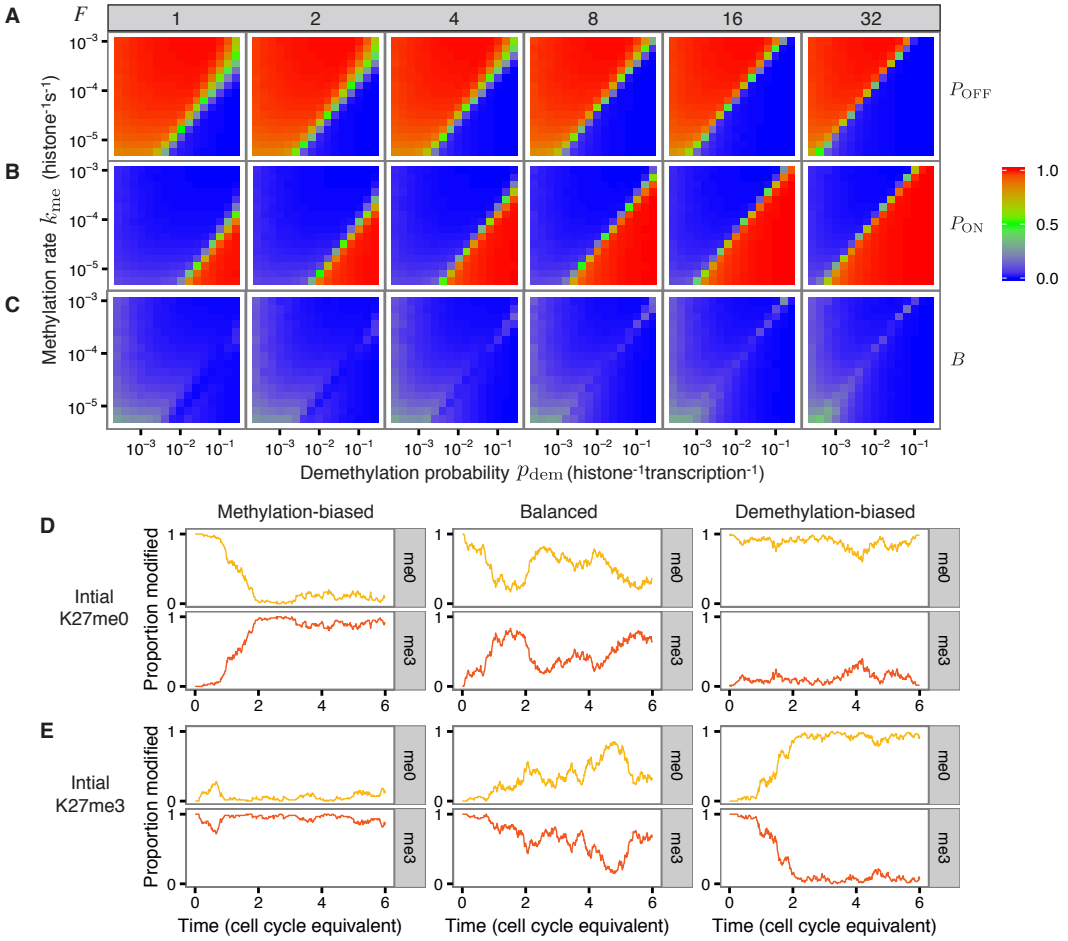


Figure 4.2: Results of simulations of the two-state model. Heat-map indicating (A) P_{OFF}, (B) P_{ON}, or (C) the bistability measure, $B = 4P_{OFF}P_{ON}$ for simulations performed with the parameters k_{me}, p_{dem} and $F = f_{max}/f_{min}$ indicated on axes and panel labels. DNA replication is not included. (D) Example simulations of a single gene initialised in the active (me0) state, for methylation-biased, balanced, or demethylation-biased parameter sets. (E) As in D with initial repressed (me3) state.

also be thought of as the ratio of signal-to-noise. There is also another signal-to-noise ratio for methylation in this model, given by k_{me}/γ_{me}: this is fixed at 20 (i.e. 5% noise).

Fig. 4.2 shows P_{OFF}, P_{ON}, and B averaged over time[c] and over a population of loci, for simulations performed over ranges of k_{me}, p_{dem} and F. Remaining parameter values are specified in Table 4.1. The population of loci were initialised as 50% uniform me0, 50% uniform me3. In agreement with earlier studies of two-state models, bistability was not

[c]Time-averaging described in Sec. 4.6.2.

observed in this two-state model, even in the absence of DNA replication [94, 98]. This is likely due to the lack of nonlinearity in both the histone modification reactions and the function linking the transcription rate to the chromatin state.

Nonlinearity has previously been added to two-state models in an arbitrary fashion and shown to contribute to generating bistability [94, 98]. In the following section, details of the PRC2 repression system are used as inspiration about how such nonlinearity may arise naturally in this system.

In summary, the linear two-state model presented here is capable of generating stable ON (low me3) and OFF (high me3) transcription states when parameters are biased in one direction or the other. However, no parameter sets were found for which both states can be stably maintained.

4.2 Non-processive model

Previous theoretical work has focused on models that include intermediate histone modification states between the marks associated with activation and repression (me0 and me3 in this case). Models with intermediate histone states provide a mechanism for implicit generation of nonlinearity, which has previously been shown to be necessary for bistability, as discussed in Sec. 1.4 [28, 95, 96, 103, 213].

In the PRC2 system, H3K27 methylation could, in principle, be accomplished by either processive or non-processive catalysis. In processive catalysis, the enzyme remains bound to a particular substrate molecule until all methylations have taken place, while in non-processive catalysis, the enzyme dissociates from the substrate after each methyl group is added [254]. The me0 to me3 reaction in the two-state model (Fig. 4.1) depicts H3K27me3 as arising in a single step, i.e. processive addition of three methyl groups. In contrast, non-processive methylation of H3K27 could generate four distinct H3K27 methylation states

(me0/1/2/3) (Fig. 4.3), which may serve as a natural mechanism to generate nonlinearity.

Evidence for non-processivity. For SET-domain histone methyltransferases, such as the catalytic subunit of PRC2, there are examples of both processive and non-processive enzymes [255, 256]. It is known that recombinant mammalian PRC2 can monomethylate H3K27me0, H3K27me1 and H3K27me2 substrates *in vitro*, and actually displays a preference for H3 peptides carrying less K27-methyl groups [257–259]. *In vivo*, PRC2 activity is required for H3K27me2/me3, whereas H3K27me1 levels are reduced but not abolished in embryonic stem cells lacking functional PRC2 [260–263], suggesting the existence of a H3K27me1-specific methyltransferase independent of PRC2. In mammals, H3K27me1 is primarily localised to active genes and enhancers [264, 265] and intragenic H3K27me1 accumulation at Polycomb target genes in mouse embryonic stem cells is dependent on PRC2 [263]. This could be explained by either PRC2-catalysed monomethylation at PRC2 targets, or non-processive demethylation of H3K27me2/me3. Heavy isotope labelling mass spectrometry analysis of H3K27 methylation in cell culture showed that H3K27me3 is mostly formed from monomethylation of existing H3K27me2 substrates and also found evidence that H3K27me2 can arise through monomethylation of H3K27me1 [266, 267]. Collectively, these data suggest that PRC2 can act non-processively both *in vitro* and *in vivo* to methylate H3K27.

Less is known about the processivity of H3K27-demethylation *in vivo*. The demethylation reaction is catalysed by jumonji-C domain containing proteins, which use a dioxygenase reaction for demethylating mono, di and tri-methylated residues [268]. The reaction mechanism occurs one step at a time, so these enzymes, like methytransferases, could act non-processively. The human H3K27-demethylase UTX has been observed to sequentially remove single methyl groups from H3K27me3 peptides *in vitro* [269].

4.2.1 *Non-processive model formulation*

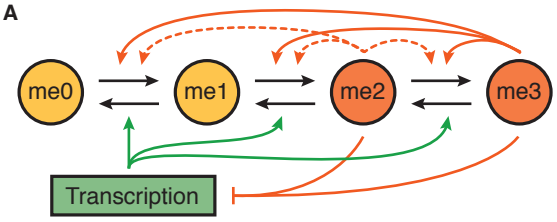
The non-processive model assumes that both methylation and demethylation occur non-processively (Fig. 4.3). Parameters are defined in Table 4.2. This model can be seen as an extension to the two-state model, in which transcription represents the opposing state to PRC2 silencing. As for the two-state model, H3K27me3 positive feedback is motivated by the binding and allosteric activation of PRC2 by H3K27me3 [60]. *In vitro* studies indicate that H3K27me2 is also able to bind to and activate PRC2 —albeit to a lesser extent than the H3K27me3, while H3K27me1 does not activate PRC2 [60]. In the model, H3K27me2 can also therefore recruit and activate PRC2, although with 10-fold reduced efficacy as compared to H3K27me3. H3K27me1 does not activate PRC2 in the model. The me0/me1 modification states can therefore be grouped as 'neutral marks' and me2/me3 as 'repressive marks'. The firing rate function (Fig. 4.3) reflects this change, with the assumption being that that H3K27me2 histones are equally as repressive for transcription as H3K27me3 histones.

The non-processive model also takes into account the relative catalytic activity of PRC2 on H3K27me0, me1 and me2 substrates (k_{cat}/K_M (me0 : me1 : me2) = 9 : 6 : 1) from *in vitro* studies [258], which is captured with the parameters $k_{\text{me0-1}}$, $k_{\text{me1-2}}$, $k_{\text{me2-3}}$, where $k_{\text{me2-3}} = k_{\text{me}}$. Noisy methylation rates $\gamma_{\text{me0-1}}, \gamma_{\text{me1-2}}, \gamma_{\text{me2-3}}$ are set at 5% of the rate of allosterically-activated PRC2 (Table 4.2).

Like in the two-state model, demethylation is coupled to transcription, so that each transcription event can remove a single methyl group on each histone, with probability p_{dem} per histone. In addition there is also a noisy rate of demethylation γ_{dem}, which is set to be equal to $f_{\text{min}}p_{\text{dem}}$ so that the noisy demethylation rate is equal to the transcription-coupled demethylation rate when the gene is maximally repressed.

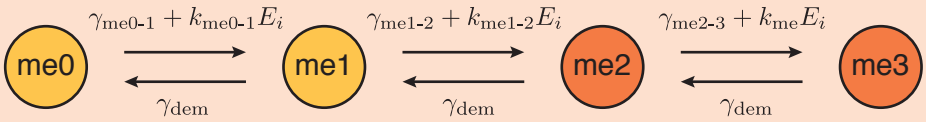
Histone turnover. For the two-state model, H3K27 demethylation and histone turnover were equivalent because me3 histones were replaced

A



B

Histones



$$E_i = \sum_{j \in \text{neighbours of } i} \left(\rho_{\text{me2}} \delta_{S_j,\text{me2}} + \delta_{S_j,\text{me3}}\right) \qquad S_j \in \{\text{me0}, \text{me1}, \text{me2}, \text{me3}\}$$

Transcription

$$f = f_{\max} - \frac{1}{N}\left(\sum_{i=1}^{N} (\delta_{S_i,\text{me2}} + \delta_{S_i,\text{me3}})\right)\left(f_{\max} - f_{\min}\right)$$

$$\text{With probability } p_{\text{dem}}: \quad S_j \to \begin{cases} \text{me0} & \text{if } S_j = \text{me1} \\ \text{me1} & \text{if } S_j = \text{me2} \\ \text{me2} & \text{if } S_j = \text{me3} \end{cases}$$

$$\text{With probability } p_{\text{ex}}: \quad S_j S_{j+1} \to \text{me0}_j \text{me0}_{j+1}, \text{ for } j = 1, 3, 5, \ldots, N-1$$

DNA replication

$$\text{With probability } \frac{1}{2}: \quad S_j S_{j+1} \to \text{me0}_j \text{me0}_{j+1}, \text{ for } j = 1, 3, 5, \ldots, N-1$$

Figure 4.3: Non-processive model. (A) Model Schematic. Black arrows represent state transitions, while coloured arrows represent feedback interactions. Dashed lines indicate a weaker interaction. 'Repressive marks' are coloured orange, while 'neutral marks' are coloured yellow. (B) Mathematical description. Parameter values defined in Table 4.2. The sum over 'neighbours' in E_i includes the other histone on the same nucleosome, and the four histones on neighbouring nucleosomes.

Parameter	Description	Value
k_{me}	Directed methylation rate (me2 to me3) (histone^{-1}s^{-1})	Free
f_{max}	Maximum firing rate (s^{-1})	Free
p_{dem}	Transcription-induced demethylation rate (histone^{-1}transcription^{-1})	Free
p_{ex}	Transcription-induced histone turnover rate (histone^{-1}transcription^{-1})	Free
ρ_{me2}	Relative activation of PRC2 by H3K27me2	$\rho_{me2} = 0.1$
$k_{me0\text{-}1}$	Directed methylation rate (me0 to me1) (histone^{-1}s^{-1})	$k_{me0\text{-}1} = 9k_{me}$
$k_{me1\text{-}2}$	Directed methylation rate (me1 to me2) (histone^{-1}s^{-1})	$k_{me1\text{-}2} = 6k_{me}$
$\gamma_{me0\text{-}1}$	Noisy methylation rate (me0 to me1) (histone^{-1}s^{-1})	$\gamma_{me0\text{-}1} = \frac{1}{20}k_{me0\text{-}1}$
$\gamma_{me1\text{-}2}$	Noisy methylation rate (me1 to me2) (histone^{-1}s^{-1})	$\gamma_{me1\text{-}2} = \frac{1}{20}k_{me1\text{-}2}$
$\gamma_{me2\text{-}3}$	Noisy methylation rate (me2 to me3) (histone^{-1}s^{-1})	$\gamma_{me2\text{-}3} = \frac{1}{20}k_{me}$
f_{min}	Minimum firing rate (s^{-1})	$f_{min} = 10^{-4}$
γ_{dem}	Noisy demethylation rate (histone^{-1}s^{-1})	$\gamma_{dem} = f_{min}p_{dem}$
N	Number of histones	$N = 60$
	Cell cycle duration (hours)	22

Table 4.2: Parameters in the non-processive model. Parameter definitions and values.

with me0 histones in a one-step processive demethylation. In the non-processive model, removal of me2 or me3 due to replacement of a H3/H4 tetramer results in incorporation of a H3/H4 tetramer that is unmethylated at H3K27. This represents a direct conversion of a pair of neighbouring histones to me0/me0, regardless of their state (Fig. 4.3). Therefore, unlike in the two-state model, histone turnover and active demethylation are no longer equivalent. In the non-processive model, each transcription event can therefore induce histone turnover with probability p_{ex} (H3/H4 tetramer exchange) and can also cause active demethylation with probability p_{dem}.

Many studies have attempted to quantify rates of histone turnover (reviewed in [246]). Metabolic labelling experiments in *S. cerevisiae* indicated that H2B turns over faster than H3, and that turnover is correlated with gene expression level [115–117]. These studies found that up to 50% of H3 over the coding region could be replaced within one hour, but failed to detect turnover at inactive genes. Similarly, pulse-chase ex-

periments in Drosophila cell culture estimated mean histone residence times of a few hours at actively transcribed genes [118]. All of these measurements were, however, limited to a short time-window of labelling, preventing accurate determination of slow rates of turnover. Canonical histone H3.1 is synthesized and incorporated specifically during S-phase, whereas histone H3.3 is incorporated throughout the cell cycle, independently of DNA replication [270]. The accumulation of H3.3 can therefore be used as a marker of histone turnover outside of S-phase in replicating cells. Genome-wide profiling of histone variant H3.3 in Drosophila and human cell lines revealed that H3.3 is not incorporated at high levels in inactive genes, suggesting that H3 is not replaced independently of replication in rapidly proliferating cells [271–275]. In relation to H3K27-methylated histones, pulse-chase mass spectrometry experiments indicate that H3K27me3 often arises through monomethylation of H3K27me2, which is inherited from a previous generation [266, 267, 276]. This suggests that histones in H3K27me2/me3-rich domains may be turned over very slowly, and perhaps not at all during a cell cycle. Consistent with this, quantitative imaging of H3K27me3 in *Caenorhabditis elegans* (*C. elegans*) embryos over multiple cell cycles did not detect replacement of H3K27me3 histones independently of replication[d] [236]. Taken together, these results indicate that H3 turnover at genes is correlated with transcriptional activity, and that inactive genes (including PRC2 targets enriched in H3K27me2/me3) have very low levels of H3 turnover independent of DNA replication.

DNA replication. DNA replication consists of replacement of each nucleosome with a new me0/me0 nucleosome with a probability of one half, once per cell cycle [28] (Fig. 4.3). This simulates the random distribution of parental histones between daughter DNA strands [25]. After each replication, only one of the daughter DNA strands is followed in the simulation. During replication, transcription must stop to allow

[d] At this developmental stage in *C. elegans*, transcription is yet to be activated [236].

passage of the replication fork [277]. Therefore DNA replication is also modelled as a discrete event that occurs instantaneously.

4.2.2 *Bistability in the non-processive model*

For each set of parameters, 100 loci were initialised in each of the uniform me0 or me3 states and simulated for 50 cell cycles. DNA replication was included every 22 hours. Similar to simulations of the two-state model, the quantities P_{ON}, P_{OFF}, and B were measured from simulations. After fixing the minimum firing rate f_{min} and the noisy methylation and demethylation rates γ_{me}, γ_{dem}, there are four free parameters: k_{me}, p_{dem} p_{ex} and f_{max}. Figure 4.4 shows the bistability measure, B for simulations performed over a range of values of the free parameters. In contrast to the two-state model, bistability is observed when methylation and demethylation processes are balanced and the fold-change in firing rate between the active and repressed states F is sufficiently large. Increasing histone turnover is not intrinsically detrimental to bistability for the parameter values assessed here, provided it is balanced by a corresponding increase in methylation rate k_{me}. However, it can be seen that the minimum methylation rate for which bistability is obtained increases as histones are exchanged more often. Intuitively, this is because for lower methylation rates, H3K27me2/me3 is not being replaced quickly enough to counteract histone replacement and DNA replication.

Average replication-independent histone lifetimes were also computed directly from these simulations and are shown in Figure 4.5. Histone lifetimes are computed directly from simulations as the total simulation time divided by the average number of histone exchange events per histone. It can be seen that the histone turnover rates are dependent on both p_{ex} and f_{max}(Fig. 4.5). F provides an upper bound on the fold-change in histone lifetimes between the active and repressed states, because histone turnover is coupled directly to transcription. To break this linear coupling would require a more complicated function relat-

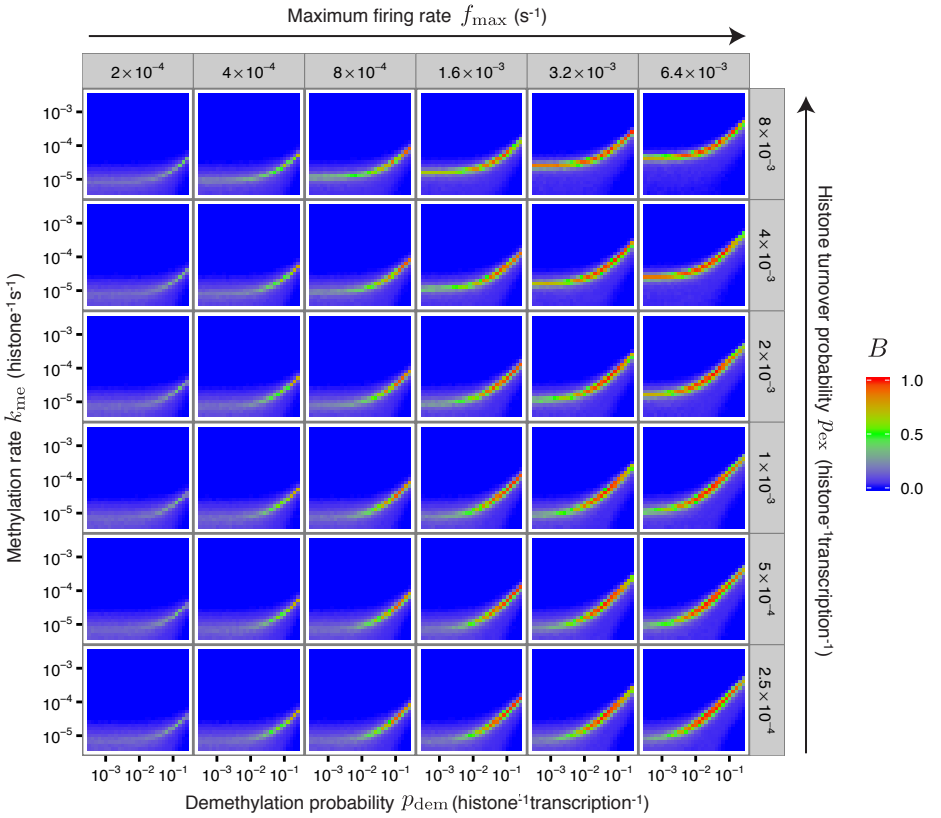


Figure 4.4: Bistability in the non-processive model. Heat map showing the bistability measure, $B = 4P_{OFF}P_{ON}$ for simulations performed with the parameters k_{me} (histone^{-1}s^{-1}), p_{dem} (histone^{-1}transcription^{-1}), p_{ex} (histone^{-1}transcription^{-1}) and f_{max} (s^{-1}) indicated. Each panel shows B as a function of k_{me} and p_{dem}, for a pair of (f_{max}, p_{ex}) values shown in the panel labels to the top and right.

ing transcriptional firing and histone turnover. For further simulations the parameter values $F = 40$ and $p_{ex} = 10^{-3}$ histone^{-1}transcription^{-1} were chosen. These parameters generate almost negligible replication-independent turnover in the repressed state, but histone lifetimes comparable with the length of the cell cycle in the active state. If it is assumed that histone H3.3 is incorporated in place of H3.1 during these transcription-coupled turnover events, and that H3.1 is incorporated exclusively during replication [270], then with these parameter values the model shows accumulation of H3.3 in the active but not the repressed state (Fig. 4.6A,B). By tracking the methylation status on the histones that are removed by histone turnover processes, it was observed

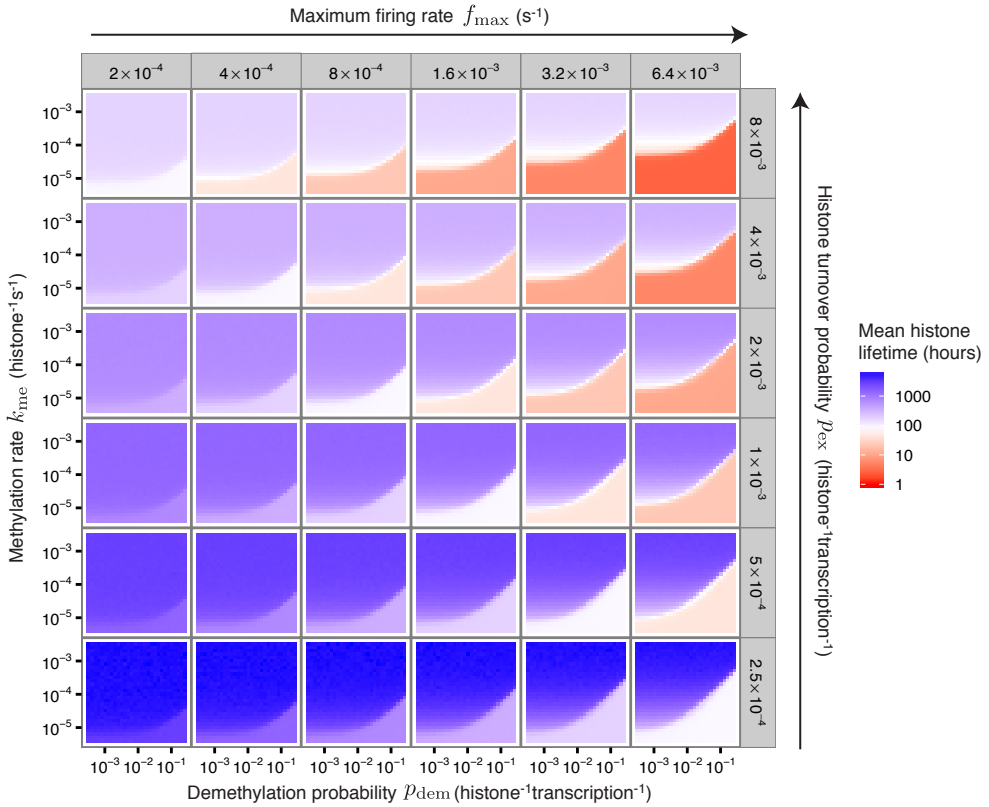


Figure 4.5: Histone lifetimes in the non-processive model. Heat map showing mean histone lifetime, for simulations performed with the parameters k_{me} (histone^{-1}s^{-1}), p_{dem} (histone^{-1}transcription^{-1}), p_{ex} (histone^{-1}transcription^{-1}) and f_{max} (s^{-1}) indicated. Histone lifetime is computed directly from simulations as the total simulation time in hours divided by the average number of histone exchange events per histone. Each panel shows histone lifetime as a function of k_{me} and p_{dem}, for the pair of (f_{max}, p_{ex}) values shown in the panel labels to the top and right.

that the majority of histones are in low H3K27 methylation states when they are lost from chromatin (Fig. 4.6C). This means that although histone turnover can, in principle, cause conversion of me3 to me0, this actually seldom occurs because histone turnover rates are correlated with transcription and therefore anti-correlated with me2/me3 levels.

In summary, linearity between transcription and histone turnover allows model bistability with plausible histone turnover rates. The model is able to reproduce two qualitative experimental results: transcription-dependent H3.3 accumulation and negligible histone turnover in the repressed state.

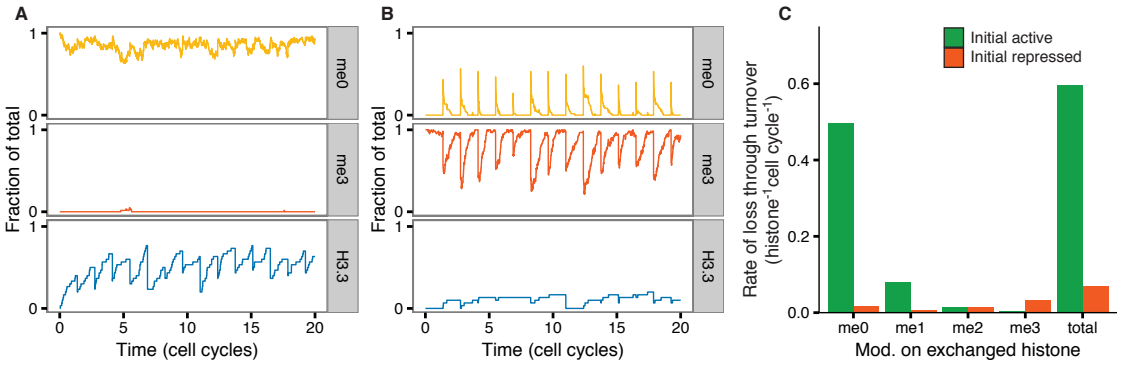


Figure 4.6: Validation of histone turnover rate. Example simulations showing levels of me0, me3 and histone variant H3.3 over time for (A) an active (me0) or (B) repressed (me3) initial state. Parameter values are $k_{me} = 3 \times 10^{-5}$ histone^{-1}s^{-1}, $f_{max} = 40 f_{min}$ s^{-1}, $p_{dem} = 2 \times 10^{-2}$ histone^{-1}transcription^{-1}, and $p_{ex} = 10^{-3}$ histone^{-1}transcription^{-1}. Only the repressed state shows the characteristic 'halving' of histone modifications during DNA replication, because me0 marks are randomly inserted at DNA replication. (C) The calculated rate of loss of specific histone modifications (histone^{-1}cell cycle^{-1}) through histone turnover events, for systems initialised in the active or repressed state. 'total' shows the overall rate independent of H3K27 modification state, while 'me0', 'me1', 'me2', 'me3' show the rates of loss for the individual modification states. Results are averaged over 20 cell cycles for 1000 loci initialised in each of the uniform me0- and me3-states.

4.2.3 *Determining the in vivo K27-methylation rate*

To further constrain parameter values, the model was fit to published quantitative mass spectrometry data. Time-resolved stable isotope labelling by amino acids in cell culture (SILAC) together with nascent chromatin capture was recently used to study the kinetics of histone post-translational modification accumulation after DNA replication [26]. Consistent with prior studies [267, 278, 279], these data demonstrate that H3K27me3 accumulates extremely slowly on newly incorporated histones. In fact, H3K27me3 does not reach steady-state within a cell cycle [26]. Most previous mathematical models of histone-modification-based epigenetic memory have employed time-scales for histone modification reactions which are orders of magnitude faster than this [28, 83, 98, 99, 103, 213, 230, 232]. Typically, each histone tail undergoes many modification reactions within a cell cycle. This fast time-scale is motivated by the assumption that models need to be robust to high levels of

noisy histone turnover and also recover rapidly after DNA replication. In the model presented here, histone turnover rates in the high H3K27-me3 state are very low, allowing the possibility of providing robust epigenetic memory and also reproducing the observed slow time-scale of H3K27me3 accumulation.

The triple-SILAC experiment used to fit the non-processive model is illustrated schematically in Fig. 4.7A. 'Old' histones (yellow) are distinguishable by mass differences from 'new' histones (blue) and 'unlabelled' histones (grey)[e]. New histones are incorporated during the first DNA replication, at which time newly-synthesized DNA is also labelled to allow specific isolation of this 'nascent' chromatin at different time points after the first DNA replication [26]. In the experiments, cells underwent a further two DNA replications in the 48 hours after the initial release into S-phase. Accordingly, the levels of both new and old histones on the nascent chromatin were diluted approximately 4-fold by incorporation of unlabelled histones [26]. To determine whether the non-processive model could fit this data, a computational simulation of the triple-SILAC experiment was performed. The relative levels of H3K27me3 on old and new histones were extracted from the simulations 0, 10, 24 and 48 hours after replication, and were directly compared with the experimental data. Further details on the fitting procedure are provided in Sec. 4.6. Strikingly, the model was unable to fit the slow rate of H3K27me3 accumulation whilst simultaneously providing robust epigenetic memory (Fig. 4.7B,C). The problem with fitting the model to this data was that the model requires 100% H3K27me3 coverage to achieve maximal gene repression (see the equation for firing rate, f, in Fig. 4.3). On the slow time-scales required to fit the SILAC data, H3-K27me3-saturation is not achieved within a cell cycle and consequently transcriptional repression is eventually lost after a few cell cycles.

The qualitative experimental result that H3K27me3 levels continue to increase on histones more than a cell cycle after they are incorporated

[e]These three distinguishable species: old, new, and unlabelled, make this a *triple*-SILAC experiment.

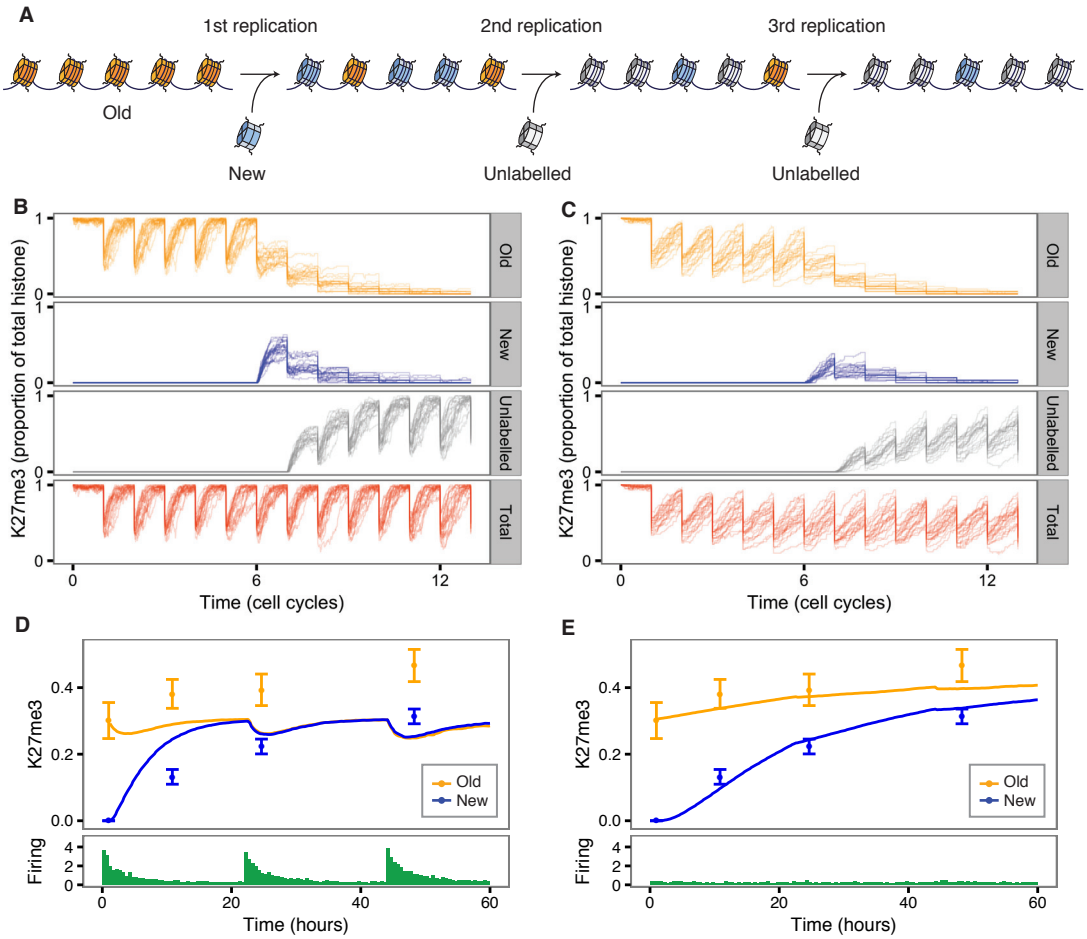


Figure 4.7: Fitting the model to slow H3K27me3 accumulation. (A) Schematic illustration of the triple-SILAC experiment used to determine the H3K27me3 accumulation rate. 'Old' histones (yellow) are diluted by incorporation of 'new' histones (blue) at the first DNA replication. Chromatin from this replication is then followed through another two DNA replications. Both old and new histones are diluted by 'unlabelled' histones during the second and third replications. (B,C). Over-plotted trajectories for 20 loci from computational simulations of the triple-SILAC experiment. H3K27me3 levels as a proportion of total histone are shown for old, new, unlabelled, and total histones. B shows the slowest bistable model with $P_T = 1.0$, while C shows the best-fit model with $P_T = 1/3$. (D,E). H3K27me3 levels on old and new histones as a proportion of old and new histone incorporated, respectively. Points are experimental data from [26], with error bars representing s.e.m. ($n = 3$). Solid lines show the results of simulations shown in B and C averaged over 1000 loci. Simulation data are normalised so that the mean cell-cycle end value of H3K27me3 obtained from the simulation is equal to the experimental mean initial level on old histones, as described in Sec. 4.6.3.

in chromatin [26] suggests that there are many non-K27me3 H3 tails present within a H3K27me3 domain. If the H3K27me3 marks within such a domain are also responsible for gene repression, this implies that maximal gene repression must be achieved without H3K27me3 on every histone tail of the gene. While chromatin immunoprecipitation (ChIP) experiments are useful for understanding the genome-wide distributions of post-translational modifications, they can provide only information on the relative levels of such modifications. The absolute proportion of histones tails carrying a H3K27me3 mark within a H3K27me3-enriched domain has therefore remained enigmatic. To allow transcriptional repression at sub-saturating H3K27me3 levels, a threshold proportion of H3K27me2/me3 marks, $P_T \leq 1$ was introduced into the model. The transcription firing rate f was reformulated to linearly scale between $f_{\min}$ and $f_{\max}$ only for H3K27me2/me3 levels below the threshold,

$$f = \begin{cases} f_{\max} - \frac{P_{\text{me2/3}}}{P_T}\left(f_{\max} - f_{\min}\right) & , P_{\text{me2/3}} < P_T, \\ f_{\min} & , P_{\text{me2/3}} \geq P_T, \end{cases} \tag{4.2}$$

where,

$$P_{\text{me2/3}} = \frac{1}{N}\sum_{i=1}^{N}\left(\delta_{S_i,\text{me2}} + \delta_{S_i,\text{me3}}\right). \tag{4.3}$$

As anticipated, including this threshold caused the region of bistability in parameter space to extend to much lower values of k_{me} (Fig. 4.8B). Using the fixed parameter values shown in Table 4.2, simulations were performed for a range of values of P_T, k_{me} and p_{dem}. Fig. 4.8A,B shows the mean cell-cycle-end value of H3K27me3 and the bistability measure, B, obtained from these simulations. The sum of squared errors (SSE) between all experimental data points and the corresponding model prediction was used to quantitatively characterise the model fit (Fig. 4.8C). The best fit to the data was obtained for $k_{\text{me}} = 8{\times}10^{-6}$ histone^{-1}s^{-1}. However, models with $P_T > 0.6$ did not generate robust bistability at this low methylation rate. The model fit to the data over a range of parameters for $P_T = 1/3$ is shown in Fig. 4.9. Figures 4.7B,D show stochastic sim-

ulations for 20 loci and average results fitted to the data for a bistable model with $P_T = 1$ ($k_{me} = 3 \times 10^{-5}$ histone^{-1}s^{-1}). In contrast, Figures 4.7C,E show the fit obtained with $P_T = 1/3$ ($k_{me} = 8 \times 10^{-6}$ histone^{-1}s^{-1}). In both cases p_{dem} remains a free parameter and was chosen to maximise B ($p_{dem} = 0.02$ and 0.004 histone^{-1}transcription^{-1}, respectively for Fig. 4.7B/D, and C/E). The quantitative fit to the data is greatly improved for the model incorporating the firing threshold. However, more importantly, when the threshold, P_T is included, the model is able to recapitulate the qualitative experimental result that H3K27me3 remains under-represented on new histones versus old histones for more than a cell cycle (Fig. 4.7C,E).

How H3K27me3 can reach the same level at the end of each cell cycle on total chromatin without ever reaching steady-state is somewhat counterintuitive. Consideration using some actual numbers for the average deterministic case is helpful. Assume a gene containing 30 nucleosomes (60 H3 tails), has approximately 2/3 of its histones carrying H3K27me3 at the end of each cell cycle (40 total H3K27me3 marks). Immediately after DNA replication, on average only 20 of these H3K27me3 marks will remain associated with each daughter chromosome, and all of these will be on old histones. Therefore after replication, 20 out of 30 old histones, and 0 out of 30 new histones will carry H3K27me3. H3K27me3 will then be slowly added to the 30 new histones and also the 10 old histones that are still devoid of H3K27me3. To reach the same level of 40 H3K27me3 marks in total chromatin at the end of each cell cycle, a total of 20 H3K27me3 marks must be added to the inherited 20 marks from the previous cell cycle. Therefore, assuming no distinction between old and new histones, these 20 marks will be divided proportionally between the 30 new and 10 unmodified old histones[f]. Therefore, on average, 15 H3K27me3 marks will be added to the new histones and 5 marks will be added to the old histones. Thus at the

[f]For the non-processive model presented here, the old histones are actually more likely to accumulate H3K27me3 than the new histones because some of the old histones will already be marked with H3K27me2 from the previous cell cycle. This complication is neglected for clarity of the explanation.

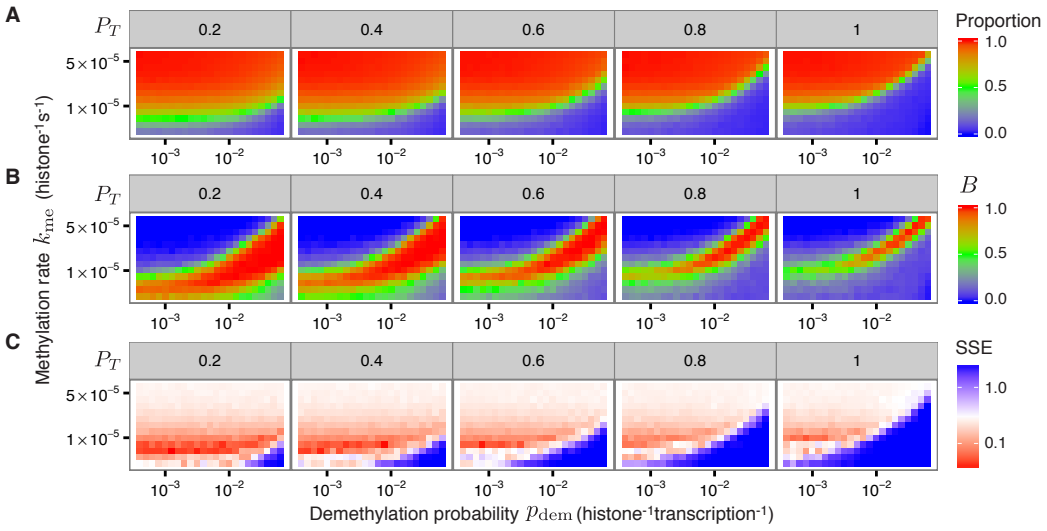


Figure 4.8: Introduction of a firing threshold P_T. (A) Average proportion of H3K27me3 at the end of each cell cycle. (B) Bistability measure, B. (C) Sum of squared errors (SSE) between the experimental SILAC data and the model prediction, for a computational simulation of the triple-SILAC experiment, similar to the example in Fig. 4.7. The fitting procedure is described in Sec. 4.6.3. All values calculated from simulations of the non-processive model with DNA replication, for a range of p_{dem} and k_{me}. Each panel represents a specified value of P_T, the threshold level for H3K27me2/me3 above which transcriptional firing rate is equal to f_{min} (Eq. 4.2). $f_{max} = 40 f_{min}$ and $p_{ex} = 10^{-3}$ histone^{-1}transcription^{-1}. Other parameter values are shown in Table 4.2.

end of the cell cycle, the 30 old histones now carry 25 H3K27me3 marks, while the 30 new histones carry just 15 H3K27me3 marks. Nonetheless, the total number of H3K27me3 remains at 40 out of 60 histones, just as it was at the end of the previous cell cycle.

In summary, slow increases in H3K27me3 levels within a H3K27me3-enriched domain imply that H3K27me3 levels are not saturated throughout the cell cycle. By allowing a non-saturated H3K27me3 domain to repress transcription equally effectively as a saturated domain, the non-processive model can provide epigenetic memory of both active and repressed states through many cell divisions and simultaneously fit the slow accumulation of H3K27me3 over several cell cycles.

However, it is important to remember that these data represent averages over the entire genome. It is therefore not trivial that the average time-scale extracted through the analysis presented here reflects that

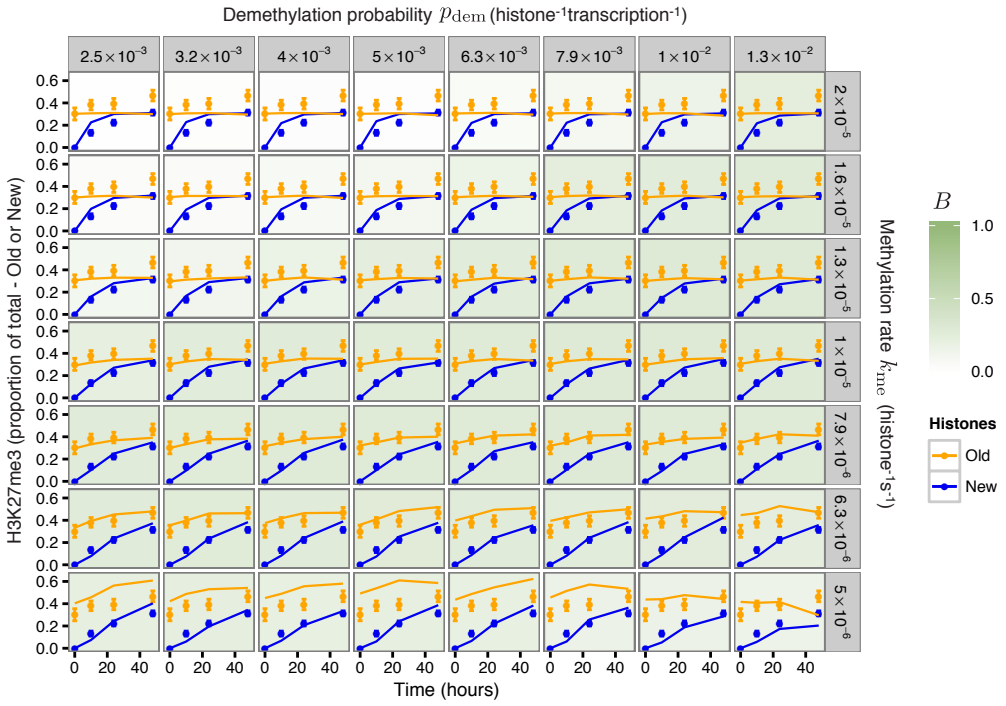


Figure 4.9: Fitting H3K27me3 dynamics from the triple-SILAC experiment. Each panel shows the experimentally determined H3K27me3 level on old and new histones 0, 10, 24 and 48 hours after incorporation into nascent chromatin. Data are represented as the fraction of H3K27me3 on old and new histones, respectively. Solid lines indicate the model prediction, linearly interpolated between 0, 10, 24, and 48 hour time-points, which are each averages over 1000 loci. Simulation data are normalised so that the mean cell-cycle end value of H3K27me3 obtained from the simulation is equal to the experimental mean initial level on old histones, as described in Sec. 4.6.3. Each panel shows the results of simulations for a single pair p_{dem}, k_{me} parameters. p_{dem} increases from left to right, while k_{me} increases from bottom to top. $P_T = 1/3$, $f_{\mathrm{max}} = 40 f_{\mathrm{min}}$ and $p_{\mathrm{ex}} = 10^{-3}$ histone^{-1}transcription^{-1}. Other parameter values are shown in Table 4.2. The background shading of the panel (green) represents the bistability parameter, B, calculated using the same parameters.

of a gene which is actually depending on H3K27me3 for gene repression. For this reason, faster H3K27-methylation dynamics (similar to Fig. 4.7B) cannot be excluded in all cases.

4.2.4 *A role for slow dynamics in buffering regulatory noise*

The presence of regulatory factors at their target genes can vary over time. This occurs due to noisy expression of the regulatory factors themselves [280–282] and also due to fluctuations in the occupancy of these factors at their targets [283, 284]. This section contains an investigation of the hypothesis that the transcription/chromatin state could help to buffer these fluctuations and thereby provide a less variable transcription output.

Incorporating the transcriptional firing rate directly in the chromatin-based epigenetic model allows consideration of the interplay between trans and cis regulation. To this end, it is necessary to introduce external transcriptional control into the non-processive model developed above. External trans-regulation is included as a multiplicative factor $\alpha(t)$ in the firing rate function,

$$f = \begin{cases} \alpha(t)\left(f_{\max} - \frac{P_{\mathrm{me2/3}}}{P_T}\left(f_{\max} - f_{\min}\right)\right) & , P_{\mathrm{me2/3}} < P_T, \\ \alpha(t) f_{\min} & , P_{\mathrm{me2/3}} \geq P_T. \end{cases} \tag{4.4}$$

$\alpha(t)$ can be interpreted as externally driven 'gene activation'. $\alpha = 1$ is 'neutral' and gives the firing rate function used earlier (Eq. 4.2). In addition to this function, the firing rate is limited to once per minute on average ($f \leq 1/60\ \mathrm{s}^{-1}$) to simulate a physical limitation to maximum transcription rate.

A stochastic model of gene expression[g] [285] was used to simulate a variable concentration of a transcriptional regulator $r(t)$. The concentration $r(t)$ was normalised to give a time-dependent gene activation function, $\alpha(t) = r(t)/\langle r(t)\rangle$, where $\langle\rangle$ indicates an average over time. $\alpha(t)$ was then used to calculate the firing rate according to Eq. 4.4.

[g] Model and parameter values are described in Sec. 4.6.4.

Keeping $\langle r(t) \rangle = 1000$ molecules and varying other parameters in this model allows modulation of the variability of the signal without affecting its mean (i.e. $\langle \alpha(t) \rangle = 1$) . The 'noisiness' of the signal was quantified from the simulations as the coefficient of variation, $\mathrm{CV} = \sigma / \langle \alpha(t) \rangle = \sigma$, where σ is the standard deviation. With this noisy input function, simulations were again performed over a range of k_{me} and p_{dem}. To quantify how well the chromatin is able to buffer noise in the input signal, the combined first passage time measure, *FP* was used. *FP* is defined in Sec. 4.6 (Eq. 4.14) and ranges from 0 to 1. $FP = 1$ indicates that no state changes occurred throughout the simulation for both the initial active and initial repressed states, while $FP = 0$ indicates that at least one of the states was unstable.

Strikingly, systems with fast dynamics that were bistable ($FP \approx 1$) when noise was low, showed a marked decrease in *FP* as noise was increased (Fig. 4.10) while systems with slower dynamics had a greater ability to buffer noisy inputs. This contrasts sharply with the state lifetimes observed for constant α, where fast systems are generally preferred due to their ability to rapidly recover from the perturbations of histone turnover and DNA replication. In fact, when the model is parameterised to fit the slow accumulation of H3K27me3 in the SILAC data (Sec. 4.2.3), there is an impressive ability to buffer noise in the input signal, when compared to fast systems (Fig. 4.10). The increase in stability for slow systems was observed for both repressed (Fig. 4.11) and active (Fig. 4.12) initial states. The modelling therefore suggests a possible explanation for why H3K27me3 dynamics are so slow: genes that change H3K27me3 levels slowly when trans-factor inputs are fluctuating offer an advantage over genes with faster chromatin dynamics because neither prolonged absences nor pulses of regulatory trans-factors are sufficient to change chromatin states.

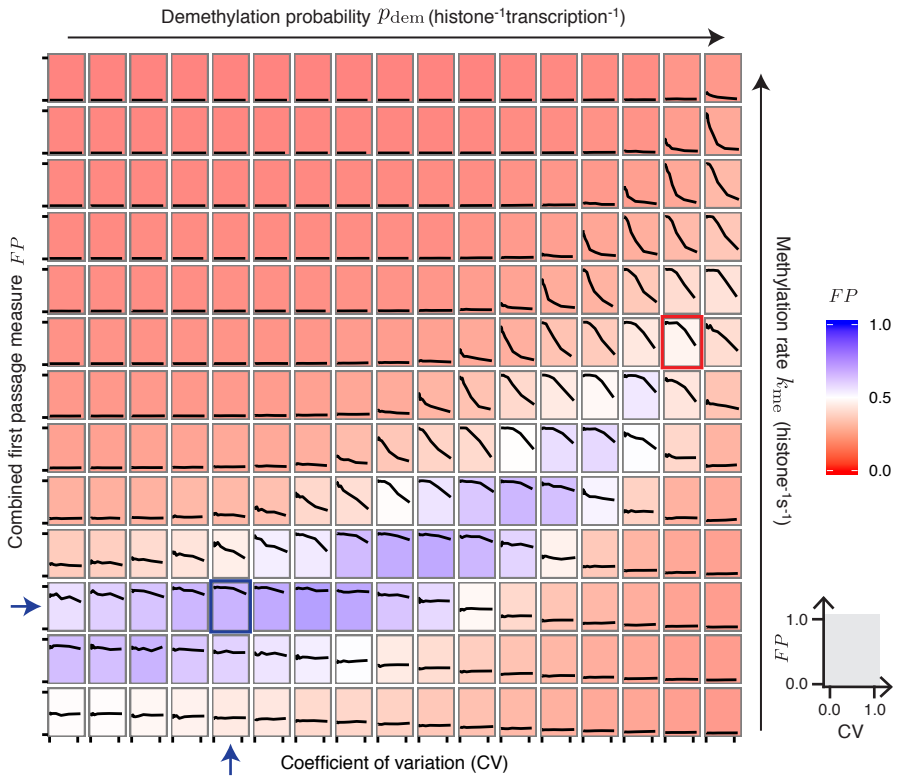


Figure 4.10: Effect of noisy input signal on memory-storage capability. Each panel shows the combined first passage time measure, *FP* as a function of the noisiness of the gene activation input signal $\alpha(t)$, as measured by the coefficient of variation (CV). The schematic panel shown in grey to the right indicates the values of *FP* and CV, represented by the axis ticks in each panel. Each panel represents simulations performed using different values k_{me} and p_{dem}. k_{me} increases from bottom to top (log scale with range $\approx [10^{-6}, 10^{-4}]$ histone^{-1}s^{-1}) while p_{dem} increases from left to right (log scale with range $\approx [10^{-3}, 10^{-1}]$ histone^{-1}transcription^{-1}). The k_{me} value determined by fitting the SILAC data (Sec. 4.2.3), and the corresponding p_{dem} value chosen for bistability are indicated with blue arrows and a blue box. These k_{me}, and p_{dem} values are used for example simulations of 'Slow chromatin dynamics' in Figs. 4.11, 4.12. The red box represents the k_{me}, and p_{dem} values used for example simulations of 'Fast chromatin dynamics' in Figs. 4.11, 4.12. The background colour of each panel represents *FP* for a noisy input signal (CV $\approx$ 1), as indicated in the colour guide.

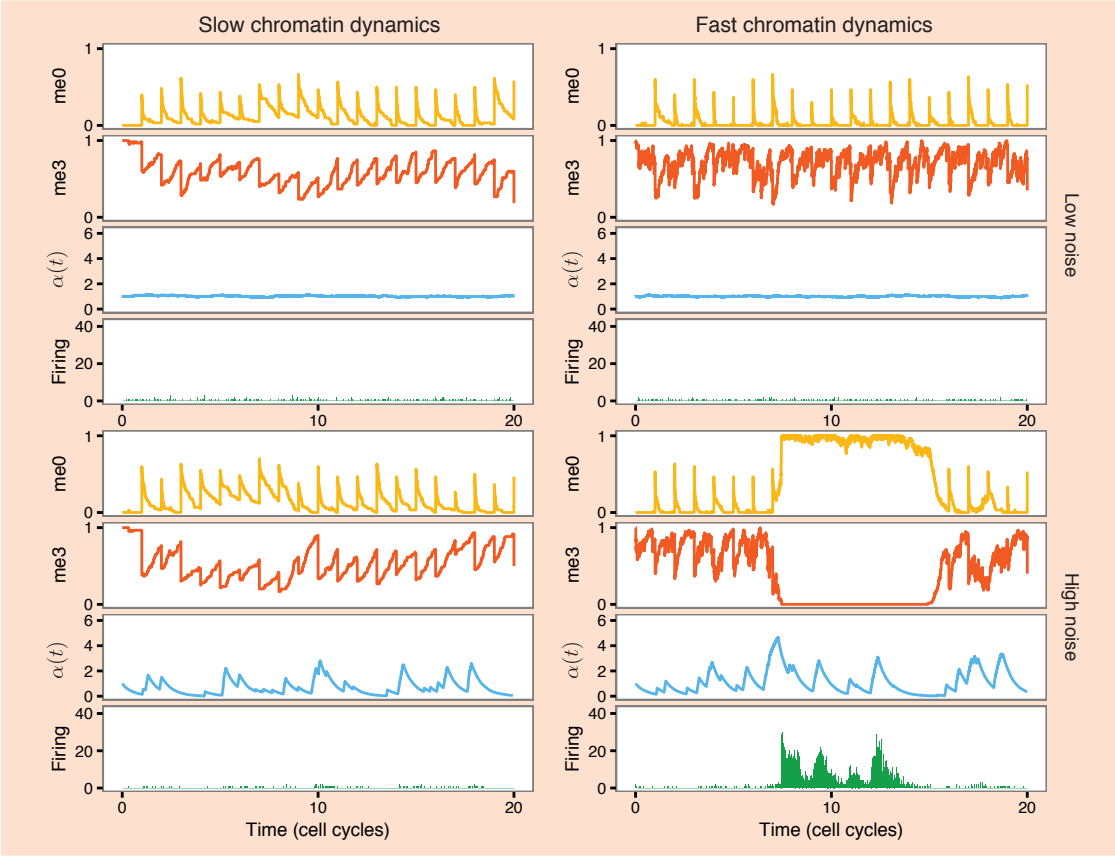


Figure 4.11: Noise buffering in the repressed state. Simulations of individual loci initialised in the repressed state with variable transcriptional activation signals $\alpha(t)$ (shown). $\alpha(t)$ has low noise in upper plots ($b = 1$) and high noise in lower plots ($b = 1000$), where b is defined in Sec. 4.6.4. Left plots show the results for a model with slow dynamics ($k_{\mathrm{me}} = 8 \times 10^{-6}$ histone^{-1}s^{-1}, $p_{\mathrm{dem}} = 4 \times 10^{-3}$ histone^{-1}transcription^{-1}). Right plots show a model with fast dynamics ($k_{\mathrm{me}} = 4 \times 10^{-5}$ histone^{-1}s^{-1}, $p_{\mathrm{dem}} = 2 \times 10^{-1}$ histone^{-1}transcription^{-1}). In both cases, $P_T = 1/3, f_{\mathrm{max}} = 40 f_{\mathrm{min}}, p_{\mathrm{ex}} = 10^{-3}$ histone^{-1}transcription^{-1}. *FP* as a function of noise strength for these 'slow' and 'fast' parameter sets are highlighted in Fig. 4.10.

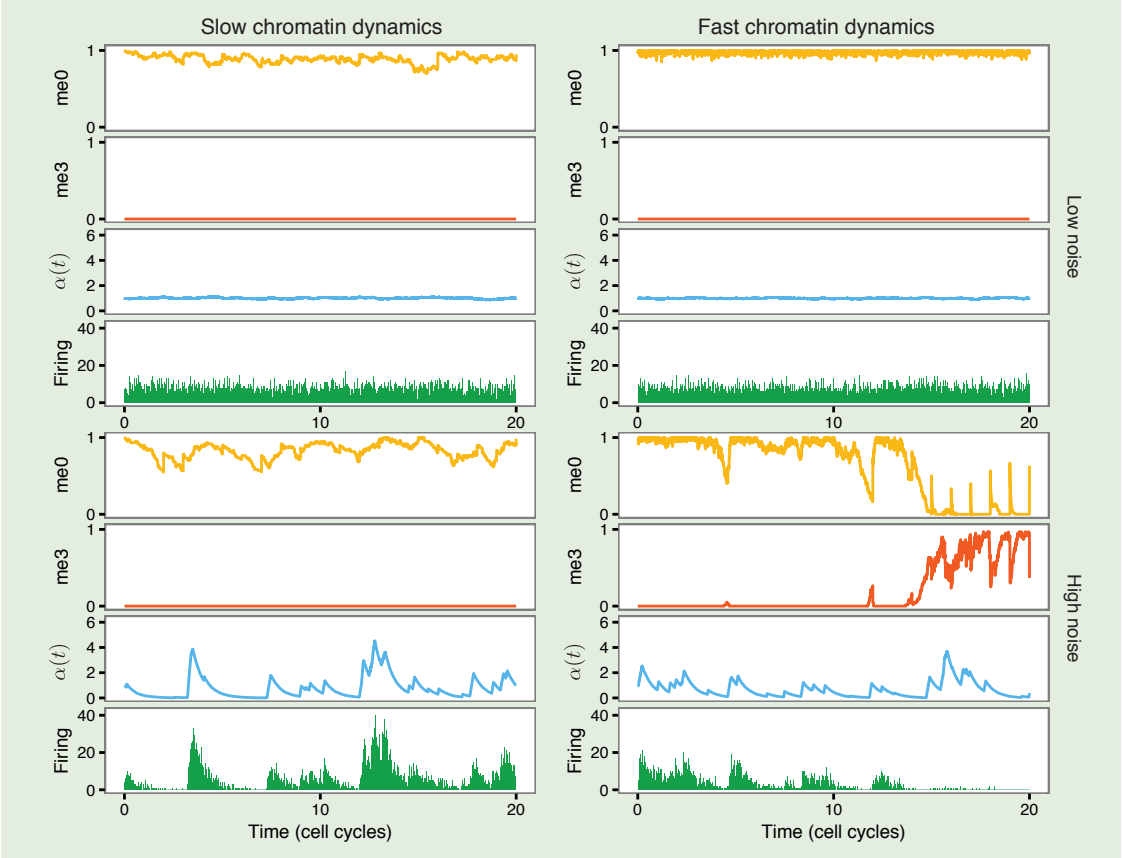


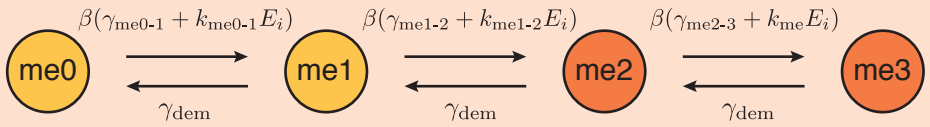
Figure 4.12: Noise buffering in the active state. Simulations of individual loci initialised in the active state with variable transcriptional activation signals $\alpha(t)$ (shown). Details provided in Fig. 4.11 caption.

4.3 Dynamic regulation of chromatin

Having established a role for slow chromatin dynamics in buffering noise in trans-acting regulators, it is interesting to consider how the model can be switched between states by either directly modifying transcription or chromatin. This is particularly relevant to the observation that H3K27me3 can accumulate when PRC2 target genes are repressed by direct perturbation of transcription.

All simulations in this section incorporate the non-processive model with firing threshold, using parameter values as fitted to the experimental data in Section 4.2.3. However, to also allow modulation of the local PRC2 activity away from the fitted value of k_{me}, the noisy and di-

Histones



$$E_i = \sum_{j \in \text{neighbours of } i} \left(\rho_{\text{me2}}\delta_{S_j,\text{me2}} + \delta_{S_j,\text{me3}}\right) \qquad S_j \in \{\text{me0}, \text{me1}, \text{me2}, \text{me3}\}$$

Transcription

$$P_{\text{me2/3}} = \frac{1}{N}\sum_{j=1}^{N}\left(\delta_{S_j,\text{me2}} + \delta_{S_j,\text{me3}}\right)$$

$$f = \begin{cases} \alpha\left(f_{\text{basal}} - \frac{P_{\text{me2/3}}}{P_T}\left(f_{\text{basal}} - f_{\text{min}}\right)\right) & , P_{\text{me2/3}} < P_T \\ \alpha f_{\text{min}} & , P_{\text{me2/3}} \geq P_T \end{cases}$$

$$\text{With probability } p_{\text{dem}}: \quad S_j \rightarrow \begin{cases} \text{me0} & \text{if } S_j = \text{me1} \\ \text{me1} & \text{if } S_j = \text{me2} \\ \text{me2} & \text{if } S_j = \text{me3} \end{cases}$$

$$\text{With probability } p_{\text{ex}}: \quad S_j S_{j+1} \rightarrow \text{me0}_j\text{me0}_{j+1}, \text{ for } j = 1, 3, 5, \ldots, N-1$$

DNA replication

$$\text{With probability } \frac{1}{2}: \quad S_j S_{j+1} \rightarrow \text{me0}_j\text{me0}_{j+1}, \text{ for } j = 1, 3, 5, \ldots, N-1$$

Figure 4.13: Non-processive model with variable α and β. Mathematical description incorporating α, β and firing rate threshold P_T. Parameter values are given in Table 4.3. Similarly to Fig. 4.3, the sum over 'neighbours' in E_i includes the other histone on the same nucleosome, and the four histones on neighbouring nucleosomes. In addition to the firing rate function f, the firing rate is limited to once per minute on average ($f \leq 1/60\ \text{s}^{-1}$) to simulate a physical limitation to maximum transcription rate.

rected methylation rates are multiplied by β ($k_{\text{me}} \rightarrow \beta k_{\text{me}}$, $\gamma \rightarrow \beta\gamma$). For $\alpha = \beta = 1$, the model corresponds to the unstimulated case studied earlier. For completeness, a mathematical description of the model is provided in Fig. 4.13. Since the actual firing rate is now dependent on α, f_{max} in Eq. 4.4 is rewritten as f_{basal} (Fig. 4.13), representing the maximum firing rate in the unstimulated 'neutral' case ($\alpha = 1$).

After initialisation in either the repressed or active state and pre-equilibration of the model for 5 cell cycles with α=β=1, one of the α or β parameters was then changed permanently and the time-evolution of the chromatin state studied. Changes to α simulate recruitment of

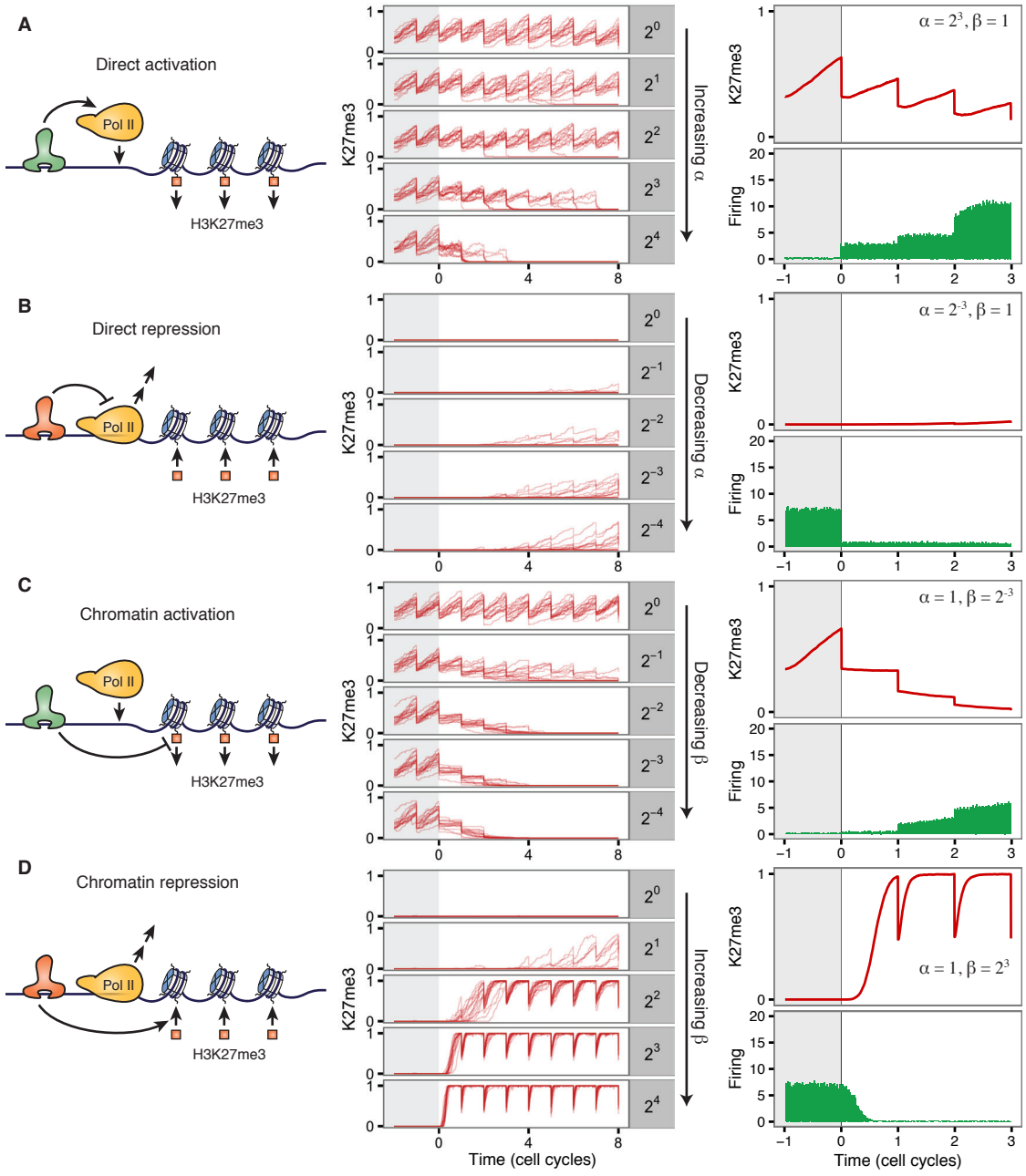


Figure 4.14: Dynamic changes to chromatin and transcription. (A) Direct activation, (B) Direct repression, (C) Chromatin activation, (D) Chromatin repression. Left panels show a schematic of the process being simulated. Central panels show over-plotted H3K27me3 levels over time for 20 loci simulated after a dynamic change to either α (A, B) or β (C, D). Parameter change is incorporated after 5 cell cycles of equilibration at $\alpha = \beta = 1$ ($t = 0$) and then simulated for a further 8 cell cycles. Right panels show population-average H3K27me3 levels for 3 cell cycles after activation or repression and corresponding firing rate changes. Transcription is measured in number of events per 30 minute interval.

Parameter	Description	Value
α	Externally-driven transcription activation	varied around 1
β	PRC2 recruitment bias	varied around 1
k_{me}	Directed methylation rate (me2 to me3) (histone^{-1}s^{-1})	8×10^{-6}
f_{basal}	Basal firing rate (s^{-1})	$20 f_{\text{min}}$
p_{dem}	Transcription-induced demethylation rate (histone^{-1}transcription^{-1})	8×10^{-3}
p_{ex}	Transcription-induced histone turnover rate (histone^{-1}transcription^{-1})	10^{-3}
P_T	Fraction of H3K27me2/me3 required for gene repression	1/3

Table 4.3: Parameters for dynamic chromatin regulation. Description of parameters and values used in dynamic simulations of the non-processive model.

an activator or repressor that directly modulates transcriptional activity (direct activation/repression, Fig. 4.14A,B), while changes to β simulate activators or repressors which modify the activity of PRC2 at the gene (chromatin activation/repression, Fig. 4.14C,D). When transcription is activated directly, increases in transcription precede chromatin state changes. Transcription leads to stochastic loss of the silenced state through H3K27-demethylation in the hours following gene activation (Fig. 4.14A, middle panel) and gradual reductions in H3K27me3 at the population level (Fig. 4.14A, right panel). Conversely, when transcription is directly down-regulated from an active initial state (Fig. 4.14B), stochastic switching to the silenced state and accumulation of H3K27-me3 at the population level is very slow, taking several cell cycles. This is due to the slow intrinsic time-scale of H3K27me3 addition. These results are reminiscent of time-course experiments in which it has been observed that accumulation of H3K27me3 occurs slowly after cessation of transcription [111, 121, 286, 287]. In contrast, when the chromatin state is directly modified in the model, via modulation of the rate of H3K27-methylation, changes to the chromatin and transcription states occur synchronously (Fig. 4.14C,D).

The non-processive model was parameterised based on a requirement for bistability, in order that epigenetic memory can be stored lo-

cally in the proportion of histone modifications. With this requirement, it is natural that the chromatin state is resistant to change through transcriptional perturbation. Nonetheless, the model predicts that switching the chromatin state of PRC2 target gene will occur very slowly (taking several cell cycles), if transcription is directly modulated without recruitment of additional chromatin modifiers.

That H3K27me3 follows transcriptional down-regulation has been interpreted as indicating that H3K27me3 is a consequence rather than a cause of gene repression. This conclusion is justified where 'cause' refers to the initiation of transcriptional down-regulation. However, referring to H3K27me3 as a 'consequence of repression' understates its importance in the subsequent maintenance of repression. For example, both the active and repressed states are stably maintained with $\alpha = \beta = 1$. Therefore, in the model, the chromatin state can be switched with a strong, persistent pulse of gene activation or repression and can then be maintained using the internal transcription and chromatin dynamics once this pulse has finished. In this sense, while the chromatin does not 'cause' the initial transcriptional down-regulation, it does 'cause' re-establishment of the repressed state after each DNA replication.

4.3.1 *Integration of cis- and trans-regulation*

The non-processive model of a PRC2 target gene formulated in this chapter has the capability to store epigenetic memory locally in the methylation status of H3K27. However, when the activation level of the gene is dynamically reduced for a gene that is initially expressed, the chromatin state can also respond by accumulating H3K27 (Fig. 4.14B). This is in agreement with experimental results [111], and indicates that the chromatin state is responsive to an externally specified transcription state. For less extreme reductions in activation, however, it was observed that the chromatin state is maintained by the internal chromatin/transcription dynamics. In this latter case, the chromatin may be described as instructive. Taken together, these results suggest that,

depending on the extent of gene activation, the chromatin state can either be responsive or instructive.

To understand this in more detail, simulations were performed at different values of α, starting from either the repressed or active initial state. Simulations are similar to those above (Section 4.3). After equilibration for 5 cell-cycles at $\alpha = \beta = 1$, the value of α was permanently changed and the system was simulated for a further 20 cell cycles. The gene output was then measured as the average number of transcription firing events in the final cell cycle. This 'gene output' is plotted as a function of the 'gene activation' α in Fig. 4.15A, for both active and repressed initial states. For extreme values of α, it can be seen that the transcriptional output is completely independent of the initial chromatin state, with the H3K27 methylation status determined entirely by the concentration of the trans-acting regulators. For intermediate values of α (around 1), however, the transcriptional output can depend strongly on the initial state. For this intermediate range of α values, the chromatin has a tendency to be maintained in its initial state by the internal chromatin/transcription dynamics and, in doing so, determines the transcriptional output of the gene.

This intermediate range of α can be thought of as 'window' of cis memory. In this modelling framework, the difference between responsive and instructive chromatin is therefore not categorical but quantitative. When external inputs to transcription are relatively balanced, the chromatin state can act to regulate gene expression, but when transcription is increased or decreased beyond these limits the chromatin state becomes purely responsive.

The mean first passage time (Eq. 4.13) was also calculated as a function of α for the initially uniform me0 and me3 states. At $\alpha = 1$, it takes on average around 150 cell cycles to change from the me0 to me3 state and vice versa (Fig. 4.15B). As α is increased or decreased, this alteration favours the active or repressed state respectively.

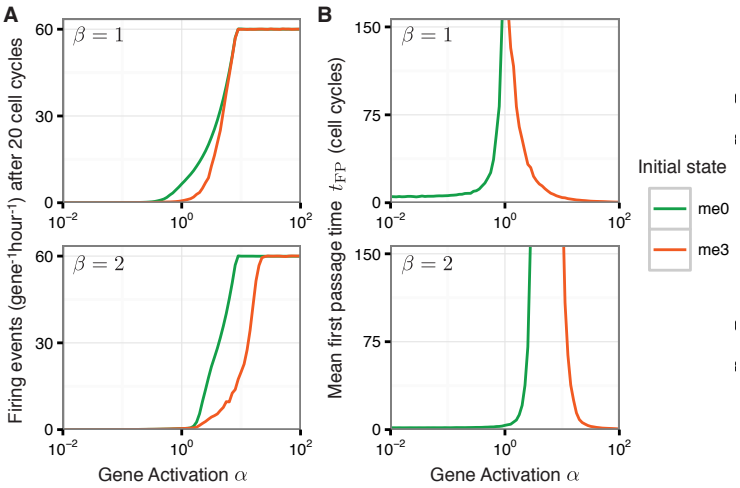


Figure 4.15: Response of gene to transcriptional perturbation. (A) Gene output measured as the average number of transcription events (gene^{-1}hour^{-1}) in the 20th cell cycle after activation or repression. Simulation data averaged over 1000 loci for each value of α. Green lines indicate an initially active gene (me0), while red lines indicate an initially repressed gene (me3). Upper panel: $\beta = 1$ throughout, $\alpha = 1$ during 5 cell-cycle equilibration and then α as indicated on the x-axis for a further 20 cell cycles. Lower panel, $\beta = 2$ throughout, $\alpha = 5$ during 5 cell-cycle equilibration and then α as indicated on the x-axis for a further 20 cell cycles. Transcription firing rate is capped at an average rate of once per minute. (B) Mean first passage time (Eq. 4.13) as a function of α averaged over 640 loci simulated for 1000 cell cycles each, starting from either the active (me0) or repressed (me3) initial state. Other parameter values are given in Table 4.3.

The range of α over which chromatin is instructive could potentially be increased or decreased for different PRC2 target genes, or for the same gene in different cell types, genetic backgrounds or environmental conditions. To illustrate this, simulations were performed as described above, except with a two-fold increase in the local activity of PRC2 ($\beta = 2$). In this case, the transcriptional output shows dependence on the initial chromatin state over a greater range of α, and the difference in gene output between the two initial states occurs at proportionally higher α values. This indicates that chromatin can instruct gene expression over a wider range of transcriptional activation levels (Fig. 4.15B). Furthermore, the mean first passage times were greater for $\beta = 2$ than $\beta = 1$, for both initial states (Fig. 4.15A). In this example, an

increase in local activity of PRC2 was used to illustrate that the ability of chromatin to instruct expression can be quantitatively modulated. The parameter β represents both the enzymatic activity and the propensity of the locus to recruit PRC2. Other factors affecting the 'width' of the 'cis memory window' are the same as those that influence robustness of bistability, such as the number of histones in the gene [28].

Situations could also be imagined in which the cis memory window is so small that chromatin effectively always responds to external transcriptional inputs. For this class of genes, repression is associated with high H3K27me3 but these genes remain repressed only by virtue of a trans-memory system, rather than because of the H3K27me3/PRC2 feedback. It is interesting to consider what proportion of H3K27me3-enriched genes identified in genome-wide studies actually use H3K27 methylation for storage of epigenetic memory.

4.4 Discussion

Large regulatory regions are required for bistability. In the model developed here, it is assumed that every H3K27me2 and me3 mark at the gene contributes equally to reducing the transcription firing rate. Several mechanisms can be imagined for how this could occur. One possibility is that H3K27me2/me3 could drive a dynamic change in the physical position of the gene in the nucleus to a region in which the availability of the transcription machinery is reduced [288, 289]. In this case, the precise location of the marks relative to the transcription start site would not be important in determining their ability to influence the transcription rate. Another possibility is that gene body H3K27me2/me3 could allow binding of factors which mediate a repressive effect on initiation of transcription. This is conceptually similar to the proposed (repressive) function of gene body H3K36me3 in yeast, which directly recruits histone deacetylases [290].

The non-processive model was also tested with smaller regulatory regions (not shown). The increased noise associated with having fewer

components controlling the transcription firing rate led overall to less stable models. This is analogous to the effect of decreasing system size in the M-U-A model (Section 3.3.2).

Mitosis. In this chapter, models were introduced without reference to possible molecular events that occur during the chromosome condensation associated with mitosis. During mitosis, histones are retained at similar locations and their H3K27-methylation status is maintained [27, 236, 279]. It is also known experimentally that transcription is actively repressed [291] and that the majority of Polycomb group proteins dissociate from chromatin [292]. This suggests that both transcription and H3K27-methylation occur with lower probability on condensed chromatin during mitosis. Based on these data, it is assumed that mitosis does not substantially bias the model towards either the high or low transcription state. With this assumption, mitosis effectively represents a 'pause' in the state of the system and is therefore not included in the model.

Local interactions are sufficient for bistability. Previous theoretical models of histone-modification-based epigenetics have concluded that bistability requires modified histones to recruit enzymatic complexes that act beyond neighbouring nucleosomes [28]. For example, in a three-state M-U-A model with purely local interactions, 'patches' of histones in one state can accumulate in a region predominantly covered by the opposing mark. This leads to long-lived intermediate states and therefore poor bistability. Action beyond neighbouring histones is referred to as 'long-range interactions', and are commonly attributed to DNA looping or higher-order chromatin structure [28, 98], which allows distant nucleosomes in the one-dimensional chain to actually be in close physical proximity. While this proposal is plausible, particularly if the 'interaction length' is reasonably short, it becomes more problem-

atic to prevent such models from exhibiting uncontrolled spreading of histone modifications [98].

In contrast to earlier models, the non-processive model formulated here contains only local interactions between histones and their modifying complexes. That is, PRC2 recruited to one nucleosome can act only on its neighbours. The reason that bistability is still observed in this model is two-fold. First, the model contains no self-reinforcing opposing mark, so the problem of an opposing mark nucleating and spreading within a repressed domain does not exist. Second, although histone modifications recruit complexes that act only on neighbouring nucleosomes, the opposing state of transcription can act anywhere within the gene and thereby effectively generates a demethylation rate determined by the average chromatin state of the entire gene. In this sense, the process of transcription and the mechanism by which it is regulated by H3K27me fulfil the requirement for long-range interactions in this model. Nonetheless, the mechanism outlined here has considerable advantages. Firstly, the chromatin state of the entire gene is naturally coordinated by the process of transcription, which prevents one part of the gene from adopting a repressive configuration while another part is active. Secondly, the DNA sequence used to control the initiation and termination of Pol II can also be used to naturally delimit the boundaries of chromatin-activation by transcription.

The bistable model presented here, which requires only local recruitment and action of PRC2 makes the possibility of insulator elements at the ends of chromatin domains a plausible mechanism for limiting the spread of H3K27me2/me3. PRC2 could also be prevented from spreading into highly-transcribed neighbouring genes through the intrinsic properties of transcription that antagonise Polycomb silencing.

As an alternative to insulators, it is also interesting to consider the possibility that the region of chromatin which has the capacity to recruit PRC2 and H3K27me2/me3 could also be defined by the process of transcription itself, which occurs infrequently in the repressed state.

That is, if transcription were to somehow mark the chromatin template as a pre-requisite for H3K27 methylation by PRC2, then H3K27me3 would be naturally limited to the transcription unit. This limitation of H3K27me3 to the transcribed region is known to occur genome-wide in Arabidopsis [293] and also at *FLC* (Fig. 3.1) [83, 123]. H3K27me3 domains in mammals are more variable but enrichment over genic regions is nonetheless observed [294].

'Active' chromatin marks. The process of transcription was shown to constitute sufficient antagonism of PRC2 silencing for robust stability of both the active and silenced states. However, there is considerable molecular and genetic evidence that Polycomb repression is antagonised by the Trithorax group of proteins [295, 296]. This antagonism is thought to occur in part through the histone modifications catalysed by the Trithorax group, which include H3K4 and H3K36 methylations [61, 62]. Histones carrying these so-called 'active marks' are commonly associated with highly transcribed genes and are thought be refractory to PRC2-mediated H3K27 tri-methylation [61, 62]. Why then are these marks not included in the model? To date, there is no compelling evidence that any of these 'active marks' participate in a cis-acting positive feedback mechanism independently of transcription - similar to that suggested for PRC2 [60, 107]. If these feedback mechanisms do not exist, then the 'active marks' alone are insufficient to direct their own maintenance. One possibility is that these 'active marks' are laid down by transcription-coupled processes and function primarily to antagonise Polycomb silencing, rather than directly recruiting complexes required for their own reinforcement. That is, if 'active marks' simply help to mark transcribed regions, this could provide another mechanism by which transcription could antagonise PRC2 silencing. Such an effect could be included in the modelling framework outlined above and would only act to strengthen the bistability.

How the non-processive model with transcription opposing silencing could be coupled to the H3K36me3/H3K27me3 model of the *FLC* nucleation region (Section 3.3) is considered in Chapter 6 (Sec. 6.3).

4.5 Summary

In this chapter, mathematical modelling was used to explore how transcription could form an opposing state to PRC2-mediated gene silencing. It was found that a model incorporating non-processive methylation and demethylation of H3K27, is capable of generating bistability without an explicit 'A'-mark. The model was shown to be bistable with a rate of histone turnover that allows incorporation of H3.3 in a transcription-dependent manner. Fitting the model to time-resolved mass spectrometry data, allowed determination of an average *in vivo* K27-methylation rate. This analysis also suggested that sub-saturated H3K27me3 levels at a chromatin domain are sufficient to cause gene repression.

The model was then challenged with dynamically varying 'gene activation' inputs. It was shown that the slow dynamics of H3K27 methylation reported in HeLa cells [26] provide a distinct advantage when buffering noisy transcriptional inputs. Finally, simulations with fixed changes to either the transcription rate or the local rate of H3K27 methylation were used to study the interaction between cis-epigenetic models of chromatin and diffusible trans-regulators. This led to a synthesis of cis and trans memory and the concept of a 'cis memory window' - the range of trans-factor input strengths that a cis-encoded chromatin state can sustain before changing states.

4.6 Methods

Programming languages and libraries were described in Sec. 3.5.3, and computing hardware was detailed in Sec. 3.5.2.

4.6.1 *Stochastic simulations of two-state and non-processive models*

A general introduction to the Gillespie algorithm was provided in Sec. 3.5.4.

Stochastic simulations of histone methylation and transcription were simulated at non-constant time intervals according to the 'direct' Gillespie algorithm [212]. For a system of N histones, there is a total of $2N+1$ possible stochastic reactions (N histone methylations, N histone demethylations and transcription). After each system update, all reaction propensities were recalculated before selecting the next time interval and subsequent reaction. Each transcription event resulted in a one-step demethylation on H3K27 with probability per histone p_{dem}, and nucleosome turnover with probability per histone p_{ex}. Typically $p_{\mathrm{dem}}, p_{\mathrm{ex}} < 1/N$, so on average less than one histone turnover or demethylation occurred per transcription event. A transcription event did not otherwise change the chromatin state.

In addition to this Gillespie algorithm simulation, DNA replication was implemented at regular intervals, every 22 hours. To replicate DNA, the Gillespie algorithm simulation was interrupted if the projected time for the next reaction exceeded the time at which DNA would have been replicated. In this case, system time was updated to the precise time of DNA replication and the Gillespie algorithm was repeated for another cell cycle. A similar approach has been used when incorporating reactions with delays in Gillespie algorithm simulations [297].

4.6.2 *Quantities calculated from simulations*

Time-averaging. For an individual simulation time-course comprising K reactions, the Gillespie algorithm determines the state of the system at a series of K simulation time-points t_i (the trajectory). The duration between time points $\Delta t = t_{i+1} - t_i$ is not constant. Time-averaging for a quantity x_i (e.g. P_{OFF} or P_{ON}) between t_0 and t_{K-1} was performed

using the formula,

$$\sum_{i=0}^{K-1} x_i \frac{t_{i+1} - t_i}{t_{K-1} - t_0}. \tag{4.5}$$

Bistability measures. The quantity introduced in [213] to determine the time-averaged 'state' of the gene is equivalent to P_{OFF}, the probability that the number of repressive me3 marks exceeds the number of neutral me0 marks by at least half the total number of histones,

$$P_{\mathrm{OFF}} = \Pr\left(n_{\mathrm{me3}} + n_{\mathrm{me2}} - n_{\mathrm{me1}} - n_{\mathrm{me0}} > \frac{N}{2}\right), \tag{4.6}$$

with $N = n_{\mathrm{me3}} + n_{\mathrm{me2}} + n_{\mathrm{me1}} + n_{\mathrm{me0}}$, this reduces to,

$$P_{\mathrm{OFF}} = \Pr\left(n_{\mathrm{me3}} + n_{\mathrm{me2}} > \frac{3N}{4}\right). \tag{4.7}$$

Similarly,

$$P_{\mathrm{ON}} = \Pr\left(n_{\mathrm{me3}} + n_{\mathrm{me2}} < \frac{N}{4}\right), \tag{4.8}$$

and the bistability measure is given by,

$$B = 4P_{\mathrm{OFF}}P_{\mathrm{ON}}. \tag{4.9}$$

Since the histone type that is randomly inserted during DNA replication is identified with the high transcription state, it was necessary to allow the system to recover from this perturbation before assessing the stability of the state through DNA replication. For this reason, results were typically calculated only for the last hour of each cell cycle. This allowed systems with slow recovery times to attain high values of B, consistent with their long-term stability.

After introduction of the firing rate threshold, P_T, these definitions of P_{ON} and P_{OFF} no longer accurately reflect the chromatin state in terms of its control on expression. In this case, the gene is defined as being in the OFF-state if the chromatin influence on the firing rate function is in its lower quartile,

$$P_{\mathrm{OFF}} = \Pr\left(f_{\max} - \frac{n_{\mathrm{me2}} + n_{\mathrm{me3}}}{NP_T}\left(f_{\max} - f_{\min}\right) < f_{\min} + \frac{f_{\max} - f_{\min}}{4}\right), \tag{4.10}$$

which can be simplified to,

$$P_{\mathrm{OFF}} = \Pr\left(n_{\mathrm{me2}} + n_{\mathrm{me3}} > \frac{3NP_T}{4}\right), \tag{4.11}$$

and likewise for P_{ON},

$$P_{\mathrm{ON}} = \Pr\left(n_{\mathrm{me2}} + n_{\mathrm{me3}} < \frac{NP_T}{4}\right). \tag{4.12}$$

Note that with $P_T = 1$, Eqs. 4.11 and 4.12 reduce to 4.7 and 4.8, respectively. These latter definitions are therefore consistent with earlier usage of the bistability measure B [213].

First passage times. Mean first passage times, $t_{\mathrm{FP(me0)}}$ and $t_{\mathrm{FP(me3)}}$, are defined as the average time taken for the system to change to the opposite chromatin state, when initialized in the uniform me0 or me3 state, respectively. E.g., for the initial active state,

$$t_{\mathrm{FP(me0)}} = \min\left\{ t \;\middle|\; n_{\mathrm{me2}}(t) + n_{\mathrm{me3}}(t) > \frac{3NP_T}{4} \right\} \tag{4.13}$$

In the simulations, mean first passage times were bounded above by the total simulation time (Either 20 or 50 cell cycles). This allowed the introduction of a quantity to measure the mutual stability of the two states, the 'combined first passage',

$$FP = \left(t_{\mathrm{FP(me0)}} t_{\mathrm{FP(me3)}}\right) \Big/ T^2, \tag{4.14}$$

where T is the total simulation time. Since $t_{\mathrm{FP(me0)}}, t_{\mathrm{FP(me3)}} < T$, then $0 < FP \leq 1$.

4.6.3 *Fitting triple-SILAC mass spectrometry data*

Data were generated in the laboratories of Anja Groth (Biotech Research and Innovation Centre, Copenhagen) and Axel Imhof (Ludwig-Maximilians Universität, Munich), and obtained from Carsten Marr (Institute of Computational Biology, Helmholtz Zentrum, Munich).

Mass spectrometry data were analysed and normalised as described in [26] to yield H3K27me3 levels on 'old' and 'new' histones as a proportion of the total old and new labelled peptides measured at each timepoint (c.f. Fig. 3E in [26]). Simulation data for H3K27me3 levels on old and new histones were initially also expressed as a proportion of

the levels of old and new histones, respectively. Because the mass spectrometry data represent a genome-wide average, and the simulation represents a single PRC2-target gene, simulation data must be scaled in order to make quantitative comparison with experiments. To do so, the simulation data were further normalised so that the average simulated cell-cycle-end value of H3K27me3 on total histones, $P_{\text{me3-end}}$ was equal to the proportion of H3K27me3 on old histones at $t = 0$ (0.301), obtained experimentally. That is, each simulation time-point was multiplied by the factor $0.301/P_{\text{me3-end}}$. This is valid because all histones are labelled as old at $t = 0$, so the value 0.301 also represents the relative amount of H3K27me3 on total histones at the end of each cell cycle.

After this normalisation, the $t = 0, 10, 24, 48$ hour time-points for old and new histones were compared with equivalent model time-points using the sum of squared errors. Three biological replicates were available for each time-point [26].

The normalisation procedure requires that the model is epigenetically stable over many cell cycles in the repressed state in order that the extracted $P_{\text{me3-end}}$ correctly normalises the simulated data at the start of the cell cycle in which 'new' histones are added. In Figure 4.9, it can be seen that the normalisation fails for some of the unstable models for low values of k_{me} (bottom right). This is because the repressed (high-me3) state is generally not maintained through the equilibration cell-cycles before new histones are added.

4.6.4 *Stochastic model of a noisy transcriptional regulator*

The following model was used in [285] to investigate how rates of transcription and translation affect the variability in protein number over time. In the present work it is used as an arbitrary 'noisy' input function representing a trans-regulator.

$$\text{DNA} \xrightarrow{s_R} \text{mRNA} \xrightarrow{d_R} \varnothing \tag{4.15}$$

$$\text{mRNA} \xrightarrow{s_P} \text{Protein} \xrightarrow{d_P} \varnothing \tag{4.16}$$

In steady state, $\langle \text{mRNA} \rangle = s_R/d_R$ and $\langle \text{Protein} \rangle = s_R b/d_P$, where $b = s_P/d_R$ is the average number of proteins synthesised per mRNA transcript [285]. The 'noise' in protein level is controlled by the value of b, with larger b giving a more variable output.

To simulate a transcriptional regulatory protein with variable concentration $r(t)$, the following parameter values were used,

$$d_R = 1/2, \quad d_P = 1/12, \quad s_R = d_P \langle r(t) \rangle / b, \quad s_P = d_R b, \tag{4.17}$$

where the unit of each rate is hours^{-1}. Specifying the mean number of regulatory proteins as $\langle r(t) \rangle = 1000$, the noise can then be varied using the single parameter b. Higher values of b indicate a greater noise. The noisy gene activation $\alpha(t)$ is then given by $\alpha(t) = r(t)/\langle r(t) \rangle$.

The numbers of protein and RNA molecules were explicitly simulated using the Gillespie algorithm according to the model specified in Eqs. 4.15, 4.16. These simulations to generate $\alpha(t)$ were performed concurrently with simulations of the chromatin state (described in Sec. 4.6.1).

5

RNA-BINDING BY LHP1 IS REQUIRED FOR POLYCOMB SILENCING

Understanding how transcription and epigenetic memory are coordinated by chromatin is a central goal of modern epigenetics. The previous chapter considered the role of transcription in direct modulation of chromatin through increasing histone turnover rates and recruitment of chromatin modifiers. This chapter focuses on how RNA, the direct product of transcription, could also contribute to defining chromatin states.

Protein-RNA interactions have recently emerged as a mechanism for recruitment, removal, and modulation of activity of proteins associated with chromatin (reviewed in [141, 298, 299]). In the case of PRC2, RNA-binding of some subunits has been shown to regulate enzymatic activity *in vitro* [139, 144] and biological function *in vivo* [139, 143].

Individual proteins that can interact with both modified histones and RNA provide an opportunity to investigate how RNA-binding could affect the function of chromatin-associated protein complexes more generally. It is well-known that LIKE HETEROCHROMATIN PROTEIN 1 (LHP1) can recognise H3K27me3 [150, 151], and it has been reported that mammalian [146] and yeast [147] homologues of LHP1 are also able to bind RNA. LHP1 is required to maintain repression of Polycomb target genes in Arabidopsis [150, 151], including *FLC* [148]. This chapter contains an investigation into the RNA-binding ability of LHP1 and the importance of this RNA-binding for Polycomb silencing.

The chapter begins with *in vitro* characterisation of LHP1, in which it is shown that LHP1 can bind to RNA through positively-charged residues in a flexible region of the protein. These *in vitro* studies are used to generate separation-of-function mutants that are defective in either H3K27me3-recognition or RNA-binding. Mutant proteins are

then expressed in Arabidopsis plants lacking a functional LHP1 protein, to allow characterisation of the relative importance of these two biochemical activities in biological function.

5.1 Introduction to LHP1 and RNA-protein interactions in epigenetics

Heterochromatin is generally perceived as transcriptionally inactive because RNA does not accumulate from reporter genes integrated at heterochromatic loci. However, Pol II can be detected in heterochromatic regions [300–302], and transcription run-on experiments indicate that RNA can be produced from these regions at low levels [302]. More generally, genome-wide studies have suggested that transcription is more pervasive than previously anticipated, with regulation of RNA accumulation being achieved partly at the co-transcriptional or post-transcriptional level [303, 304]. Accordingly, the mathematical model presented in Chapter 4 contains low levels of transcription even in the repressed state, though not all of these RNAs may survive to maturity if the machinery necessary for splicing, polyadenylation, and export of the pre-mRNA are not efficiently recruited to these 'repressed' loci [303, 305, 306].

In some cases, the maintenance of heterochromatin structure can itself be dependent on the RNA-interference pathway, which uses small ~21-25 nucleotide RNAs bound in Argonaute-containing complexes to recognise nascent transcripts in heterochromatin, and thereby recruit chromatin modifiers to maintain the heterochromatic state (reviewed in [307]).

It has been reported that part of this mechanism in *S. pombe* involves RNA-binding by the fission yeast Heterochromatin Protein 1 (HP1) homologue, $HP1^{Swi6}$ [147]. RNA-binding by $HP1^{Swi6}$ is proposed to target nascent transcripts arising from within heterochromatin for degradation [147], explaining why infrequent transcription events do not result in accumulation of heterochromatic RNA. It was also shown that

RNA-binding of HP1^{Swi6} is necessary to prevent the spreading of heterochromatin into neighbouring genomic regions [138]. In this model, the high levels of nascent RNA from transcription units that border heterochromatin act as a sponge to 'mop-up' any HP1^{Swi6} that spreads linearly along the chromatin fibre. Thus HP1^{Swi6} prevents transcript accumulation [147], and transcription prevents HP1^{Swi6} spreading [138]. Both of these activities depend on RNA-binding.

The latter model (prevention of spreading) is similar to the proposed role of RNA-binding of PRC2. PRC2 binds to RNA in a non-sequence-specific manner [139, 144, 145] and interacts with nascent RNA across the mammalian genome [143, 145, 308, 309]. Nascent RNA inhibits PRC2 activity and may therefore help to prevent targeting of transcriptionally active loci by the PRC2 machinery [144, 145].

5.1.1 *LHP1: a euchromatic HP1 protein*

As suggested by the name Heterochromatin Protein 1, members of the conserved HP1 family of proteins were originally identified as major components of heterochromatin (reviewed in [310]). Arabidopsis LHP1, on the other hand, is predominantly localised in euchromatin [149, 311]. Specifically, LHP1 binds to repressed Polycomb target genes, such as *AGAMOUS*, *APETALA3*, *FT*, *WUS* and *FLC* [150, 151, 312, 313].

Arabidopsis *LHP1* was first discovered in screens for mutants with altered leaf glucosinolate levels, where it was named *TU8* [314, 315]. Independently, it was isolated as an enhancer of the *terminal flower 1* (*tfl1*) mutant phenotype and was named *TERMINAL FLOWER 2* [222]. *lhp1* mutants were also identified as early-flowering from a T-DNA mutant collection [316] and the *TFL2* gene was later shown to encode LHP1 [223]. The *lhp1* (*tfl2*) mutant phenotype is pleiotropic, which is thought to be due to mis-expression of many PRC2 target genes.

5.1.2 *The HP1 family of proteins*

As shown in Figure 5.3A, the defining structural features of the HP1 family are the chromodomain [317], which is capable of binding to methylated lysines [318–321], and the chromoshadow domain [322], which is distantly related to the chromodomain and is involved in protein-protein interactions, including HP1 homodimerisation [323–325]. The HP1 family also shares an N-terminal region of variable length, often containing an acidic patch, comprised of aspartic and glutamic acid residues [326]. Finally, the chromodomain and chromoshadow domain are separated by a poorly conserved flexible region of variable length referred to as the 'hinge'. The hinge region has been previously implicated in nucleic acid binding [146, 147].

The chromodomain is shared by many chromatin-associated proteins, including HP1 and Drosophila Polycomb (Pc), and is able to recognise methylated lysines on histone H3 tails with high specificity [321]. This is proposed to allow targeting of chromodomain-containing complexes to specific sites in chromatin (reviewed in [327]). However, while HP1 co-localises with H3K9me2/me3, Arabidopsis LHP1 co-localises with H3K27me2/me3 [150, 151] - more similar to Pc in Drosophila. On histone H3, lysine-9 and lysine-27 are both found in the consensus peptide sequence Ala-Arg-Lys-Ser. It has been suggested that the LHP1 chromodomain is not able to differentiate between H3K9me and H3K27me peptides [150]. In contrast, the chromodomains of Pc and HP1 from Drosophila show higher affinity binding of H3K27me and H3K9me, respectively [321]. It is worth noting that only a 56 amino acid fragment of LHP1 (amino acids 104-160) was used in these *in vitro* assays, compared to 59 and 98 amino acids for HP1 (17-76) and Pc (1-98), respectively. Therefore, it remains possible that discrimination of H3K27me and H3K9me is indeed achieved by full-length LHP1 in its native context. Mutations in the LHP1 chromodomain have been shown to

eliminate LHP1 recruitment to Polycomb target genes, suggesting that recognition of H3K27me3 is necessary for targeting *in vivo* [312].

Although it is structurally a HP1 family member, LHP1 may be functionally more analogous to Pc, which is part of Polycomb Repressive Complex 1 (PRC1) in Drosophila. A physical link between LHP1 and PRC2 has also been identified, via the core PRC2 subunit MSI1 [153] (Fig. 1.3, p. 26).

HP1 proteins have been shown to bind without sequence specificity to both DNA [328, 329] and RNA [146, 147] *in vitro*. In particular, basic residues within the flexible hinge region of the *S. pombe* HP1 homologue $HP1^{Swi6}$ were shown to be important for the ability of this protein to recognise RNA [147].

5.2 *In vitro* characterisation of LHP1 activity

As discussed in Chapter 3, the Arabidopsis HP1 homologue LIKE HETEROCHROMATIN PROTEIN 1 (LHP1) is important for maintenance of *FLC* repression after vernalisation (Fig. 3.5, p. 97). Specifically, it was shown that *lhp1* mutants fail to spread H3K27me3 from the nucleation region to the gene body at *FLC* (Fig. 3.6, p. 98). Thus LHP1 has structural similarity with $HP1^{Swi6}$ and is required for chromatin state changes at *FLC*. It is intriguing to consider if LHP1 is able to bind nascent RNA, and how this affects its function *in vivo*.

To investigate possible roles of RNA-binding by LHP1 in repression of Polycomb target genes, the ability of LHP1 to bind RNA was investigated using *in vitro* assays.

5.2.1 *LHP1 binds to RNA*

Glutathione S-transferase (GST) LHP1 fusion protein was expressed in *Escherichia coli* (*E. coli*) and purified using glutathione sepharose followed by anion exchange chromatography (Fig. 5.1, Methods: Secs. 5.6.3, 5.6.4). Electrophoretic mobility shift assays (EMSA) were then performed to investigate binding of this recombinant protein to nucleic

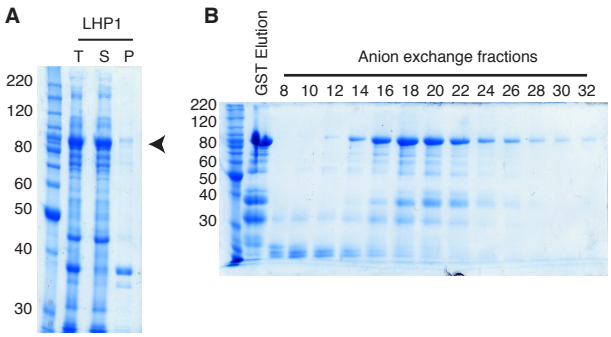


Figure 5.1: Expression of LHP1 in *E. coli* and purification. Coomassie-stained SDS-PAGE. (A) Bacterial lysate after induction and expression of GST-LHP1. T = Total lysate, S = Soluble lysate, P = Pellet (insoluble fraction). Molecular size markers are shown in kDa. Predicted size of GST-LHP1 is 76.4 kDa. Putative LHP1-containing band indicated with an arrowhead. (B) GST Elution is the glutathione elution after purification of GST-LHP1 using glutathione sepharose beads. Also shown are fractions containing GST-LHP1 after anion exchange chromatography. The protocol used for expression and purification is described in detail in Sec. 5.6.4 (p. 209).

acids *in vitro* (Method: Sec. 5.6.5). In EMSA, a fluorescent-labelled nucleic acid probe is incubated with recombinant protein and the reaction mixture is separated by gel electrophoresis. Because the protein-nucleic acid complex is less mobile in the gel than the free probe, interaction with the protein causes the probe to run more slowly through the gel, and appears as a 'shift' to higher apparent molecular weight. In these assays, GST-LHP1 was able to bind to a 40 nt single-stranded RNA probe but not to a single-stranded DNA probe of the same sequence (Fig. 5.2B). Double-stranded DNA binding by LHP1 was detected by EMSA, but this interaction was much lower affinity than LHP1 single-stranded RNA binding (Fig. 5.2B).

In vitro, $HP1^{Swi6}$ is thought to exist as a dimer in solution at concentrations above 17 nM [330]. This dimerisation is mediated through the chromoshadow domain [325, 330]. Conversely, human HP1β exists in equilibrium between monomer and dimer up to concentrations of 1-2 µM *in vitro*, and is predominantly a dimer above this concentration [104]. It is currently unknown if LHP1 exists as a dimer *in vitro*,

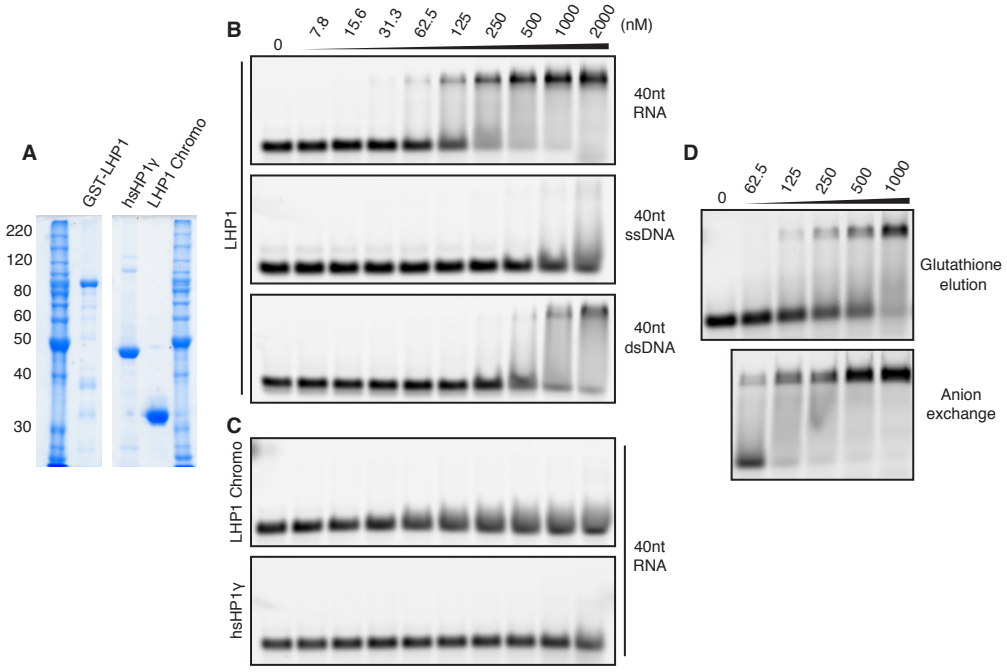


Figure 5.2: EMSA with purified LHP1. (A) Coomassie-stained SDS-PAGE showing purified GST-LHP1 used in EMSA assays. (B) EMSA showing GST-LHP1 binding to nucleic acids. Concentration of protein in nM shown above. LHP1 shows higher affinity for RNA than double-stranded DNA (ds-DNA) or single-stranded DNA (ssDNA). In each case, the labelled nucleic acid probe is a 40 nt sequence from *FLC* exon 1. (C) Negative control EMSA using the same concentrations of GST-HP1γ or GST-LHP1$_{105-160}$ (chromodomain) together with the 40 nt RNA probe. (D) EMSA with GST-LHP1 purified after glutathione elution or after further purification using anion exchange.

although the chromoshadow domain appears to contain the conserved residues at the dimerisation interface [325].

It is notable that only one LHP1-RNA complex is present in EMSA, and that the apparent molecular weight of this band is independent of LHP1 concentration. This suggests that only one LHP1 species (monomer or dimer) binds to RNA in these assays. Since the stoichiometries of LHP1-LHP1 interaction and the LHP1-RNA interaction are unknown, the dissociation constant for LHP1-RNA binding cannot be determined from these assays. Instead, the half-saturation constant $K_{1/2}$, equal to the LHP1 monomer concentration at which half the RNA is bound will be referred to as the 'affinity' for the remainder of this work. For GST-LHP1, $K_{1/2} \approx 200$ nM (Fig. 5.2B).

GST-LHP1$_{105-160}$ (chromodomain) showed no binding to the RNA probe (Fig. 5.2C), indicating that the RNA-binding of GST-LHP1 is specific to the full-length LHP1 protein. RNA-binding was also not detected for human HP1γ fused to GST[a] (Fig. 5.2C). This lack of interaction suggests that not all HP1 proteins have the ability to bind RNA. Positively charged residues in the hinge region of *S. pombe* HP1^{Swi6} have been shown to be important for its RNA-binding. HP1^{Swi6} has 25 K/R residues in its 137 amino acid hinge region, and Arabidopsis LHP1 has 35 K/R in its 222 amino acid hinge (Fig. 5.3). By contrast, the hinge of human HP1γ is only 44 amino acids in length and contains 14 K/R residues. That HP1^{Swi6} and LHP1 appear to bind RNA but human HP1γ does not, suggests that the length and amino acid composition of the flexible hinge region are important for mediating this interaction.

RNA-binding of the *S. pombe* HP1 homologue HP1^{Swi6} was previously shown to be independent of RNA sequence [147]. The 40 nucleotide sequence used in the assays presented here was derived from the *FLC* nucleation region. The sequence-specificity of LHP1 RNA-binding was not investigated other than to verify that binding could also be detected for two other sequences derived from the *FLC* 3′-end (data not shown).

Interestingly, the apparent affinity of GST-LHP1 for RNA was observed to be 2-4 times higher after including anion exchange chromatography as an additional purification step after glutathione elution. This is likely to be because LHP1 co-purifies with bacterial nucleic acids, which are effectively removed by anion exchange. For this reason, it was important to purify LHP1 proteins from *E. coli* using glutathione sepharose purification followed by anion exchange chromatography.

5.2.2 *Designing separation-of-function LHP1 mutants*

An LHP1 mutant which is defective in RNA binding but not H3K27-me3-binding would be a valuable tool for dissecting the contribution

[a] Purified GST-fusion protein kindly provided by Veronika Ostapcuk, Friedrich Miescher Institute, Basel, Switzerland

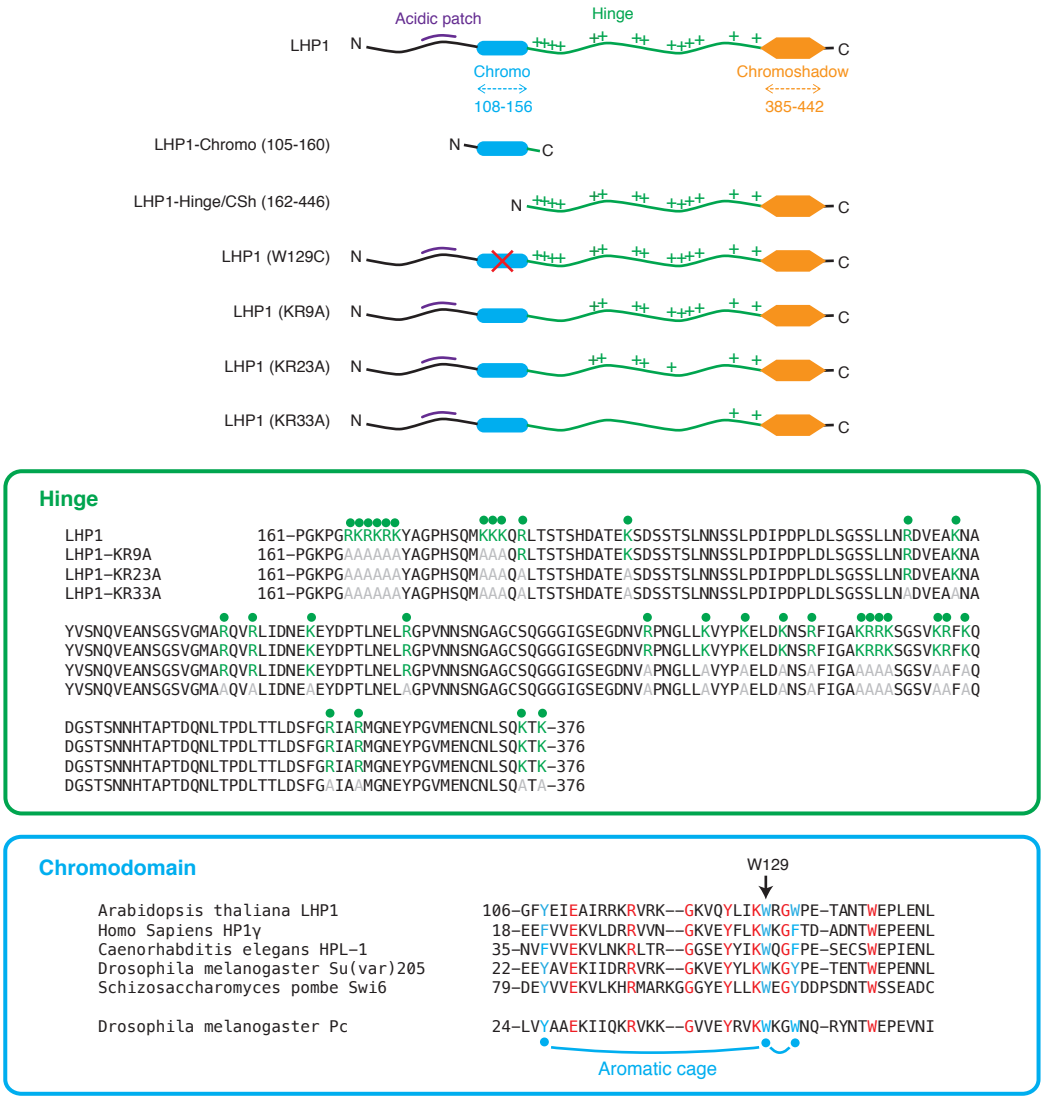


Figure 5.3: Structure of LHP1 protein and design of mutants. LHP1 has an extended N-terminal domain containing an acidic patch, a chromodomain and a C-terminal chromoshadow domain. The chromo- and chromoshadow domains are separated by a flexible 'hinge', which is not well conserved in length or primary amino acid sequence. Positively charged residues in the hinge region are indicated schematically with green '+' symbols. Middle panel (Hinge) shows the position of lysine (K) and arginine (R) amino acids in the hinge, and their mutations in putative RNA-binding mutant proteins. Lower panel (Chromodomain) shows a seqence-based alignment of part of the chromodomain with that of HP1 homologues from human, *C. elegans*, Drosophila, *S. pombe* and also the chromodomain from the Drosophila protein, Polycomb (Pc). Alignment of annotated chromodomains performed in Geneious, using the Geneious alignment tool. Aromatic cage residues are indicated in blue, while other residues conserved among these HP1 chromodomains are indicated in red.

of RNA-binding to LHP1 function *in vivo*. Previous work on *S. pombe* HP1 protein $HP1^{Swi6}$ highlighted the importance of positively charged K/R residues in the hinge region for binding to negatively charged RNA [147]. In the $HP1^{Swi6}$ study, mutation of 25 K/R residues in the hinge region was shown to abolish RNA binding. Importantly, nuclear magnetic resonance demonstrated that, outside the hinge region, the *in vitro* structure of $HP1^{Swi6}$ was not affected by these mutations. That is, the N-terminus, chromodomain and chromoshadow still adopted the same folds and interactions in the presence or absence of these 25 lysine and arginine residues [147].

The hinge region of LHP1 is highly basic (pI = 10.10), and contains 20 lysine (K) and 15 arginine (R) residues (Fig. 5.3). LHP1 expression constructs encoding three putative RNA-binding mutants were generated by conversion of 9, 23 or 33 of these K/R residues to alanine (A). These are referred to as KR9A, KR23A, and KR33A (Fig. 5.3).

A putative H3K27me3-binding mutant protein was also generated, via mutation of the chromodomain. An alignment of the LHP1 chromodomain with the chromodomains of other HP1 proteins, and that of Drosophila Polycomb (Pc) is shown in Figure 5.3. The highly conserved 'aromatic cage' residues (Y108, W129, W132), which form the binding pocket for H3K27me3 in Pc (H3K9me3 in HP1) [321], are present also in LHP1. Mutation of these residues in HP1 proteins has previously been shown to result in loss of binding to H3K9me3 [318, 319, 326]. A previous study in Arabidopsis also identified a splice-site mutation in LHP1, resulting in W129 being converted to CCER [312]. This mutation was observed to disrupt LHP1 H3K27me3-binding *in vitro* [312] and LHP1 localisation to Polycomb target genes *SEPALATA3* and *AGAMOUS in vivo*. W129C was therefore selected as a putative H3K27me3-binding mutant.

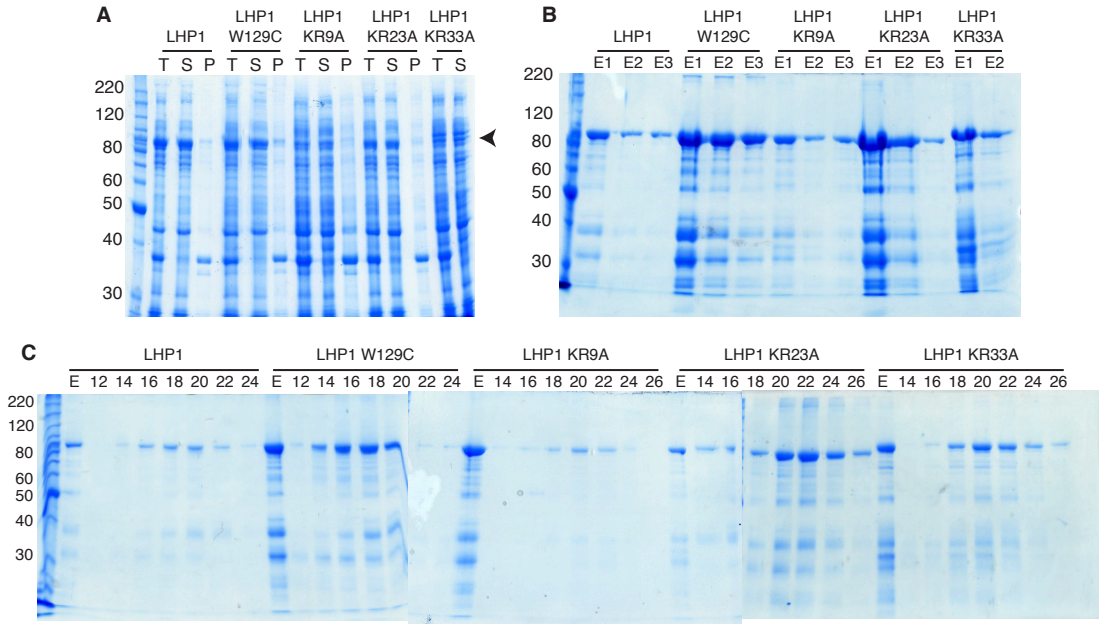


Figure 5.4: Expression of LHP1 mutants in *E. coli* and purification. Coomassie-stained SDS-PAGE. (A) Bacterial lysate after induction and expression of GST-LHP1, wild-type and mutants. T = Total lysate, S = Soluble lysate, P = Pellet (insoluble fraction). Molecular size markers are shown in kDa. Predicted size of GST-LHP1 is 76.4 kDa. (B) Glutathione elutions (E1 = Elution 1, etc.) after purification of GST-LHP1 proteins using glutathione sepharose beads. (C) Fractions containing GST-LHP1 protein after anion exchange chromatography.

5.2.3 *LHP1 RNA-binding is disrupted in hinge mutants*

LHP1(KR9A), LHP1(KR23A) and LHP1(KR33A) were purified using the protocol developed for LHP1 (Fig. 5.4, Method: Sec. 5.6.4) and assayed for RNA-binding by EMSA (Fig. 5.5). GST-LHP1(KR23A) and GST-LHP1(KR33A) were unable to bind to RNA at the highest concentration of protein used in this assay (2µM), while GST-LHP1(KR9A) showed a 4-8 fold reduction in affinity compared to wild-type. This demonstrates that LHP1 binds RNA through its hinge region and that basic residues are required for the interaction.

5.2.4 *LHP1 recognises H3K27me3 through its chromodomain*

To investigate the histone-binding properties of the full-length recombinant protein, peptide pulldown assays were performed. In these experi-

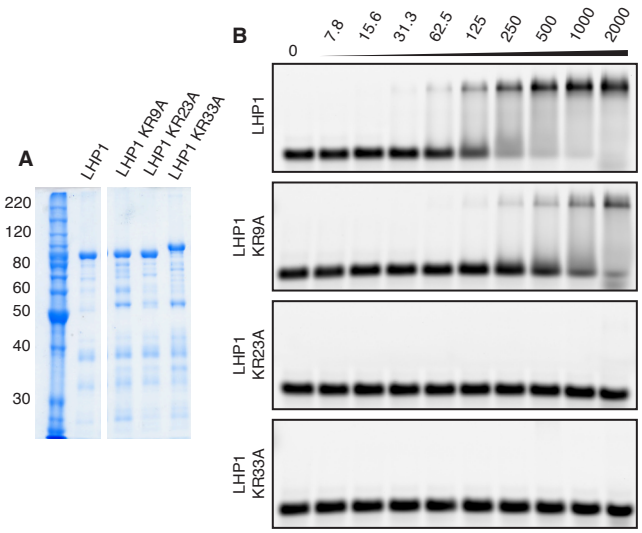


Figure 5.5: EMSA with LHP1 RNA-binding mutants. (A) Coomassie-stained SDS-PAGE showing purified wild-type and mutant GST-LHP1 used in EMSA assays. (B) EMSA showing that RNA-binding by LHP1 is partly disrupted in the KR9A mutant and completely disrupted in KR23A and KR33A mutants. Concentration of protein in nM shown above.

ments, a biotinylated peptide corresponding to amino acids 21-44 of histone H3 was conjugated to streptavidin beads and incubated with the recombinant protein. After washing to remove non-specific interacting proteins, reactions were then separated by SDS-PAGE and GST-fusion proteins detected by immunoblot with an anti-GST antibody. A strong interaction was detected between the full length protein and H3 peptide carrying a tri-methylated lysine at position 27 (H3K27me3) but not unmodified H3 or H3K36me3 peptide of the same sequence (Fig. 5.6B). This demonstrates that recombinant full-length LHP1 is able to recognise H3K27me3. GST fusions of the chromodomain (105-160) and the hinge-CSh domain were also tested. In agreement with earlier studies [150], the chromodomain alone was sufficient for specific interaction with H3K27me3. No interaction was detected between GST-Hinge-CSh and any of the peptides.

Surface plasmon-resonance (SPR) experiments involve conjugating a 'bait' ligand to the surface of a sensor and injecting a solution contain-

ing the analyte at constant flow rate. Interaction between the analyte and the bait is then detected as a change in mass bound to the surface of the sensor. The signal in response units (RU) is linearly proportional to the total mass of bound analyte. H3K27me3 peptide was conjugated to the surface of the sensor and then full-length GST-LHP1 fusion proteins were injected to detect interactions. For full-length LHP1 and LHP1(KR23A), fast association and slow dissociation were observed, indicating strong binding to the immobilised H3K27me3 peptide (Fig. 5.6C). The response units are proportional to the mass of the bound protein and the expected response of LHP1(KR23A) is therefore less than that of wild-type LHP1, as the KR23A mutant has a 4% lower molecular weight. Because the stoichiometry of these complexes is unknown, these relative responses cannot be used to determine the relative affinity of LHP1 and LHP1(KR23A) for the H3K27me3 peptide.

When an equal concentration of GST-LHP1(W129C) was instead injected, the interaction was greatly reduced, consistent with this mutation disrupting the ability of LHP1 to recognise H3K27me3. For full-length proteins, none of these curves were well fit by a single-exponential, indicating complex interactions at the sensor surface or complexes of variable stoichiometry in the solution. In addition, it was difficult to find regeneration conditions which removed all proteins from the previous sample while maintaining the integrity of the immobilised peptide. Furthermore, significant non-specific interactions were detected from the sensor ChIP surface. These problems prohibited calculation of the affinity constant for the interaction without further optimisation of the assay. Nonetheless, these results can be regarded as an indication that H3K27me3-binding is maintained in the RNA-binding mutant LHP1 KR23A and perturbed in the chromodomain mutant LHP1(W129C).

EMSA with the same GST-LHP1 and GST-LHP1(W129C) proteins indicate that LHP1(W129C) is able to bind RNA with similar affinity as wild-type LHP1 (Fig. 5.6E).

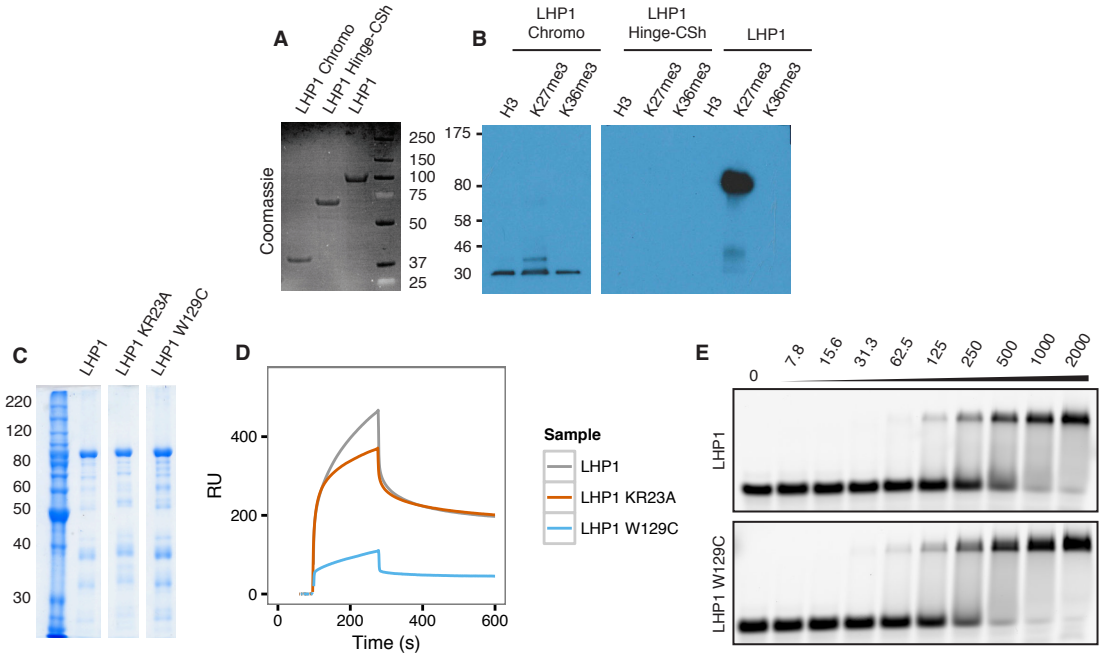


Figure 5.6: LHP1 recognition of H3K27me3. (A) Coomassie-stained SDS-PAGE showing purified wild-type and mutant GST-LHP1 used in peptide-pulldown assays. (B) Visualisation of GST-fusion proteins by anti-GST immunoblot after peptide-pulldown assay with a biotinylated H3 peptide (amino acids 21-44) or the same peptide carrying K27me3 or K36me3. (C) Coomassie-stained SDS-PAGE showing purified wild-type and mutant GST-LHP1 used in Surface-plasmon resonance (SPR) assays and EMSA. (D) SPR experiment showing the response units (RU) after sample was injected over a flow surface with immobilised H3K27me3 peptide. 120 s injection is followed dissociation. All protein samples were concentrated to 340 nM. (E) EMSA showing binding of GST-LHP1 and GST-LHP1(W129C) to a 40nt RNA probe. Concentration of protein in nM shown above.

In summary, LHP1 KR23A is an RNA-binding mutant, which maintains its ability to recognise H3K27me3. Conversely, LHP1(W129C) is impaired in its ability to recognise H3K27me3, yet is still able to bind RNA. These mutants therefore represent separation-of-function alleles of LHP1.

5.3 Separation-of-function *LHP1* mutants in Arabidopsis

This section describes the generation of Arabidopsis plants expressing the separation-of-function LHP1 mutants, which are used to test the

biological functions of LHP1 H3K27me3-binding and RNA-binding *in vivo*.

5.3.1 *Design of transgenic LHP1 constructs*

To maintain the endogenous regulation of LHP1, the full genomic LHP1 sequence was used to generate transgenic LHP1 constructs. Previous work showed that both an 11 kb SalI fragment and a 5.5 kB SpeI/SnaBI fragment of Col-0 LHP1 (Fig. 5.7) were both able to complement the *lhp1-1* mutation equally effectively in the Wassilewskija accession [316]. The shorter 5.5 kB SpeI/SnaBI fragment, containing 2266 bp upstream of the TSS and 797 bp downstream of the TES (Fig. 5.7) was therefore used as the basis for constructing all LHP1 transgenes.

Conveniently, the entire hinge region exists within a single exon of LHP1 (exon 4), which allowed simple cloning of the mutated hinge domains from bacterial expression constructs. W129 spans the exon 2/3 junction, so a single point mutation at the first nucleotide of exon 3 was used to create the W129C mutant. To allow visualisation of fusion proteins, a C-terminal eGFP tag was inserted between the native stop codon and the 3′-UTR. Over-expressed LHP1-eGFP was previously shown to complement the *lhp1-1* mutation [149], demonstrating that a C-terminal eGFP does not significantly disrupt LHP1 function. All constructs were also made independently with a C-terminal 3xHA tag for use in future immunoprecipitation experiments.

5.3.2 *Characterisation of subcellular localisation*

Replacement of lysine (K) and arginine (R) residues with alanine (A) in the KR9A and KR23A mutants disrupts the bi-partite nuclear localisation signal (NLS) at the N-terminal end of the hinge region (Fig. 5.3). It was therefore expected that subcellular localisation may be disrupted in these mutants. To test the localisation of the wild-type and mutant LHP1 proteins, LHP1 transgenes were transiently expressed in *Nicotiana benthamiana* (*N. benthamiana*) plants. As shown in Fig. 5.8B,

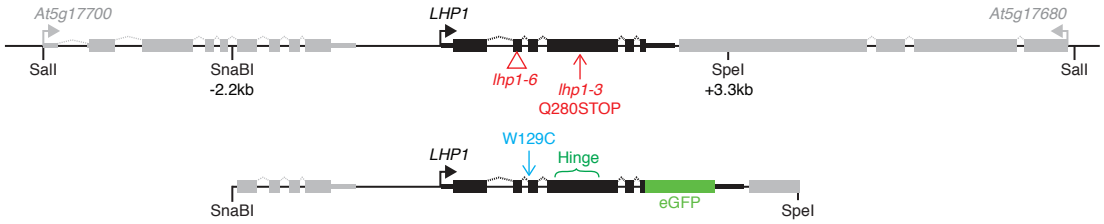


Figure 5.7: Genomic LHP1 structure. LHP1 exons are represented as black boxes and the splicing pattern indicated with dashed lines. Narrow black boxes represent the 3′- and 5′-UTR. The positions of the *lhp1-6* T-DNA insertion and *lhp1-3* point mutation are shown in red. A single point mutation in exon 3 was used to generate the W129C mutation. The hinge region is contained within exon 4. SalI, SpeI and SnaBI sites are indicated. Genes neighbouring LHP1 are depicted in grey. The SnaBI/SpeI fragment was used in transgenic LHP1 constructs such as LHP1-eGFP depicted below.

Arabidopsis LHP1-eGFP and LHP1(W129C)-eGFP are localised to nuclei when expressed in *N. benthamiana*. While LHP1(KR9A)-eGFP also showed nuclear localisation, it was observed that LHP1(KR23A)-eGFP was located in both the nucleus and cytoplasm. This suggests that the bi-partite NLS disrupted in KR9A mutations perhaps functions redundantly with another NLS. Indeed a monopartite NLS located further towards the C-terminus is disrupted by the KR23A mutation but not the KR9A mutation (Fig. 5.3).

To re-localise LHP1(KR23A) exclusively to the nucleus, a nuclear localisation signal (NLS) was added to either the N-terminus, or to the C-terminus. For the C-terminus two positions were tried: either between LHP1 and eGFP, or else after eGFP (Fig. 5.8D). NLS sequences often contain K/R residues. Adding an NLS to the LHP1(KR23A) mutant could potentially therefore cause the re-establishment of RNA-binding. Therefore, in addition to the commonly-used SV40 NLS (KKKRK), two putative NLS sequences that contain fewer K/R residues [331] were tested —to minimise the number of K/R residues added (Fig. 5.8C). These NLS peptides are referred to as class III and class V, respectively. The class V NLS is reported to be specific to plants [331]. Confocal images of *N. benthamiana* plants expressing LHP1(KR23A)-NLS-eGFP with these different NLS sequences are shown in Figure 5.8D. The SV40 NLS was able to re-localise LHP1(KR23A) exclusively to the nucleus regardless of its

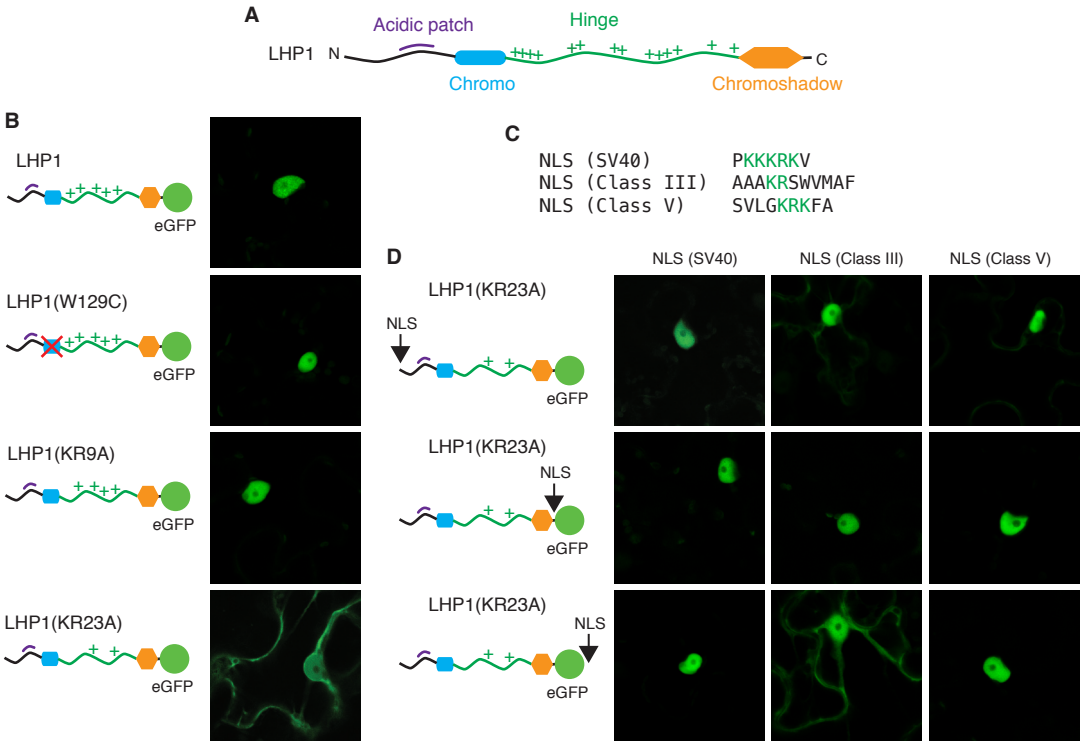


Figure 5.8: Subcellular localisation of Arabidopsis LHP1 in *N. benthamiana*. (A) Schematic of LHP1 protein. (B) Confocal microscopy images of LHP1, wild-type and mutants, transiently expressed in *N. benthamiana* leaves. Schematic diagrams represent positively charged residues in the hinge region with green '+' symbols. Mutation to the chromodomain residue W129 is indicated with a red cross. (C) Sequences of nuclear localisation signal (NLS) peptides used. (D) Restoration of exclusive nuclear localisation of LHP1(KR23A) by adding a nuclear-localisation signal peptide. Top panels indicate N-terminal NLS. Central panels indicate NLS between LHP1 and eGFP. Lower panels indicate C-terminal NLS after eGFP.

position in the fusion protein. By contrast, the class III NLS was only functional when placed between LHP1 and eGFP. The class V NLS was functional in all positions, but weak cytoplasmic localisation was detected for the N-terminal position. Based on these results, LHP1 constructs with either the SV40 or class V NLS sequence inserted between LHP1 and eGFP were both selected for transformation into Arabidopsis.

5.3.3 *Selection of parental line for transformation*

To ensure high expression of *FLC* in non-vernalising conditions, a genetic background containing an active *FRIGIDA* (*FRI*) was chosen (Sec. 1.2, Fig. 1.2, p. 24).

lhp1-6 contains a T-DNA insertion in exon 2 at the genomic LHP1 locus (Fig. 5.7). *lhp1-6* plants display an *lhp1* null phenotype, consistent with complete loss of function or lack of expression. In contrast, *lhp1-3* is a point mutation that generates a stop codon in the hinge region, in place of Q280 [223] (Fig. 5.7). While *lhp1-3* plants also show the same null phenotype, it is possible that a truncated LHP1 protein could be expressed in these lines. This truncated LHP1 would carry a wild-type chromodomain, which is capable of binding to H3K27me3 *in vitro* [150]. To avoid the complication associated with this truncated LHP1 protein interfering with an LHP1 protein expressed from a transgene, constructs were transformed into *FRI lhp1-6*.

lhp1 null mutants show pleiotropic phenotypes including curled leaves, early flowering, small plant size, reduced root growth, reduced cell expansion in leaves, and conversion of the shoot apical meristem to a terminal flower [222, 223, 316]. These phenotypes are assumed to be due to a failure to repress many Polycomb target genes in *lhp1* mutants. In particular, early flowering has been attributed to a failure to repress *FLOWERING LOCUS T* (*FT*) [332], and the curled leaf phenotype is related to ectopic expression of *APETALA3* (*AP3*) and *AGAMOUS* (*AG*) [151, 223]. Unlike in PRC2 mutants such as *curly leaf* (*clf*), inflorescence and floral organisation are normal in *lhp1* mutants .

5.3.4 *Flowering time phenotypes*

FRI lhp1-6 plants were transformed with *LHP1* constructs in two batches. Seeds from the first batch were sown on soil for selection in December 2014. After selection, resistant plants were vernalised for 3 weeks to accelerate flowering (Fig. 5.9A). The second batch were sown for selection in February 2015. Warm growth conditions and long days in the green-

house throughout March and April meant that plant development was fast and vernalisation was not required for the second batch (Fig. 5.9B). Transgenic *LHP1* plants showed a range of flowering-time phenotypes at the T1 generation - from very early, like parental *FRI lhp1-6*, to very late, like Col-FRI (Fig. 5.9). The control lines, Col-FRI and *FRI lhp1-6*, did not undergo selection for herbicide resistance and consequently showed more rapid development than corresponding transgenic lines. The absolute flowering time should therefore not be directly compared between non-transgenic and transgenic lines. However, the relative differences between Col-FRI and *FRI lhp1-6* indicate the impact of a functional copy of *LHP1* on flowering time in the Col-FRI background (Fig. 5.9).

In the first batch of transformants, lines containing wild-type *LHP1-eGFP* flowered significantly later than equivalent lines with the W129C or KR23A mutations (*t*-test, $p < 10^{-15}$) (Fig. 5.9A). This indicates that transgenic LHP1 is able to rescue the *lhp1* flowering time phenotype and that complementation requires the tryptophan at position 129 (W129C), and also the positively charged residues in the hinge region (KR23A). Based on subcellular localisation results in *N. benthamiana*, LHP1(KR-23A) is not expected to be exclusively nuclear-localised. This likely accounts for the inability of *LHP1(KR23A)* to complement the *lhp1* mutation. However, plants containing *LHP1(KR23A)-NLS(SV40)-eGFP* and *LHP1(KR23A)-NLS(Class V)-eGFP* also flowered significantly earlier that those carrying the corresponding wild-type LHP1 constructs (*t*-test, $p < 10^{-15}$) (Fig. 5.9A). Similar results were obtained when comparisons were made between transgenic *LHP1* and *LHP1(KR23A)-NLS* lines with the 3xHA tag (Transformation batch two, Fig. 5.9B). Since early flowering in *lhp1* plants is due to a failure to repress *FT* [333], this suggests that LHP1 RNA-binding is critical for repression of the Polycomb target gene *FT*.

In the second batch of transformants, early flowering was also observed for *LHP1(KR9A)* lines, carrying either the 3xHA or eGFP tag

(Fig. 5.9B). LHP1(KR9A) showed nuclear localisation in *N. benthamiana* (Fig. 5.8) and bacterially expressed LHP1(KR9A) showed a 4- to 8-fold reduction in affinity for RNA in EMSA (Fig. 5.5B, p. 188). Together with the results of the KR23A mutation, this suggests that the RNA-binding ability of LHP1 is crucial for function.

Confirmation of subcellular localisation in Arabidopsis. To confirm earlier results obtained in *N. benthamiana*, confocal microscopy was used to observe the subcellular localisation of LHP1-eGFP fusions in root meristems of T2 *LHP1* transgenic Arabidopsis plants. eGFP-tagged protein was easily detected and subcellular localisation agreed with previous results obtained in *N. benthamiana*: wild-type, W129C, and KR9A showed nuclear localisation whereas KR23A localisation was nuclear and cytoplasmic (excluded from the nucleolus) (Fig. 5.10). As observed in *N. benthamiana*, Both the SV40 and the Class V NLS, were able to restore the exclusive nuclear localisation of LHP1(KR23A) in Arabidopsis.

5.3.5 *Morphological phenotypes*

Figures 5.11 and 5.12 show representative plants from non-vernalised *LHP1* transgenic lines relative to wild-type and parental *FRI lhp1-6* after 4 and 8 weeks of growth, respectively. The plant size phenotype is fully rescued in *LHP1* and *LHP1-NLS* transgenic plants, and partly rescued in *LHP1(W129C)* and *LHP1(KR9A)* plants. However, *LHP1(KR23A)-NLS* plants are indistinguishable from parental *FRI lhp1-6* plants. In contrast to the flowering time results, this suggests that *LHP1(W129C)* and *LHP1(KR9A)* are able to partially complement the *lhp1-6* mutation with respect to the plant size phenotype. It can also be seen that parental *FRI lhp1-6* and *LHP1(KR23A)-NLS* plants display a downward curled leaf phenotype, which appears to be completely rescued in *LHP1(W129C)* plants and almost completely rescued in *LHP1(KR9A)* plants.

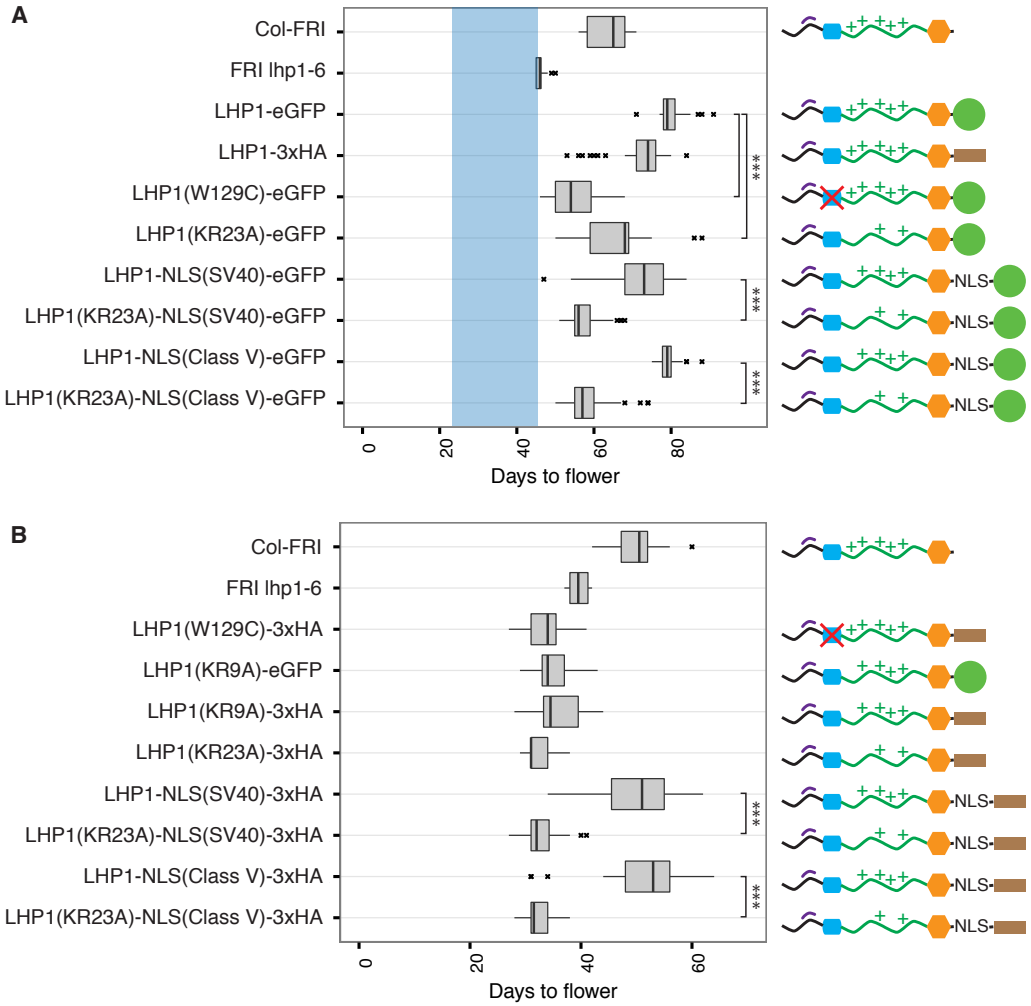


Figure 5.9: Flowering time for T1 LHP1 transgenic lines. Flowering time measured in days to flower, as described in Sec 2.5.2 (p. 71). For each transgenic construct, $n > 40$. For Col-FRI and *FRI lhp1-6*, $n = 12$. (A) Transformation batch 1. After selection of transformants using herbicide resistance, plants were grown for a further 3 weeks in short days at 8°C (indicated with shaded blue box) before being returned to glasshouse growth conditions ($\approx$ 16 – 25°C, Dec. 2014 - Feb. 2015, Norwich). Several *FRI lhp1-6* plants flowered during vernalisation treatment. This flowering was not measured so is indicated at the end of the cold treatment. (B) Transformation batch 2. Plants remained in glasshouse growth conditions throughout selection until flowering (Feb.-Apr. 2015, Norwich). *** indicates highly significant difference in flowering time (t-test, unpooled variance estimates, $p < 10^{-15}$). Schematic diagrams of LHP1 protein represent positively charged residues in the hinge region with green '+' symbols. Mutation to the chromodomain residue W129 is indicated with a red cross. The brown box represents the 3xHA tag, while the green circle represents eGFP.

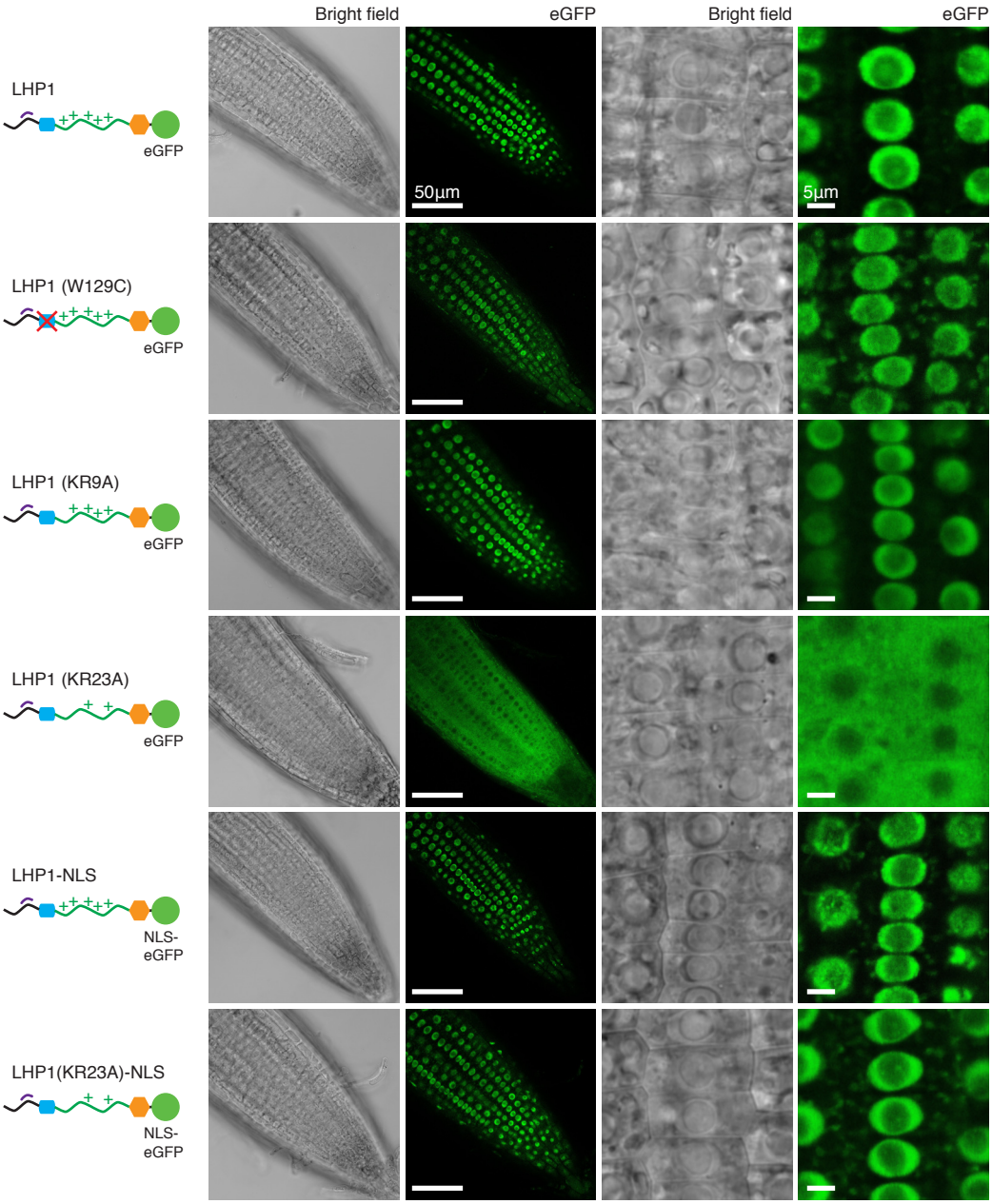


Figure 5.10: Subcellular localisation of transgenic LHP1-eGFP in Arabidopsis. Left to right, Diagram of the LHP1 protein (as in Fig. 5.9); bright-field and confocal eGFP images of Arabidopsis root meristem; zoomed bright-field and confocal eGFP images of meristematic cells. Images are individual focal planes. Sizes of scale bars (50 µm and 5 µm, respectively) are indicated in the top panels.

W129C is a mutation in one of the core residues in the conserved aromatic cage of the chromodomain (Fig. 5.3). SPR assays suggest that this greatly reduces the ability of LHP1 to bind to H3K27me3 *in vitro* (Fig. 5.6). The observation that *LHP1(W129C)* plants are slightly larger and have flatter leaves than parental *FRI lhp1-6* suggests that LHP1 protein may not absolutely require the ability to bind H3K27me3 to be functional. This result is in contrast to the interpretation of previous studies, which reported a null *lhp1* phenotype for the *lhp1-7* mutation —a splice-site mutation causing W129 to be converted to CCER [312]. One possibility is that the *lhp1-7* mutation causes a more dramatic change in structure to the LHP1 chromodomain than was anticipated in [312]. This could explain the difference in phenotype between the W129 to CCER mutation (*lhp1-7*) and the W129C mutation studied here. It is also possible that the W129C mutation considered here does not completely abolish binding of H3K27me3. However, previous work showed that the equivalent mutation in *S. pombe* $HP1^{Swi6}$ eliminates the ability of the protein to specifically recognise a H3 peptide tri-methylated on Lys-9 [326]. Yet another explanation for this discrepancy is that the phenotype reported in the present work may only be apparent in a background with functional *FRIGIDA*. The transgenic lines developed in the present study are in *FRI lhp1-6*, whereas the *lhp1-7* mutation was obtained in a *fri* mutant background (Col-0) [312]. Flowering is delayed in *FRI lhp1-6* when compared to *fri lhp1-6* (data not shown), indicating that *lhp1* phenotypes can be altered in a *FRIGIDA*-dependent manner. Further experiments will be required to distinguish these possibilities, as discussed in Sec. 5.4. In this context, it is interesting to note that H3K9-methylation is necessary but not sufficient to recruit HP1 in Drosophila [334, 335]. In the case of Drosophila, protein-protein interactions are also required [334]. Therefore multiple mechanisms may act in a partially redundant manner to recruit LHP1 to its targets.

Similar results were observed for *LHP1(KR9A)* as for *LHP1(W129C)*. While *LHP1(KR9A)* is non-functional in delaying flowering, these plants

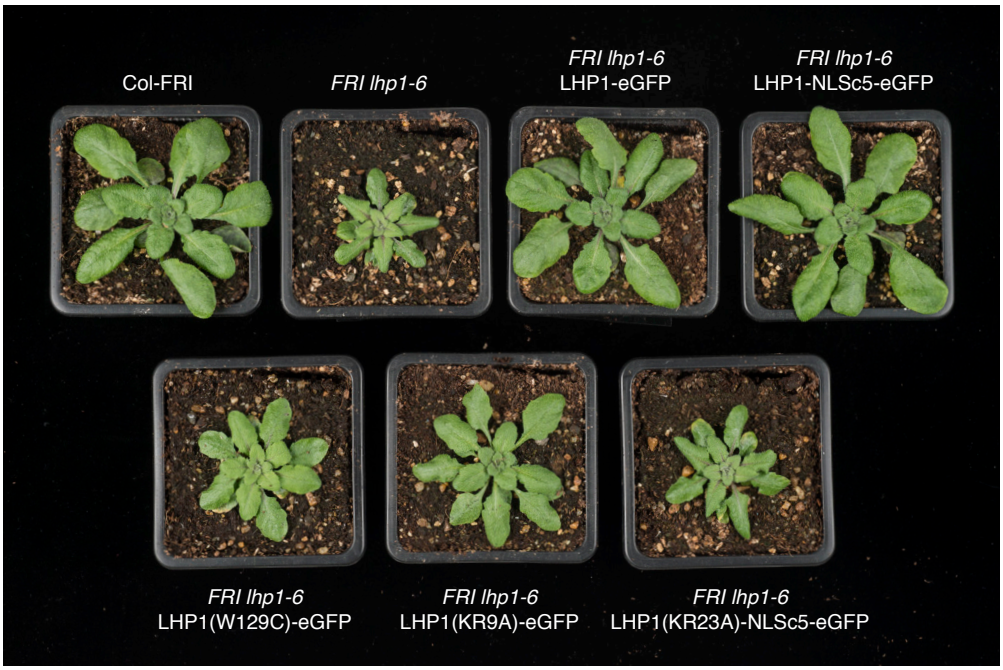


Figure 5.11: Morphological phenotypes of transgenic LHP1 plants. Representative plants after 4 weeks of growth in warm conditions.

show a larger overall size and distinct leaf morphology from parental *FRI lhp1-6* lines, and also *LHP1(KR23A)-NLS* lines (Figs. 5.11, 5.12). This may indicate that the RNA-binding ability of LHP1 is essential only for LHP1 to repress a subset of its target genes. Alternatively, the KR9A mutation may cause a small de-repression of all LHP1 targets, while a larger de-repression is observed in plants lacking *LHP1* or those carrying *LHP1(KR23A)-NLS*.

5.4 Summary and discussion

LHP1 is able to distinguish RNA from DNA *in vitro* (Fig. 5.2B, p. 183) and its ability to bind RNA through the hinge domain plays a key role in its *in vivo* function (Figs. 5.9, 5.11, 5.12). Understanding the mechanistic details of this behaviour will require further experiments. In particular, whether or not the RNA-binding mutant LHP1 proteins are able to restore wild-type spreading of H3K27me3 at *FLC* and maintenance of *FLC* repression after cold exposure is of considerable interest.

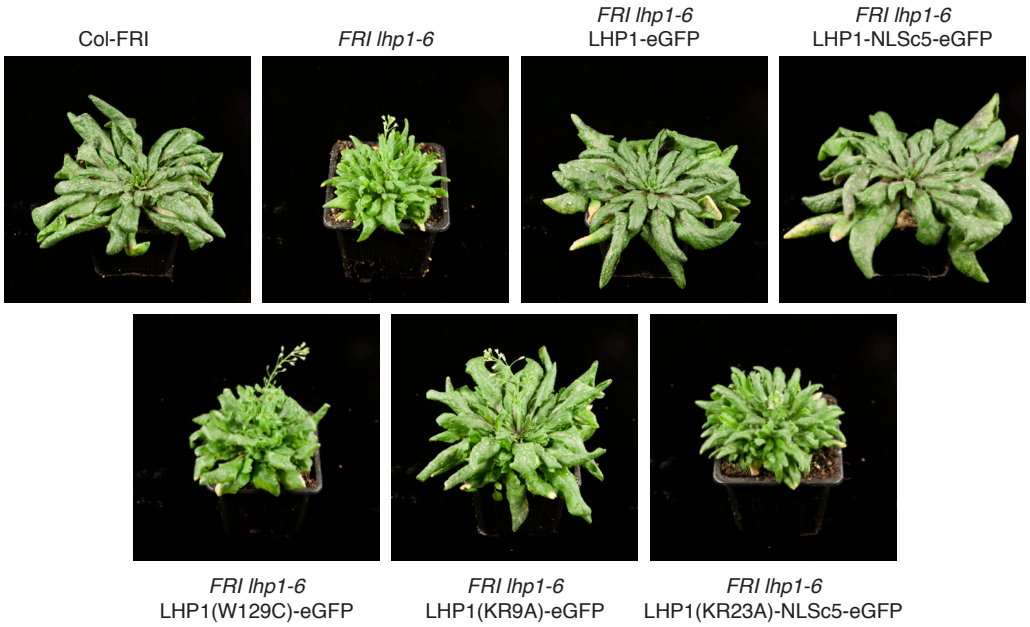


Figure 5.12: Morphological and flowering time phenotypes of transgenic LHP1 plants. Representative plants after 8 weeks of growth in warm conditions. *FRI lhp1-6*, *LHP1(W129C)*, *LHP1(KR9A)* and *LHP1(KR23A)-NLS* plants are flowering while Col-FRI, *LHP1* and *LHP1-NLS* plants are not.

The mechanistic understanding of nucleation and spreading at *FLC* provides an ideal opportunity to investigate how LHP1 RNA-binding contributes to maintenance of gene repression at a Polycomb target gene.

The unexpected results of partial complementation of the *lhp1* mutation in *LHP1(W129C)* plants also requires further investigation. If it is really the case that the *in vitro* affinity of LHP1(W129C) for H3-K27me3 is reduced to that of unmodified H3, then this mutant provides the opportunity to more precisely understand the role of H3K27-me3-recognition in LHP1 function. To pursue this, quantitative assays could be used to calculate the dissociation constants for bacterially purified LHP1 and LHP1(W129C) for different histone peptide or nucleosomal substrates. Surface-plasmon resonance (SPR) experiments presented here (Fig. 5.6D) indicate a complex mode of binding of full-length LHP1 to histone peptide immobilised on the sensor chip. This complex interaction prevented calculation of dissociation constants us-

ing SPR. Similar complexity has also been observed in *in vitro* assays with HP1^{Swi6} [330]. Fluorescence polarisation (FP) [336] has been previously used to study the chromodomain of LHP1 [150] and other HP1 proteins [321, 326]. It would be interesting to recreate the W129C mutation in the chromodomain construct and to use FP to determine the affinity of this protein for different histone peptides, including H3K27me3. Further *in vitro* characterisation of the full-length LHP1 proteins would also allow the dimerisation status of LHP1 *in vitro* to be determined, and the effect of the RNA-binding mutations on dimerisation to be assessed.

Regarding the RNA-binding of LHP1. The LHP1(KR9A) and LHP1(KR23A) mutants generated in this work show reduced affinity for RNA and abolishment of RNA-binding, respectively (Fig. 5.5). *In vivo*, these proteins were unable to delay flowering in *FRI lhp1-6* plants (Fig. 5.9). It was argued that this is likely to be due to a failure to repress the floral activator *FT*. Like *FLC*, *FT* is a PRC2 target gene that accumulates H3K27me3 in the repressed state [333].

Several possibilities can be imagined by which RNA-binding is required for LHP1 to repress PRC2 target genes. In *S. pombe*, it was proposed that nascent transcripts from heterochromatic loci are bound by HP1^{Swi6}, which causes dissociation of HP1^{Swi6} from chromatin, and targeting of the HP1^{Swi6}-bound RNA transcript for degradation [147]. When RNA-binding was perturbed in HP1^{Swi6}, H3K9-methylation remained intact but RNA from a heterochromatic reporter gene accumulated, and was even translated into protein [147]. By analogy with *S. pombe*, H3K27me3 may therefore still accumulate at LHP1 target genes when RNA-binding mutants are expressed, but the functional outcome of H3K27me3 (gene repression) may be impaired. This is illustrated in Figure 5.13A. No direct interaction between LHP1 and RNA processing complexes has yet been reported.

At *FLC*, H3K27me3 spreading but not nucleation was impaired in *lhp1* mutants (Fig. 3.6, p. 98). In addition, it has been observed that

H3K27me3 domains genome-wide in Arabidopsis are co-incident with transcription units [293]. This is tantalising evidence of a link between transcription and the spreading of H3K27me3. PRC2 has been shown to interact with LHP1 [153], and in the current work it was shown that LHP1 can interact with RNA. It is therefore possible that LHP1 proteins that are recruited primarily at sites of H3K27me3 nucleation, could be spread slowly in discrete jumps along the chromatin, facilitated by transient interactions with nascent RNA (Fig. 5.13B: Hopping and spreading). Interactions between LHP1 and RNA could lead to stabilisation of PRC2 complexes in regions adjacent to nucleation sites, and allow stochastic addition of H3K27me3 in these regions. The PRC2/H3K27me3-based feedback [337] could then stabilise the 'spread' H3K27me3 state. How LHP1 could be integrated into the mathematical model of *FLC* chromatin is discussed in Chapter 6 (Sec. 6.3).

In vitro studies of $HP1^{Swi6}$ were recently used to propose a model in which $HP1^{Swi6}$ oligomerises to spread linearly along chromatin [326]. In this model, the binding of $HP1^{Swi6}$ tetramers to one nucleosome leads to co-operativity in the binding of further $HP1^{Swi6}$ dimers on neighbouring nucleosomes. This corresponds to the popular model in which HP1 molecules coat entire regions of heterochromatin.

In human cells, however, quantitative imaging indicates that HP1α/β concentrations in heterochromatin are in the low micromolar range [338], while nucleosomes exist at a concentration of 200-300 μM [339]. This means that HP1α/β do not exist in sufficient quantities to 'cover' the chromatin fibre. Moreover, quantitative measurements of H3K9me3 indicate that this histone modification exists at only ≈ 38% of histone H3 in pericentric heterochromatin [104, 340]. Thus not all histones in heterochromatin contain H3K9me3, and not all are bound by HP1.

This is similar to the result suggested by analysis of H3K27me3 accumulation in HeLa cells in Chapter 4 (Sec. 4.2.3): that H3K27me3 levels are not saturated at H3K27me3-enriched chromatin domains. There-

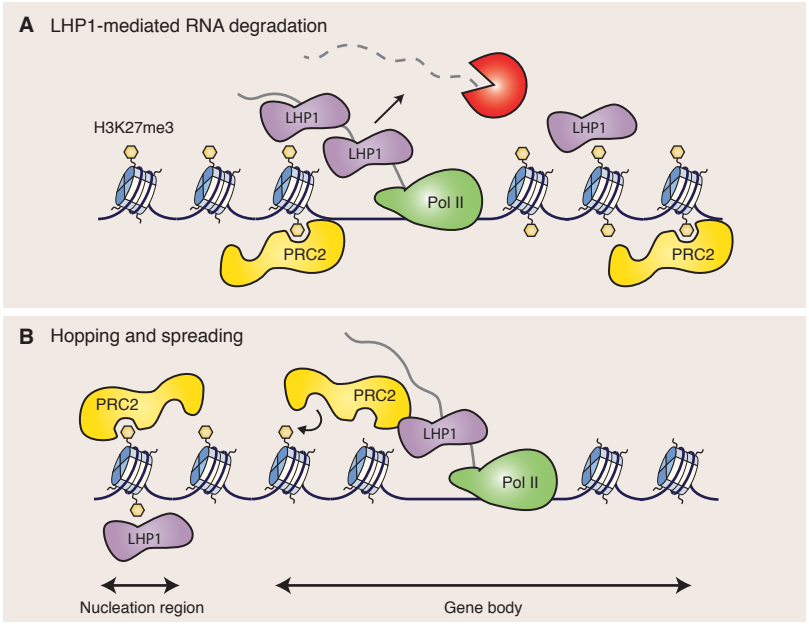


Figure 5.13: Possible roles for LHP1 RNA-binding in repression of Polycomb target genes. (A) LHP1-mediated RNA degradation. Analogous to the proposed mechanism of $HP1^{Swi6}$ in *S. pombe*, LHP1 could interact with infrequently produced transcripts at repressed loci to ensure that they are targeted for degradation. (B) Hopping and spreading. LHP1 could interact with nascent RNA to allowing spreading in discrete 'jumps' from the site of initial recruitment (e.g. nucleation region) into the gene body. The interaction with PRC2 would gradually facilitate H3K27me3 accumulation and further rounds of LHP1 recruitment through H3K27me3-binding.

fore, based on these data, a model in which densely-packed LHP1 dimers oligomerise to spread along chromatin is not favoured. Instead, individual LHP1 dimers likely make contacts with H3K27me3-containing histones individually, or form tetramers between neighbouring pairs of nucleosomes.

Another experimental result to keep in mind when considering the mechanism of LHP1-mediated stabilisation of PRC2-repression is that LHP1 binding to chromatin is highly dynamic. *In vivo* studies using fluorescence recovery after photobleaching (FRAP) indicate that the average residence time of an LHP1 molecule on chromatin is on the order of 10 seconds [341]. While quantitative differences exist among HP1 family members in other species, all are similarly dynamic in their association with chromatin [338, 342–344]. Thus any model that accounts

for the role of LHP1 in spreading and maintenance of chromatin must account for this rapid exchange of individual LHP1 molecules.

Finally, it is interesting to consider how well conserved the RNA-binding of LHP1 proteins is among plants. The hinge region of HP1 proteins is generally poorly conserved. However, the two most highly conserved regions of the hinge across the plant kingdom are those that contain the two putative nuclear-localisation (NLS) sequences [345]. An alignment of several plant LHP1 protein sequences from diverse species is shown in Figure 5.14 and summarised in Table 5.1. These two conserved NLS-containing regions within the hinge are annotated as Basic 1 and Basic 2, respectively. The KR9A mutation disrupts the Basic 1 region, while KR23A disrupts both the Basic 1 and Basic 2 regions.

It is intriguing to speculate that these NLS sequences have a dual function in both nuclear-localisation and RNA-binding. Indeed, it was shown that removing these NLS sequences in LHP1(KR23A) and then re-instating nuclear-localisation via a C-terminal NLS (in the LHP1(KR-23A)-NLS-eGFP construct) resulted in LHP1 protein that was nuclear localised but not functional in complementing the *lhp1-6* mutation.

As shown in Table 5.1, both the length of the hinge region and the overall proportion of K/R residues (≈13-19%) is reasonably well conserved. In Fig. 5.14A, it can also be seen that the positions of the two basic regions within the hinge are also reasonably conserved. Therefore, LHP1 RNA-binding has the potential to be a general phenomenon, and may contribute to PRC2 silencing across the entire plant kingdom.

5.5 Plant materials

Col-FRI was described in Sec. 2.5.3. *lhp1-6* (SALK N511762) [346] was obtained from the Nottingham Arabidopsis Stock Centre (NASC). Plants resistant to Kanamycin were crossed to Col-FRI. F2 plants homozygous for *FRI* and *lhp1-6* were obtained by PCR-based genotyping. Primer sequences and instructions for use are provided in Sec. 7.2.

Order	Species	K/R	Length	% Total
Brassicales	*Arabidopsis thaliana*	20K / 15R	222	15.8%
Cucurbitales	*Cucumis sativus* (Cucumber)	19K / 12R	197	15.7%
Fabales	*Glycine soja* (Soybean)	13K / 10R	122	18.9%
Malpighiales	*Populus euphratica* (Poplar)	22K / 10R	228	14.0%
Myrtales	*Eucalyptus grandis* (Eucalyptus)	19K / 12R	228	13.5%
Poales	*Zea mays* (Maize)	16K / 19R	199	17.6%
Rosales	*Malus domestica* (Apple)	18K / 21R	230	17.0%
Solanales	*Solanum tuberosum* (Potato)	15K / 15R	204	14.7%
Vitales	*Vitis vinifera* (Grape)	17K / 20R	213	17.4%

Table 5.1: Conservation of charged hinge regions among plants. The number of K/R residues between the chromodomain and chromoshadow domain of LHP1 in selected plant species of various taxonomic orders. Alignment method described in Sec. 5.6.11.

5.6 Methods

Transformation of Arabidopsis and selection of transgenic lines is described in Sec. 2.5.4. 48 T1 Arabidopsis plants were kept for each of the 16 LHP1 constructs transformed (Table 5.2). Flowering time measurements as described in Sec. 2.5.2.

5.6.1 *General cloning*

DH5α or TOP10 (Invitrogen) *E. coli* strains were used for cloning, while BL21 Rosetta pLysS (Novagen) was used for protein expression. Competent cells were generated and transformed by heat-shock as described in Sec. 2.5.4.

Site-directed mutagenesis was done using the Q5 site-directed mutagenesis kit (New England Biolabs). Mutagenesis primers were designed using the online tool NEBasechanger (http://nebasechanger.neb.com). Gel extraction was performed using the Wizard gel extraction kit (Promega) and cloning with type-I endonucleases (New England Biolabs or

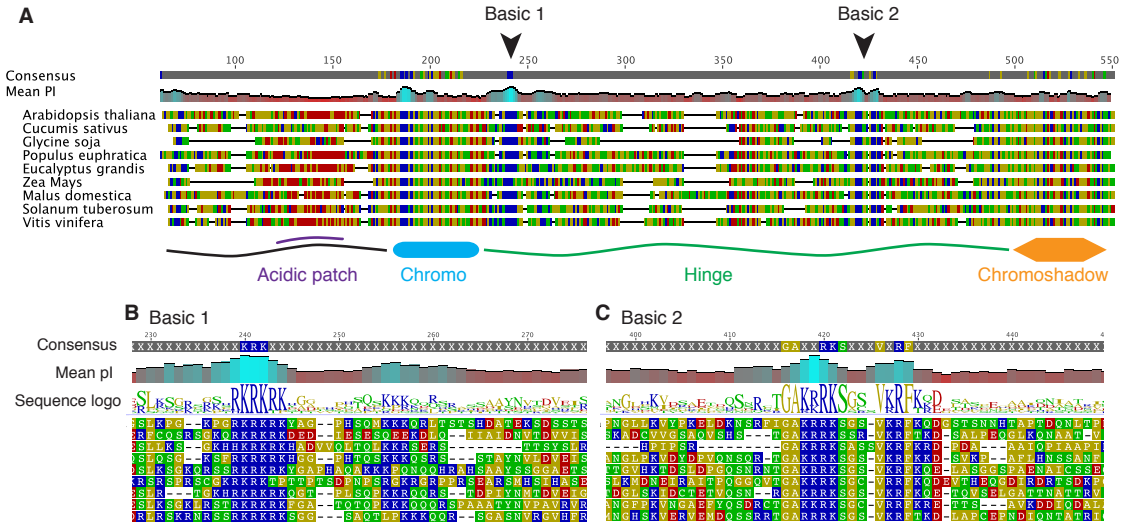


Figure 5.14: Protein sequence alignment of plant LHP1 proteins. (A) Alignment of the putative plant LHP1 homologues listed in Table 5.1, generated as described in Sec. 5.6.11. Blue residues are basic, while red are acidic, yellow are non-polar and green are polar. The mean isoelectric point (pI) over a 5 residue sliding window is shown above the alignment. The domain structure of LHP1 is indicated underneath. The *P. euphratica* sequence has an extended N-terminus that is omitted from this diagram. Two conserved basic regions (Basic 1,2) are observed in the hinge region. (B,C) Zoom of the alignment in A, with the sequence logo shown around the conserved basic regions.

Thermo Scientific) was performed according to manufacturer's instructions.

Golden Gate cloning with type-II endonucleases (New England Biolabs) [347] was performed according to the following protocol: 100 ng of each plasmid was combined with 1 µL T4 ligase (2 000 000 U/mL, M0202M, New England Biolabs) and 1uL BsaI (level one) or BpiI (level two) in T4 ligase buffer containing 1× bovine serum albumin (New England Biolabs). Reactions were transferred to a thermal cycler for 25 cycles of 37°C, 3 min; 16°C, 4 min.

5.6.2 *LHP1 constructs for bacterial expression*

The coding sequence of full-length LHP1 protein was amplified by PCR from Col-0 cDNA and cloned into the XhoI site of pGEX-4T-1 (GE Healthcare). This generated an N-terminal glutathione-S-transferase (GST)-LHP1 fusion with the endogeneous LHP1 stop codon. The chro-

modomain mutant LHP1(W129C) was generated from the wild-type LHP1 expression construct using site-directed mutagenesis. Similar mutagenesis was also used to make a synonymous mutations at +489 bp to create a SexAI site between the chromodomain and the hinge region. The sequence of the mutant hinge regions for LHP1(KR9A), LHP1(KR23A) and LHP1(KR33A) were then chemically synthesized as 750 bp fragments (GeneStrings, Invitrogen) and cloned into pGEX-4T-1 LHP1(SexAI) using SexAI/AfeI. All constructs were confirmed by test digests and sequencing.

5.6.3 *Expression of GST-LHP1 in E. coli*

E. coli BL21 Rosetta (DE3) pLysS cells were transformed with pGEX plasmid containing GST-LHP1 and incubated for 16 hours at 37°C on LB media plates containing 25 μg/mL Chloramphenicol and 100 μg/mL Ampicillin. The following day, ~10 colonies from the transformation plate were used to inoculate 10 mL liquid LB media containing antibiotics. After 4 hours rotating at 37°C, these starter cultures were diluted into 1 L of LB containing antibiotics (5 L flask, media pre-warmed to 37°C) and grown for a further 4-8 hours on shaker until OD_{600}=0.7. Cultures were then transferred to 20°C and, after cooling, protein expression was induced by adding Isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.25 mM (induction at 20°C, OD_{600}=1.0). Cultures were incubated for a further 16 hours at 20°C and cells harvested the following day by centrifugation in 1 L pots at 2500 × *g* (15 min). Cell pellets were then resuspended in 35mL phosphate-buffered saline (PBS, 10 mM PO_4^{3-} pH 7.4, 137 mM NaCl, 2.7 mM KCl) and transferred to 50 mL tubes. Cells were once again collected by centrifugation at 2500 × *g* (10 min) and pellets were weighed before flash freezing in liquid nitrogen for storage at -70°C.

5.6.4 *Purification of recombinant LHP1 from E. coli*

A description of all buffers follows these methods. Cell pellets were resuspended in 5 mL GST lysis buffer per gram of cells and disrupted by sonication on ice using a Branson sonifier (10 × 30 sec, 30-40% duty). Lysate was then cleared by centrifugation at 6000 × *g* (20 min). 1-2 mL glutathione-sepharose fast-flow resin (GE Healthcare) was equilibrated in GST lysis buffer in a disposable chromatography column (Bio-Rad). Lysate was then filtered (0.45 µm) onto the column and incubated with the resin at 4°C with gentle rocking (15 min). Protein-bound resin was washed with 5 column volumes (C.V., 1-2 mL) GST lysis buffer, 20 C.V. GST wash buffer, and 10 C.V. GST low salt buffer. GST fusion proteins were eluted by incubation of resin with 1 C.V. of GST elution buffer (10 min, 4°C). Elution was repeated 4-5 times and eluates were pooled to give a final volume of 5-10 mL.

Pooled eluates were loaded on 5 mL HiTrap Q FF column (GE Healthcare), pre-equilbrated in QA buffer using an ÄKTA FPLC protein purification system (GE Healthcare). Proteins were then eluted in 2.5 mL fractions using a 0-70% gradient of QA and QB. Chromatography was performed at 4°C. Fractions containing GST-LHP1 were identified by UV absorbance on the ÄKTA and confirmed by SDS-PAGE.

Pooled fractions were concentrated and exchanged into buffer QA using centrifugal concentrators (Amicon, 15 mL, 30 kDa molecular-weight-cut-off) with repeated 3-4 min spins at 4000 × *g*. Final concentrations were typically 0.5-1 mg/mL.

GST lysis buffer: 25 mM Tris-HCl pH 7.5, 500 mM NaCl, 1% (w/v) Triton X-100, cOmplete Protease Inhibitor Cocktail (Roche).

GST wash buffer: 25 mM Tris-HCl pH 7.5, 500 mM NaCl, 0.1% (w/v) Triton X-100, cOmplete Protease Inhibitor Cocktail (Roche).

GST low salt buffer: 25 mM Tris-HCl pH 8.0, 100 mM NaCl.

GST elution buffer: 25 mM Tris-HCl pH 8.0, 100 mM NaCl, 50 mM L-Glutathione reduced (Sigma-Aldrich, G4251)

QA buffer: 25 mM Tris pH 7.5, 100 mM NaCl, 1 mM DL-Dithiothreitol (DTT, Sigma, D9779), 0.5 mM Ethylenediaminetetraacetic acid (EDTA)
QB buffer: 25 mM Tris pH 7.5, 1M NaCl, 1 mM DTT, 0.5 mM EDTA.

5.6.5 *Electrophoretic mobility shift assay*

Concentrated proteins were diluted to 10 µM in buffer QA and then diluted further to 2 µM into EMSA buffer (20 mM HEPES-KOH pH 7.5, 100 mM KCl, 0.05% NP-40). 1 µL of 100 nM Cy5-labelled RNA or DNA (Integrated DNA Technologies) was added to 9 µL protein and the reactions incubated for 30 min at room temperature. 4 µL 50% glycerol was then added and reactions were immediately loaded onto an RNAse-free 1.6% Tris-Borate Agarose gel. Gels were run at 80V for 30 min and scanned using a Typhoon fluorescence imager (GE Healthcare) using a photomultiplier tube voltage of 600V. Images were taken with 50 µm resolution and 'Medium' sensitivity, focussed +3 mm from the scan surface.

The sequence of the single-stranded RNA and DNA probes used for EMSA was Cy5-CUCCUCCGGCGAUAAGUACGCCUUUUCCUUACCUGGGUUU, with T exchanged for U in DNA probe. All probes were synthesised and purified by high performance liquid chromatography (HPLC) (Integrated DNA Technologies). Double-stranded DNA probe was generated by annealing a complementary unlabelled oligonucleotide to the labelled single-stranded DNA probe. Annealing was done by combining equimolar amounts of each probe, heating to 95°C, and then slowly cooling to room temperature.

5.6.6 *Peptide pulldown assay*

7 µg streptavidin-coated Dynabeads (MyOne T1, Invitrogen, 65601) were washed in binding buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.1% (w/v) NP-40), according to manufacturer's instructions. 1 µg of each of the biotinylated histone peptides was then incubated with the Dynabeads for 30 min at 4°C. Peptide sequences were derived from histone

H3 residues 21-44 (ATKAARSAPATGGVKKPHRYRPGK-Biotin) and were either unmodified (Anaspec, 64440) or carried a K27me3 (Anaspec, 64367) or K36me3 (Anaspec, 64441) modification. Unbound peptides were removed by washing beads five times with 1 mL binding buffer. 5 µg of each of the GST fusion proteins in binding buffer were then added to separate tubes containing the histone peptide-Dynabead complexes. Reactions were incubated for 1 hour at 4°C. After repeatedly washing the beads, bound proteins were eluted by denaturation in Laemmli buffer (2% (v/v) SDS, 10% (w/v) glycerol, 5% (v/v) β-mercaptoethanol, 0.002% (w/v) bromphenol blue, 0.125 M Tris pH 6.8), and heated to 90°C for 10 min. Pulldown reactions were then separated by SDS-PAGE and transferred to polyvinylidene difluoride membranes. GST fusion proteins were detected by immunoblot using an anti-GST antibody (abcam) followed a secondary antibody coupled to horseradish peroxidase (anti-mouse, Santa Cruz). Bands were visualised by chemiluminescence (SuperSignal West Femto, Pierce).

5.6.7 *Surface plasmon resonance*

Biotinylated histone peptide encoding H3 residues 21-44 (ATKAARK(me3)SAPATGGVKKPHRYRPGK-Biotin, 64367, Anaspec) was loaded onto a streptavidin-coated sensor chip (Series S Sensor Chip SA, BR100531, GE Healthcare) on the Biacore X100 instrument (GE Healthcare), according to manufacturer's instructions. HBS-EP+ buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.05% (v/v) Surfactant P20 (Tween 20)) was used for all experiments. Proteins were exchanged into this buffer before injecting onto the sensor chip and their concentrations adjusted to 340nM. In the program, proteins were injected for 120 sec. Following this, 5 minutes of dissociation was monitored before regeneration of the sensor chip surface using successive washes with 2.5M NaCl and 2M $MgCl_2$.

5.6.8 *LHP1 constructs for in planta expression*

Genomic LHP1 sequence from -2406 to +3164 bp (relative to ATG) was chemically synthesized in 3 distinct modules (upstream/5′UTR, CDS: ATG to TAA, and 3′UTR/downstream). Constructs were assembled using Golden Gate modular cloning [347]. All BsaI and BpiI sites were mutated in the synthesized sequences and a synonymous G to A mutation was included in exon 4 (+1015 bp) to create a SexAI site for subcloning the mutated LHP1 hinge domain fragments.

Similarly to the bacterial LHP1 expression constructs, fragments encoding KR9A, KR23A and KR33A mutations were cloned from synthesized fragments into the LHP1 CDS module using SexAI/AfeI. The W129C mutation was generated in the LHP1 CDS fragment using site-directed mutagenesis. Nuclear-localisation sequences were inserted as either N-terminal signal peptide modules or added directly to the eGFP or 3xHA fragment using site-directed mutagenesis. All constructs were confirmed by test digests and sequencing.

Assembled LHP1 constructs were transferred to a custom-made gateway donor plasmid (pL2V-GW-EXPORT2) containing Gateway attL1 and attL2 sites flanking the Golden Gate assembly product. Gateway L/R reaction (Invitrogen) was then used to transfer the LHP1 construct to pSLJ-DEST, a gateway destination plasmid based on pSLJ755I6, previously designed and constructed by Hailong An in the Dean lab.

All plant LHP1 expression constructs are listed in Table 5.2.

5.6.9 *Transient expression in N. benthamiana*

Golden gate level one plasmids contain T-DNA left and right border sequences and can be used without a selectable marker for transient expression in *N. benthamiana*. Electro-competent *Agrobacterium tumefaciens* GV3101 cells were transformed with a level one Golden Gate plasmid (Ampicillin-resistant) by electroporation. Transformants were selected on LB media plates containing 50 µg/mL Carbenicillin, 20 µg/mL Gentamycin and 50 µg/mL Rifampicin. 10mL cultures were grown from

Native	*N. benthamiana*	*A. thaliana*
LHP1-eGFP	Yes	Yes
LHP1-3xHA	No	Yes
LHP1(W129C)-eGFP	Yes	Yes
LHP1(W129C)-3xHA	No	Yes
LHP1(KR9A)-eGFP	Yes	Yes
LHP1(KR9A)-3xHA	No	Yes
LHP1(KR23A)-eGFP	Yes	Yes
LHP1(KR23A)-3xHA	No	Yes
N-terminal NLS		
NLS(SV40)-LHP1(KR9A)-eGFP	Yes	No
NLS(SV40)-LHP1(KR23A)-eGFP	Yes	No
NLSc3-LHP1(KR23A)-eGFP	Yes	No
NLSc5-LHP1(KR23A)-eGFP	Yes	No
C-terminal NLS (before tag)		
LHP1-NLS(SV40)-eGFP	Yes	Yes
LHP1-NLS(SV40)-3xHA	No	Yes
LHP1(KR23A)-NLS(SV40)-eGFP	Yes	Yes
LHP1(KR23A)-NLS(SV40)-3xHA	No	Yes
LHP1(KR23A)-NLSc3-eGFP	Yes	No
LHP1-NLSc5-eGFP	Yes	Yes
LHP1-NLSc5-3xHA	No	Yes
LHP1(KR23A)-NLSc5-eGFP	Yes	Yes
LHP1(KR23A)-NLSc5-3xHA	No	Yes
C-terminal NLS (after tag)		
LHP1(KR23A)-eGFP-NLS(SV40)	Yes	No
LHP1(KR23A)-eGFP-NLSc3	Yes	No
LHP1(KR23A)-eGFP-NLSc5	Yes	No

Table 5.2: LHP1 Constructs. All constructs based on the genomic LHP1 sequence from -2.2 kb upstream to +3.3 kb downstream of the TSS (Fig. 5.7). Constructs were cloned and transformed into *N. benthamiana* or *A. thaliana* using Agrobacterium-mediated transformation.

re-streaked single colonies for 14 hours. Cells were then collected by centrifugation (3000 × *g*, 20 min) and resuspended in 1 mL buffer containing 10 mM MES pH 5.6, 10 mM $MgCl_2$, 150 µM acetosyringone (D134406, Aldrich). Volumes were adjusted to achieve an OD_{600} = 1.0. 0.25 mL of this resuspended culture was infiltrated using a syringe into young expanding leaves of *N. benthamiana*. Leaves were harvested and imaged 54 hours after infiltration.

5.6.10 *Confocal imaging of LHP1-eGFP*

Imaging was performed using a 20x/0.7 NA multi-immersion lens, with water as the immersion fluid on a Leica TCS SP5 confocal microscope equipped with Leica HyD Hybrid detectors. The eGFP fluorophore was excited using a laser with 458 nm wavelength and emission collected between 500-550 nm. Laser power was set to 15% (20% for Arabidopsis) and the HyD detector sensitivity was adjusted to capture individual images. *N. benthamiana* images were collected at a resolution of 1024 × 1024 pixels, while Arabidopsis images were 2048 × 2048 pixels.

5.6.11 *Alignment of LHP1 homologues*

Sequences of LHP1 homologues were retrieved by BLASTP searches of the NCBI database, as described in [345]. LHP1 homologues from diverse taxonomic order were then selected.

Alignment of sequences was performed using the ClustalW algorithm [348] within the Geneious software package [349]. The BLOSUM cost matrix was used with gap open cost of 10 and a gap extend cost of 0.1, following [345]. The number of K/R residues in the hinge region, listed in Table 5.1, was obtained by counting the number of K/R residues between the chromodomain and the chromoshadow domain.

6 DISCUSSION & CONCLUSIONS

This thesis has focused on the molecular mechanisms of how gene expression states are remembered. The classic example of vernalisation — the memory of prolonged cold exposure in plants, was used as a paradigm for exploring the interactions between chromatin and transcriptional regulation.

In concluding the thesis, links between the different chapters are explored in more detail, and results are interpreted in the context of other studies in the field of epigenetics. In particular, Chapter 2 —the observation of cis memory, is discussed in the context of the model of transcription and Polycomb silencing developed in Chapter 4, which suggested a conceptual integration of trans and cis memory. Chapters 3 and 4 considered the *FLC* nucleation region, and PRC2 repression without an 'activating' mark, respectively. How these models could be integrated into a whole-gene model of *FLC* chromatin is also discussed. Finally, the thesis concludes with an outlook for the field of epigenetics and some general comments regarding the use of mathematical modelling in biology.

6.1 Implications of the observation of cis memory

In Chapter 2, it was shown that prolonged cold exposure causes *FLC* loci to autonomously switch from an active to a repressed expression state, both of which are mitotically heritable. It was then observed that active and repressed *FLC* loci can co-exist in the same cell, and that the expression state of each locus is heritable. From these results, it was concluded that the key inherited memory elements that determine *FLC* expression reside in the vicinity of the *FLC* gene itself.

Previous studies have shown that switching this locus from the active state to the repressed state is dependent on a PHD-PRC2 complex,

which delivers H3K27me3 to *FLC* chromatin [43]. The involvement of conserved PRC2 in this locally-encoded memory makes this finding relevant well beyond Arabidopsis.

In recent years, the field of epigenetics has been divided in opinion regarding the role of chromatin in memory of gene expression states [23, 29–32]. While most researchers in the field would agree that symmetric methylation of DNA at a CG dinucleotide can instruct its own inheritance [16, 20], the same consensus is not present for patterns of histone modifications. The key difference is that DNA methylation is covalently attached to the genetic material, and conceptually simple to transmit through DNA replication using an enzyme that converts a hemimethylated CG to a symmetrically methylated CG [22] (Fig. 1.1, p. 19). In contrast, histone modifications are carried on nucleosomes, which are turned over independently of replication, and shared randomly between the two daughter DNA strands at replication [25–27, 236]. This means that the histone modifications at any particular position are lost, on average, every other DNA replication. Patterns of histone modifications must therefore be distributed over several nucleosomes in order to act as heritable memory elements [28].

Prior to this work, it had been shown that PRC2 can methylate H3K27 [53], is allosterically activated by H3K27me [60], and that histones are shared between daughter DNA strands during replication [25]. This led to a model in which the inherited H3K27me histone modifications act as epigenetic memory elements, causing recruitment of PRC2 after DNA replication, and thereby re-establishing the repressive chromatin state each cell cycle [60, 107]. During the course of the current work, two further discoveries were published that lent support to this model. First, mutagenesis experiments in Drosophila showed that Lys-27 on H3 is required for the repressive action of PRC2 [64]. Second, using elegant genetics in *C. elegans*, PRC2-knockout embryos were generated in which the paternal set of chromosomes, but not the maternal set, contained H3K27me [236]. It was observed that H3K27me was passed on

to daughter chromosomes even in the absence of PRC2, and that H3K27me did not spread between the two sets of chromosomes. These latter experiments were noteworthy as they seemed to exclude a competing hypothesis that Polycomb proteins rather than histone modifications are inherited through DNA replication [350].

However, these experiments do not address whether epigenetic memory is stored in cis. *C. elegans* embryos do not activate their own transcriptional program until the larval stage, at which time the H3K27me marks from paternally-inherited chromosomes had been diluted in these experiments, and all chromosomes were effectively lacking H3K27me [236]. When similar experiments were performed in embryos containing PRC2, the inherited chromosomes that were initially devoid of H3K27me acquired H3K27me coincident with activation of the zygotic transcription program, presumably through *de novo* histone methylation [236]. Thus the question of whether inherited H3K27me can act as a memory element to re-establish a repressive gene expression state after DNA replication remained still unresolved.

To fill this gap in understanding, the studies presented in Chapter 2 focused on *expression* of a well-characterised PRC2 target gene, which has a requirement for long-term storage of epigenetic memory. As such, these studies represent the first observation of local storage of epigenetic memory in a PRC2-based system [208].

A similar approach could conceivably be applied to the study of other genes, to determine the generality of the phenomenon of cis epigenetic memory. One caveat, however, is that establishment of silencing must be somewhat stochastic in order to generate cells containing both active and repressed copies of the gene. High throughput technologies based on sequencing of RNA have looked for allele-specific expression genome-wide in F1 hybrid animals [351–354]. However, these approaches are limited by the requirement for sequence polymorphism to distinguish the two alleles (reviewed in [355]).

FLC can now be considered alongside X chromosome inactivation (XCI) and genomic imprinting as key examples of cis epigenetic memory. It is noteworthy that in all three cases (XCI, imprinting, and *FLC*), the key evidence excluding a trans epigenetic memory is simultaneous observation of active and repressed expression states within the same nucleus, and maintenance of this state through cell division [6]. Unlike for XCI and imprinting, however, DNA methylation can be excluded as the carrier of cis epigenetic memory for *FLC* [356].

It is important to note that the results presented in Chapter 2 do not inform as to the identity of the cis epigenetic memory elements. To show decisively that H3K27-methylated histones act as inherited memory elements at *FLC*, it would be necessary to show that *FLC* reactivation occurs when H3K27me3 marks are removed.

An experiment that may soon be technically feasible would be to provide a brief pulse of a H3K27me3-demethylase specifically targeted to endogenous *FLC*. Targeted tethering of protein domains with the ability to modify histones is referred to as 'epigenome-editing' [357]. This technique has been used to deliver specific histone modifications to several genes in cultured human cells and primary neurons [358]. With the advent of CRISPR-Cas technologies[a] that readily facilitate tethering of proteins anywhere in the genome, studies are beginning to emerge in which histone modifiers such as histone acetyltransferases [360] or methyltransferases [361] are directed to specific genes or enhancers. *FLC* represents a prime candidate for such a study because so much is known already about its regulation. Based on the models proposed in this thesis, inducing targeted removal of H3K27me3 would result in loss of silencing at *FLC* loci repressed after cold exposure. Removal of the targeted demethylase after complete H3K27me3-demethylation at *FLC* would then not lead to gene re-silencing, because switching of an active to a

[a] Cas9 is a bacterial protein capable of recognising a specific DNA sequence when a guide RNA complementary to that sequence is expressed in the same cell. This is known as the CRISPR system, and has been adapted as a biotechnology from a bacterial immune system of the same name. CRISPR-Cas technology has been widely used over the last couple of years in applications ranging from genome-editing to regulating and labelling endogenous genes in a variety of organisms (reviewed in [359]).

repressed state is assumed to require VIN3-mediated nucleation of H3-K27me3 [83, 172].

6.2 Integration of cis and trans regulation

In Chapter 2, cis and trans epigenetic memory were contrasted, and it was shown that the decisive factor determining the repressed and active *FLC* states after cold exposure is located in cis. However, in addition to the changes in chromatin state at *FLC* after cold, it is possible that there are also changes to the concentrations of some of the trans factors that regulate *FLC*.

Despite analysis of many root samples, it was striking that files of *FLC*-repressed cells were never observed for samples grown entirely in warm conditions. It is therefore possible that cis-memory at *FLC* is only operational after cold exposure, when trans-acting factors driving *FLC* expression may have been reduced. This hypothesis was developed in a more general sense in Chapter 4, using a mathematical model of transcription as the opposing state. This model led to the concept of a 'window of cis memory'. Within a certain range of trans-acting 'activation signals', it was shown that the chromatin state can determine gene expression. However, when activation signals were extended beyond this range, the process of transcription was independent of the initial chromatin state (Sec. 4.3.1). This model showed how H3K27me at a single locus could either be maintained as a result of trans-acting gene repression, or by a cis-acting positive feedback, depending on the relative strengths of each of these two aspects.

With this in mind, it is interesting to return to the results of Chapter 2 and consider differences in trans-regulation in addition to the cis memory observed after cold. In Chapter 4, it was suggested that the window of cis memory depends on various parameters including the strength of the H3K27me3 feedback at the gene of interest. The original model of *FLC* repression during vernalisation assumed that, in principle, bistability exists at all times: before, during and after cold [83].

However, it is also possible that bistability does not exist before cold exposure, if the factors required for establishment of the cis-acting H3-K27me-based feedback are not present at *FLC*. That is, there are two possibilities: that bistability exists in principle at all times, but is not observed before cold because the strength of trans-activation is too strong; or, that there is no bistability before cold.

These possibilities are contrasted in Figure 6.1, which shows possible single-cell *FLC* expression levels as a function of trans-activation strength[b]. The bistable model corresponds to the situation that was observed after cold exposure, where both active and repressed *FLC* expression states are possible at the same level of trans-activation[c]. As shown in Fig. 6.1, before cold exposure (Col-FRI (NV)) may correspond to bistability-in-principle with a high level of trans-activation, or lack of bistability. In both cases, no silenced cells are observed.

Conversely to Col-FRI (NV), the *FRIGIDA* mutant, Col-0 likely corresponds to a low level of transactivation, as indicated in Fig. 6.1. In Col-0 plants, expression of *FLC* is low and H3K27me3 is constitutively high, even before cold. As shown in Figure 6.1, both the Col-0 and Col-FRI genotypes correspond to a level of trans-activation that is outside the cis-memory window and therefore do not distinguish the bistable-in-principle from non-bistable models.

What is needed to distinguish these possibilities is quantitative control over the level of trans-activation at *FLC*. This would allow the *FLC* activation strength to be dialed-down by controlling the concentration of a trans-regulator. If bistability exists in principle, then there should be some level of trans-activation at which cells begin to stochastically switch to a 'digital' silenced state without cold exposure. Conversely, if bistability is not possible without a cold treatment, a continuum of expression states, similar among all cells, will be observed as the trans-activation strength is reduced. Observation of *FLC*-expression at the

[b] This diagram is somewhat analogous to Figure 4.15 (p. 165), except that it represents individual cells rather than population averages.

[c] Indeed, both expression states were observed in the same nucleus (Sec. 2.3.2).

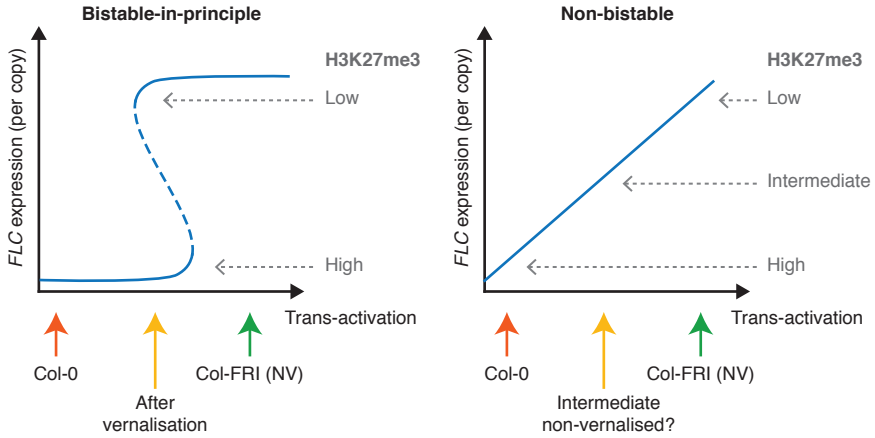


Figure 6.1: Bistable and non-bistable *FLC* chromatin. Schematic illustration showing *FLC* expression level for an individual gene copy, as a function of the strength of trans-activation. The bistable model of chromatin on the left indicates that there are two stable *FLC* chromatin/expression states in partially vernalised plants. The bistable-in-principle model on the left has stable states with high and low H3K27me3. In contrast, the non-bistable model on the right shows a continuum of states with intermediate expression and H3K27me3 levels.

level of single cells will be critical to distinguish these two possibilities. Similar to the experiments with *FLC-Venus* presented in Sec. 2.2, bistability should result in files of *FLC*-ON and *FLC*-OFF cells in Arabidopsis roots, whereas a lack of bistability should result in a similar graded reduction of expression in all cells. Observation of intermediate *FLC* expression states at the single-cell level, that are associated with an intermediate H3K27me3-level in non-vernalising conditions would be an extremely interesting result: this would indicate that H3K27me3 accumulation at *FLC* is not sufficient for inducing a stable silenced expression state.

The discussion above highlights that even after so many years of research into the mechanism of Polycomb repression, it is still unclear whether H3K27me3 acts as an epigenetic memory element, or whether cis-memory even exists at *FLC* outside of vernalisation.

This also highlights a major conceptual problem that currently exists the field of epigenetics. In the genome-wide era, it is common for all H3-K27me3-enriched genes to be grouped together as 'PRC2 target genes' and thought of as a single entity. The prevailing assumption seems to be that H3K27me3 is associated with stable epigenetic silencing. As discussed above, this is far from clear.

While genome-wide profiling is useful for determining the 'average features' of a PRC2 target gene, a deeper understanding is likely to require detailed studies of several individual cases. It is my personal feeling that PRC2 is not just a machinery for maintenance of chromatin states, but rather that different PRC2 target genes will have exploited different aspects of this elaborate apparatus to provide transcriptional control in both memory and non-memory processes. Detailed studies of several PRC2 target-genes with different memory properties would be a good starting point for understanding how memory versus non-memory functions of PRC2 are distinguished.

6.3 Combining transcription and nucleation-region models for *FLC*

In Chapter 3, experimental results that challenge the previous mathematical model of *FLC* chromatin were presented. One of these results was that candidate 'activating' histone modifications associated with the Trithorax group of proteins[d] (H3K4me3 and H3K36me3) are only present at the *FLC* nucleation region and not in the gene body (Sec. 3.2.1, p. 89). This contrasted with the prediction of the original *FLC* model, which was that an 'A-mark' at active copies of *FLC* should oppose H3K27me3 everywhere across the locus.

The second result presented to challenge the original model was that spreading of H3K27me3 from the nucleation region to the gene body after cold was extremely slow at the cell-population level (Sec. 3.2.2, 94). This could be explained in two ways: either epigenetic memory is con-

[d] Historically implicated as antagonistic to the Polycomb group of proteins [24]

tained in the nucleation region independently of the *FLC* gene body (Sec. 3.3.3, p. 104), or that spreading only occurs in replicating cells (Sec. 3.3.3, p. 114).

These two pieces of data led to the development of two alternative M-U-A models for the nucleation region only. While neither model could be excluded based on current experimental data (Sec. 3.3.4, p. 124), the second of these models —the 'two-populations model' (p. 114), was theoretically favoured.

The apparent absence of an A-mark in the *FLC* gene body motivated the development of a model in which transcription acts as an opposing state to Polycomb silencing, without a mutually-exclusive 'activating' histone modification (Chapter 4). This was shown to be capable of generating bistability.

It is now interesting to consider how these models can be brought together into a whole-gene model of *FLC*. The observation of bistability in the transcription model of Chapter 4, independently of an A-mark, begs the question as to whether the M-U-A model for the nucleation region is still required. For a general Polycomb target gene, this is unclear. For the case of *FLC*, experimental evidence indicates that H3K36me3 is indeed required for maintenance of the active state. Removal of the H3K36me3-methyltransferase SDG8 results in constitutive silencing by PRC2/H3K27me3 in non-vernalising conditions [123]. This, together with the observation that H3K36me3 and H3K27me3 rarely co-exist on the same histone [123], and *in vitro* experiments that show reduced PRC2 activity on H3K36me3-containing substrates [61, 62], suggest that H3K36me3 is a vital part of the maintenance of the active state in warm conditions.

Together, these data suggest a model in which A-marks at the nucleation region exist to prevent spurious nucleation of H3K27me3 at this position in non-vernalising conditions. If PRC2 proteins are primarily recruited at the nucleation region, rather than the gene body, then the requirement to prevent their activity in the gene body is less strong. In

fact, the low level of noisy H3K27me3-addition in the gene body could likely be prevented from accumulating using transcription-coupled H3-K27-demethylation and histone turnover, as in Chapter 4. However, H3K27me3 marks in this region must still be able to contribute to PRC2 recruitment, either directly or indirectly. If not, spreading of H3K27-me3 to the gene body in the repressed state would provide no extra stability in the model, because the gene body would simply reflect the state of the nucleation region. As seen in Sec. 3.3.3 (p. 104), storing memory in such a small region is problematic.

If the transcription-as-the-opposing-state model is to be adapted for the *FLC* gene body, then the requirement for LHP1 in spreading must also be explained. One way in which this could be done is to extend the non-processive model (Sec. 4.2) to include LHP1-K27me3 as an extra state, as shown in Figure 6.2. LHP1 binds H3K27me3 (Sec. 5.2.4) [150] and interacts with PRC2 [153], so the extra LHP1-K27me3 state in the model may also represent an extra repressive 'mark'. The dynamics of LHP1 binding are much faster than the other transitions in this models (seconds versus tens of minutes) [341]. However, based on the results presented for the non-processive model in Sec. 4.2 (Fig. 4.4), it is expected that parameter values could be found for which stability of the low expression state depends on the presence of LHP1. In the absence of LHP1, only the active (unmodified) state would be stable. *S. pombe* $HP1^{Swi6}$ dimers are proposed to form higher order complexes on nucleosomes [326], so an additional feedback representing nearest-neighbour LHP1-LHP1 interaction could also be reasonably included, which may further stabilise the repressed state. Preliminary simulations (not shown) indicate that including such an 'LHP1-me3' state can further increase the bistability in this model because a greater number of low-probability events have to occur to establish silencing. This further increases the nonlinearity in this system and makes the model less prone to noisy state changes.

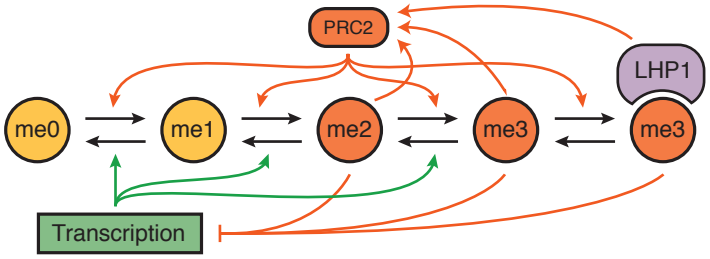


Figure 6.2: Including LHP1 in a PRC2/Transcription model. Based on the model proposed in Chapter 4 (Fig. 4.3, p. 142) with transcription as the opposing state, the above model shows how dependence on LHP1 for stability of the repressed state could be included.

Although the picture of the molecular events at *FLC* developed during this study is incomplete, it is time to 'bite the bullet' and outline a conceptual model that can explain many, but not all, of the details of *FLC* regulation in the process of vernalisation. In the following description, the 'two-populations' model (Sec. 3.3.3) will be presented in favour of the hybrid protein-histone modification model, for the theoretical reasons discussed in Sec. 3.3.4. The description is accompanied by Figure 6.3.

Updating the working model of *FLC* regulation. The active *FLC* expression state is characterised by active transcription with associated K27-demethylation and histone turnover, and nucleation region H3K36me3 to prevent spurious H3K27me3-nucleation (Fig. 6.3A). H3K36me3 is delivered in a transcription-dependent manner. Nucleation requires a PHD-PRC2 complex containing VIN3 [49, 81]. To further reduce the likelihood of spurious nucleation before cold, VIN3 levels are kept low in warm conditions. Exposure to cold leads to a slight reduction in transcription [84], reducing H3K27-demethylation, histone turnover rates, and transcription-coupled H3K36me3 addition. Induction of VIN3 expression targets a PHD-PRC2 complex to the nucleation region, driving accumulation of H3K27me3 in this region.

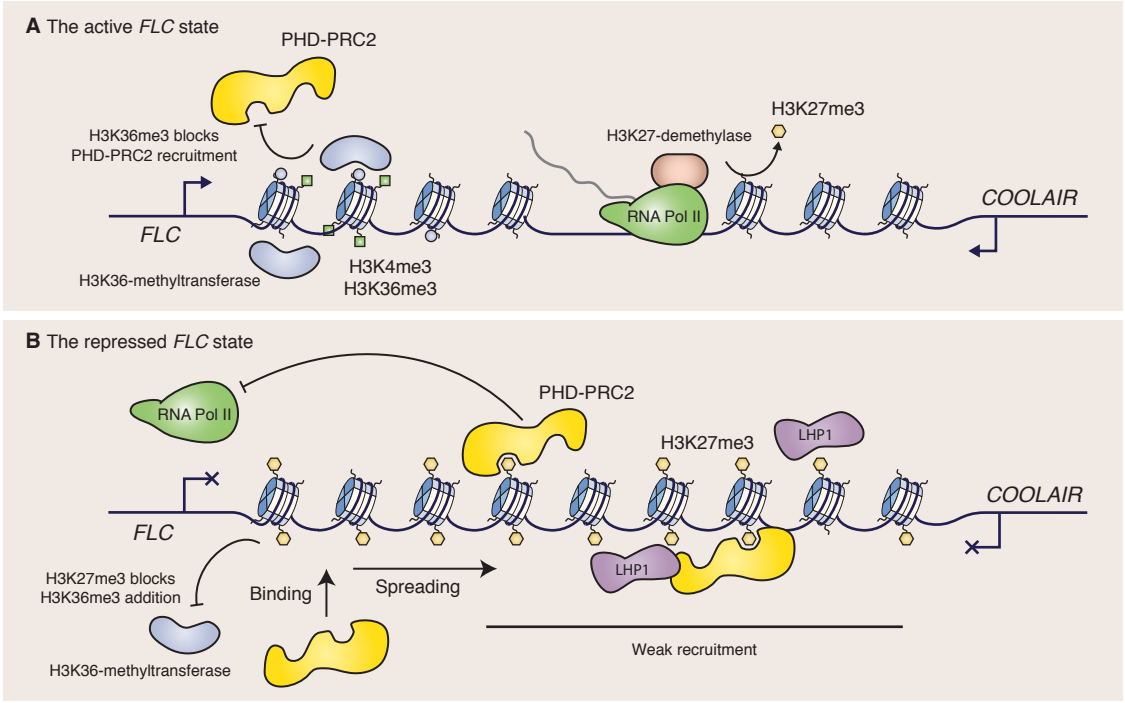


Figure 6.3: Updated working model of *FLC* through vernalisation. (A) The active state is stabilised by frequent transcription events that remove noisy H3K27me3 addition in the gene body and H3-K36me3 that is deposited at the nucleation region to prevent H3K27me3-nucleation in cis. (B) After spreading, the repressed *FLC* chromatin state is stabilised by a H3K27me3/PRC2-based feedback, dependent on LHP1. Binding of PHD-PRC2 occurs primarily through the nucleation region, though weak recruitment is possible in the gene body. The assumption of weaker recruitment in the gene body is to prevent H3K27me3-accumulation in this region without nucleation, however some recruitment, direct or indirect, is required for the gene body to stabilise the nucleation peak once spread.

During cold, the nucleation peak is sufficient to maintain a low *FLC* expression state (Fig. 1.6), and is itself maintained in trans, via high expression of VIN3, and also by cis-acting positive feedback of H3K27-me3. After cold, when DNA is replicated, PHD-PRC2 then spreads to the gene body and begins to deposit H3K27me3. H3K27me3 in the gene body is stabilised by LHP1. H3K27me3 and LHP1-H3K27me3 in the gene body can then feed back (weakly) to stabilise the nucleation region H3K27me3-peak, and maintain gene repression (Fig. 6.3B). Re-establishment of the repressed state after subsequent DNA replications is achieved using the classic H3K27me3-PRC2 based feedback [60], supplemented with an LHP1-PRC2 interaction. Failure to spread in the *lhp1* mutant results in loss of the H3K27me3-nucleation peak when DNA is replicated and leads to re-establishment of the active *FLC* state.

This model accounts for many of the experimental observations previously shown, or presented in this work. These include a genetic requirement for SDG8 [123] and transcription [121] in maintenance of the active state; for PHD-PRC2 in establishment of nucleation [45, 47–49, 80, 81]; for LHP1 in spreading (Sec. 3.2.3); and for antagonism between H3K27me3 and H3K36me3 [61, 62, 123]. This model is consistent with cis-memory after cold (Sec. 2.3) and of independent establishment of silencing at two *FLC* copies in the same cell (Sec. 2.3.2), because of the cis-determinant in stochastic establishment of H3K27me3-nucleation. The model is also consistent with H3K27me3-nucleation driving H3K27me3-spreading through cis-acting positive feedbacks, as proposed in the original *FLC* model [83]. However, recruitment within the gene body must now be significantly weaker than previously assumed as no A-marks oppose silencing in this region. This model also captures the slow time-scale of H3K27me3 spreading, because spreading is coupled to DNA replication and therefore only occurs in a subset of the population.

Perhaps the biggest unknown in the working model outlined above is how spreading is prevented in chromatin that does not replicate, or con-

versely, how spreading is coupled to replication. In the model outlined above, once transcription is repressed by a H3K27me3 nucleation peak, there is nothing to stop spreading of the H3K27me3 outwards from this region. A mathematical implementation of this model would currently require DNA replication for spreading as a hard-wired 'rule'. While several models were proposed to explain the replication-dependence of spreading (Sec. 3.3.3, p. 122), the observation that spreading indeed only occurs when DNA is replicated awaits experimental validation.

One hypothesis for the role of LHP1 RNA-binding in Polycomb silencing suggested in Chapter 5 was that RNA-binding may be required for spreading of H3K27me3 at *FLC*. If this turns out to be the case, then this observation too, must be somehow captured by the model. Figure 5.13 (p. 204) outlines conceptually how this could be included.

To summarise, this thesis has focussed on two sub-models for different aspects of *FLC* regulation. With some effort these two models could, in principle, be fused into a whole-gene model. However, the number of unknown parameters involved is likely to be large, because many interactions are now involved and the different regions of *FLC* (nucleation and gene body), seem to have different characteristics —which would require independent parameterisation. With this amount of freedom in parameter values, it is extremely likely that such a model could be made to fit the data.

In contrast to this hypothetical whole-gene mathematical model of *FLC*, the smaller sub-models developed in Chapter 3 were easily constrained by virtue of the small number of parameters and led to testable hypotheses. The minimal model in Chapter 4 was also developed in a relatively abstract manner, and generated interesting conceptual insight.

A conceptual model expressing current understanding of the chromatin dynamics at *FLC* points to gaps in understanding and will be helpful in future studies to determine how other components identified as important *FLC* regulators could integrate into the mechanism.

6.4 An outlook for epigenetics

The advent of next-generation sequencing technologies has been both a blessing and a curse for research in the field of epigenetics. On one hand, these technologies have provided unprecedented power to study the distribution of DNA methylation, histone modifications, and binding of transcription-factors or other proteins across the entire genome, in a variety of different organisms. Furthermore, genome-wide studies can rapidly provide insight into the possible functions of a protein by analysis of its binding profiles with respect to other chromatin landmarks. On the other hand, a generation of researchers in this field has now been trained in uncovering such correlations, and the skills and techniques needed to pursue the underlying mechanisms are lacking. The time to pursue a deeper understanding of this mass of descriptive data is now!

This thesis focused on trying to understand one gene in detail, with the aim of uncovering general principles that may be repeatedly employed in different biological systems. Two aspects of this study differ from the majority of others in the field of modern epigenetics: the use of techniques with single-cell resolution, and the use of mathematical modelling. Both aim to simplify the problem of determining single-cell or single-gene behaviour from measurements averaged over heterogeneous populations. This problem of sample heterogeneity is pervasive in genome-wide studies.

Single-cell sequencing technology is currently limited to RNA-seq for highly expressed genes. If this develops to enable single-cell ChIP studies, this has the potential to overcome the limitation of population-average measurements. Other microscopy-based techniques such as single-molecule FISH and *in vivo* imaging of transcription also have the potential to revolutionise studies in epigenetics.

One aspect of *FLC* regulation that has continued to prove incredibly useful for understanding the mechanism is that expression states can

be switched from active to repressed by providing a transient stimulus (i.e. cold). A key technology that could provide this ability more generally to studies of many other genes is 'epigenome-editing' (outlined in Sec. 6.1). Targeted, dynamic changes to chromatin states represent a relatively unexplored technology with the potential to greatly inform studies in epigenetics.

Another avenue that could be more deeply pursued are links between transcription and epigenetics. It is a testament to the complexity of the subject that nearly 70 years since the discovery of the Polycomb gene in Drosophila [362], the mechanism by which Polycomb complexes actually repress transcription remains unclear [363]. The field of transcription is in the midst of an upheaval as it emerges that more and more regulation is actually co-transcriptional and post-transcriptional rather than being purely regulated at the level of initiation of transcription [303, 364]. While transcriptional initiation remains a key point of control, it is clear that this is not simply controlled in the classic model of transcription factors binding to promoters and enhancers to drive assembly of pre-initiation complexes [365]. There is much to be gained by integrating cutting-edge thinking in transcription into epigenetics. As a first (baby) step towards this, Chapter 4 represents the first mathematical model in chromatin-based epigenetics to consider explicitly how transcription affects chromatin and vice versa.

The link between epigenetics and transcription was also pursued in this thesis by studying the RNA-binding of a protein associated with the repressed chromatin state, LHP1 (Chapter 5). A more complete understanding of the role of LHP1 in Polycomb silencing awaits further experiments. However, the presence of such widespread observable phenotypes (Secs. 5.3.4, 5.3.5) associated with a specific defect in RNA-binding is extremely encouraging to the belief that understanding LHP1 will help to bridge the gap between transcription and chromatin-based epigenetics.

To summarise, observation at the level of single cells and genes, together with mathematical modelling and targeted changes to chromatin will provide a powerful combination to understand the contributions of various chromatin features to gene expression output and epigenetic memory. Further integration with the field of transcription will also be pivotal in understanding how chromatin states define transcription and vice versa. Epigenetics therefore seems poised to move from descriptive to mechanistic studies.

6.5 On the use of mathematical modelling in biology

At its core, mathematics is a tool for clarity of thinking. It allows a certain set of assumptions to be made explicit, and the implications of those assumptions to be determined.

Biological systems, even those with relatively few components, can give rise to unexpected behaviour, that is not straightforward to understand intuitively. Mathematical modelling is then most useful in rigorously determining if a particular network of interactions can give rise to the observed behaviour, and to uncover the constraints that arise in doing so.

Typically, even simple mathematical models in biology are too messy to be tackled using precise analytical techniques, meaning that theoretical biologists are increasingly reliant on computer simulations. With vast computational resources, however, there is a tendency in computational biology to build ever more complex models in an attempt to capture more and more experimental detail [366, 367]. This increase in model complexity makes it difficult to determine which of the many assumptions or parameter values are actually required to generate the observed behaviour.

The most useful mathematical models are usually aggressively simple. An excellent example of this is the pioneering M-U-A model of histone-modification-based epigenetics [28] (Sec. 3.3.1). This model abstracts the thousands of possible histone modifications and histone-

modification complexes into just three nucleosome 'states'. Even the four enzymatic complexes that convert between nucleosome states are only included implicitly. Yet this model led to deep conclusions that were far from intuitive beforehand, and provided a new way of thinking about the role of histone modifications in maintaining chromatin states.

Despite the utility of this framework for conceptual understanding, however, testing the model is equally as important as formulating it. This requires comparison with experimental data, which inevitably leads to less elegant, less tractable models [83, 103, 106]. In these cases, considerable effort is required to constrain the number of free parameters, while still achieving an acceptable fit to the data.

A significant challenge when using mathematical modelling, particularly in close contact with experiments, is finding a useful level of abstraction. When confronted with quantitative data it is often difficult to distinguish the qualitative phenomena which are essential to model from the quantitative aspects of the problem, which may arise from complexities that would not be captured by even a relatively complex model. In the words of Albert Einstein,

Models should be as simple as possible, but not more so.

The guiding principle for developing models in this thesis was epigenetic memory: is this system capable of self-propagation through cell division? In Chapter 3, the small size of the *FLC* nucleation region was shown to be a major obstacle for a purely histone-modification-based memory. One proposal to overcome this was based on the notion that more epigenetic memory elements could be packed into this small region in the form of proteins. Through quantitative simulations, it was found that the success of this model depends crucially on protein-protein interactions (Sec. 3.3.3, p. 104). The other proposal to overcome the small size of this region was to assume that epigenetic memory *through DNA replication* isn't really stored in this region at all. The hy-

pothesis then developed was that much of the apparent maintenance of H3K27me3 in the nucleation region actually comes from non-dividing cells. This led to the prediction that H3K27me3 spreading only occurs in cells undergoing DNA replication (Sec. 3.3.3, p. 114).

These examples illustrate how relatively simple models can be used to generate qualitative predictions in highly-complex biological systems.

A more abstract modelling approach was taken in Chapter 4 to explain how transcription could act as an opposing state to Polycomb silencing, without relying on an activating histone modification. Development of the model was motivated by the apparent lack of an opposing mark in the *FLC* gene body, but the model was formulated in a more general sense for a generic Polycomb target gene in any organism. Following the principle of adding complexity only when required, a simple two-state model was initially tested and shown to be insufficient for generating bistability (Sec. 4.1). However, it was shown that if methylation and demethylation of H3K27 occur non-processively, then the processes of transcription-coupled histone turnover and H3K27-demethylation are sufficient for generating a bistable system. The increase in model complexity here is justified because a simpler model does not generate the qualitative phenomenon of interest: epigenetic memory.

This non-processive model is still overwhelmingly simple compared with the real biological system. Many parameter values are likely to depend on the organism, genetic background, environmental conditions and the particular oddities of the gene itself. The motivation for this model was two-fold: to analyse if transcription could act as an opposing state to PRC2-silencing, and to determine which extra behaviours arise in a model of chromatin that includes an explicit description of transcription.

It was shown that this model is able to generate transcription-dependent accumulation of histone H3.3 in rapidly cycling cells, a qualitative result reported in many organisms. By fitting quantitative mass spec-

trometry data it was also shown that H3K27me3 levels can accumulate very slowly on newly-deposited histones, taking more than a cell cycle. This slow time-scale did not present an insurmountable obstacle to cis memory. However, this observation does imply that gene repression by Polycomb is likely to be functional at genes that are not fully saturated with H3K27me3. Models with slow dynamics of H3K27-methylation were shown to be much more effective at buffering noisy transcriptional activation and repression signals, due to the sluggishness of the gene in making state changes. This suggests a biological function for such a slow time-scale. This function was not obvious before the modelling and was apparently not obvious to the authors of the study that reported the slow time-scale [26].

The mathematical models presented in this thesis are, I hope, true to their intentions: to clarify assumptions, to rigorously examine the implications of such assumptions and to propose hypotheses for experimental validation.

Mathematical modelling and experiments work in cycles. The results of experiments inform models and updated models suggest further experiments. This iterative cycle of modelling and experiments can be seen explicitly in Chapter 3, which begins with a previously developed model of *FLC* regulation, identifies key experiments that challenge the model, and then again uses modelling to propose solutions. The cycle of modelling and experiments is also present at a deeper level throughout this thesis: Chapter 2 presented experimental results that validate a core assumption of the model —that local chromatin structure underlies epigenetic memory. Then Chapter 3 proceeded with almost a complete cycle of experiments and modelling for the *FLC* system. Chapter 4 presented general modelling results for how transcription and chromatin states interact, the results of which link back to Chapter 2 and suggest a deeper understanding of the interplay between cis and trans-regulation. Finally, Chapter 5 returned to experiments to look at in-

teractions between the product of transcription, RNA, and chromatin-associated proteins in the case of LHP1.

Throughout, modelling of quantitative data and qualitative phenomena was used to pursue a deeper understanding of the mechanisms underlying how epigenetic memory can be stored in patterns of histone modifications.

6.6 Concluding remarks

The staggering complexity of life presents a fantastic puzzle to those studying its mysteries. Understanding this complexity provides us with ample opportunity to improve the quality of human life.

Regulation of gene expression underlies so much of the development and life history strategies of all organisms that continued study is vital to determine its fundamental principles. The goal of this thesis was to understand how the expression level of a single gene in the flowering plant *Arabidopsis thaliana* is maintained after winter cold. Along the way, epigenetic memory was found to be locally-encoded in a Polycomb-based system, and a requirement for RNA-binding by a conserved chromatin protein was discovered. The cis epigenetic memory system epitomised by *FLC* provided a playground for exploration of contributions of cis and trans epigenetic determinants of chromatin and gene expression.

The recent discoveries of non-coding RNA transcripts [84, 85] and interesting nucleic acid structures [173], add to the myriad of other factors known to regulate *FLC* [43]. In the face of this complexity, it is my hope that mathematical modelling will continue to strive for simplicity, and thereby help to clarify the key determinants of expression at this quintessential Polycomb target gene.

ANTIBODIES & PRIMERS

7

7.1 Antibodies

Antigen	Application	Company	Description
H3K27me3	ChIP	Merck-Millipore, 07-449	Rabbit polyclonal IgG, ChIP grade.
H3K36me3	ChIP	abcam, ab9050	Rabbit polyclonal IgG, ChIP grade.
Histone H3	ChIP	abcam, ab1791	Rabbit polyclonal IgG, ChIP grade.
GFP (Venus)	Immunoblot (1°)	Roche, 11814460001	Mixture of two monoclonal mouse antibodies (7.1 and 13.1).
GST	Immunoblot (1°)	abcam, ab92	Mouse monoclonal (3G10/1B3)
mouse IgG	Immunoblot (2°)	Santa Cruz, sc-2005	Goat IgG coupled to Horseradish Peroxidase.

7.2 Primers

Name	Sequence	Target/comments
Genotyping		
TFL2_1172_F	ctaagcggttcgagtctatt	*lhp1-3*: Use with TFL2_1411_R
TFL2_1411_R	gccattgggtcttacattat	*lhp1-3*: Use with TFL2_1172_F, amplifies a 240 bp fragment cut by AflII in *lhp1-3*: *lhp1-3* = 49 + 191 bp [223]
UJ26-FRI	agatttgctggatttgataagg	*fri*: Use with HC_FRI_R
HC_FRI_R	cttgatgttggtcgatgatg	*fri*: Use with UJ26-FRI, amplifies 167 bp on *FRI* and 151 bp on *fri*

Name	Sequence	Target/comments
pSLJ_RB3	tattcgggcctaacttttggtgtg	Targets pSLJ-755I6 [192] tDNA right border sequence. Use with FLC-V33_R or FLC-mC11_R
FLC-V33_F	acagaggatcgagtggttt	*FLC-VENUS33*: Use with FLC-V33_R
FLC-V33_R	acatcagacgaaagagagga	*FLC-VENUS33*: Use with FLC-V33_F: ampifies 253 bp on *flc-venus33*, or pSLJ_RB3: amplifies ~190bp on *FLC-VENUS33*
FLC-mC11_F	acgctatgtaaacgtgattaagt	*FLC-MCHERRY11*: Use with FLC-mC11_R
FLC-mC11_R	acctcaagatccgatacatcc	*FLC-MCHERRY11*: Use with FLC-mC11_F: ampifies 282 bp on *flc-mcherry11*, or pSLJ_RB3: amplifies ~180bp on *FLC-MCHERRY11*
LBb1.3	attttgccgatttcggaac	SALK tDNA left border primer [346]
lhp1-6_F	gtttgggaggctcgaatactc	*lhp1-6*: Use with LBb1.3 and lhp1-6_R
lhp1-6_R	tggatcaggaatgtcaggaag	*lhp1-6*: Use with LBb1.3 and lhp1-6_F
RT-qPCR		
FLC_spliced_F	agccaagaagaccgaactca	Spliced *FLC* (At5g10140)
FLC_spliced_R	tttgtccagcaggtgacatc	Spliced *FLC* (At5g10140)
FLC_unspliced_F	cgcaattttcatagcccttg	Nascent *FLC* (At5g10140)
FLC_unspliced_R	cgcaattttcatagcccttg	Nascent *FLC* (At5g10140)
UBC_qPCR_F	ctgcgactcagggaatcttctaa	*UBC* (At5g25760)
UBC_qPCR_R	ttgtgccattgaattgaaccc	*UBC* (At5g25760)
FLC-VENUS_1247_F	cacatggtcctgctggagtt	*FLC-Venus*
FLC-VENUS_1388_R	cggagatttgtccagcaggt	*FLC-Venus*
FLC-GUS_R	tccacagttttcgcgatcca	*FLC-GUS*
Set6_new_LP	tgtatgtgttcttcacttctgtcaa	Total *COOLAIR*
Set6_new_RP	gccgtaggcttcttcactgt	Total *COOLAIR*
Set2_new_LP	tcatcatgtgggagcagaag	Proximal *COOLAIR*
Set2_new_RP	tctcacacgaataaggtggcta	Proximal *COOLAIR*
Set4 RT	aatatctggcccgacgaag	Gene-specific RT only

Name	Sequence	Target/comments
Set4-new-F-195	gtatctccggcgacttgaac	Distal *COOLAIR*
Set4-new-R-195	ggatgcgtcacagagaacag	Distal *COOLAIR*
ChIP		
SDB_FLC_-2320_F	atccagaaaagggcaaggag	*FLC* -2285
SDB_FLC_-2249_R	cgaatcgattgggtgaatg	*FLC* -2285
SDB_-1204_F	caaaacttcttggcacagctc	FLC -1142
SDB_-1079_F	actcggagtgggtgaaactg	FLC -1142
FLC_-501_F	actatgtaggcacgactttggtaac	FLC -321
FLC_-381_R	tgcagaaagaacctccactctac	FLC -321
FLC_48_F	cgacaagtcaccttctccaaa	FLC 245
FLC_205_R	agggggaacaaatgaaaacc	FLC 245
SDB_FLC_416_F	ggcggatctcttgttgtttc	FLC 470
SDB_FLC_524_R	cttcttcacgacattgttcttcc	FLC 470
FLC_679_F	tcattggatctctcggatttg	FLC 868
FLC_817_R	aggtccacagcaaagataggaa	FLC 868
FLC_1424_F	ttgacaatccacaacctcaatc	FLC 1612
FLC_1561_R	tcaatttcctagaggcaccaa	FLC 1612
SDB_FLC_1993_F	agccttttagaacgtggaacc	FLC 2093
SDB_FLC_2192_R	tcttccatagaaggaagcgact	FLC 2093
SDB_FLC_3197_F	ggggctgcgtttacatttta	FLC 3275
SDB_FLC_3352_R	gtgatagcgctggctttgat	FLC 3275
FLC_3899_F	ctttttcatgggcaggatca	FLC 4103
FLC_4069_R	tgacatttgatcccacaagc	FLC 4103
SDB_FLC_4322_F	agaacaaccgtgctgctttt	FLC 4405
SDB_FLC_4488_R	tgtgtgcaagctcgttaagc	FLC 4405
FLC_5030_F	ccggttgttggacataactagg	FLC 5089
FLC_5135_R	ccaaacccagacttaaccagac	FLC 5089
FLC_5534_F	tggttgttatttggtggtgtg	FLC 5598
SDB_FLC_5666_R	atctccatctcagcttctgctc	FLC 5598
FLC_5948_F	cgtgtgagaattgcatcgag	FLC 6013
FLC_6066_R	aaaaacgcgcagagagagag	FLC 6013
FLC_6521_F	atcgtcagtgttaaaatgcactc	FLC 6558
FLC_6582_R	tgaaattgctggttagcttcg	FLC 6558
FLC_6768_F	ttgtaaagtccgatggagacg	FLC 6809
FLC_6838_R	actcggcgagaaagtttgtg	FLC 6809
SDB_FLC_7059_F	gttttggttctcccacttgc	FLC 7121
SDB_FLC_7183_R	tacggaccgcggatatattg	FLC 7121

Name	Sequence	Target/comments
SDB_STM_92_F	gcccatcatgacatcacatc	STM
SDB_STM_196_R	gggaactactttgttggtggtg	STM
ACTIN_728_F	gatattcagccacttgtctgtg	ACTIN
ACTIN_812_R	cttacacatgtacaacaaagaagg	ACTIN
tDNA copy number		
Bar-F	ggccgagtcgaccgtgta	*BAR*
Bar-R	ttgggcagcccgatga	*BAR*
Bar-Probe	FAM-cgccaccagcggacggga-TAMRA	*BAR*
AtCO-F	gtccgggtctgcgagtca	*CONSTANS* (control)
AtCO-R	gctgtgcatagagaggcatcatc	*CONSTANS* (control)
AtCO-Probe	VIC-tgctccggctgctttttgtgtgag-TAMRA	*CONSTANS* (control)

ABBREVIATIONS

Units

bp (nucleotide) base pairs

kb kilobases

kDa kiloDalton

Chemicals

CTAB Cetyltrimethylammonium bromide

DMSO Dimethylsulfoxide

EDTA Ethylenediaminetetraacetic acid

HEPES 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid

IPTG Isopropyl β-D-1-thiogalactopyranoside

LB Lysogeny broth media

MS Murashige and Skoog media

PBS Phosphate-buffered saline

PIPES Piperazine-1,2-bis[2-ethanesulfonic acid]

PPT Phosphinothricin

SDS Sodium dodecyl sulfate

Proteins and protein complexes

ACT2 ACTIN 2 (At3g18780, H3K36me3 ChIP control gene)

AG AGAMOUS (At4g18960, LHP1 and Polycomb target gene)

AP3 APETALA3 (At3g54340, LHP1 and Polycomb target gene)

ATXR7 ARABIDOPSIS TRITHORAX-RELATED 7 (At5g42400)

ATX1 ARABIDOPSIS TRITHORAX 1 (At2g31650)

CLF CURLY LEAF (At2g23380, E(z) homologue, PRC2)

E(z) (Ezh2) Enhancer of Zeste 2 (Drosophila, PRC2)

EMF2 EMBRYONIC FLOWER 2 (At5g51230, Su(z)12 homologue, PRC2)

Esc (Eed) Extra Sex Combs (Drosophila, PRC2)

FIE FERTILISATION-INDEPENDENT ENDOSPERM (At3g20740, Esc homologue, PRC2)

FIS2 FERTILIZATION INDEPENDENT SEED 2 (At2g35670, Su(z)12 homologue, PRC2)
FLC FLOWERING LOCUS C (At5g10140)
FT FLOWERING LOCUS T (At1g65480, LHP1 and Polycomb target gene / flowering activator)
GUS β-glucuronidase
HAT Histone acetyltransferase
HDAC Histone deacetylase
LHP1 LIKE HETEROCHROMATIN PROTEIN 1 (At5g17690, TERMINAL FLOWER 2 / TFL2) (HP1 family member)
MEA MEDEA (At1g02580, E(z) homologue, PRC2)
MSI1 MULTICOPY SUPRESSOR OF IRA1 (At5g58230, Nurf55 homologue, PRC2)
Nurf55 (RbAp46, RbAp48) Nucleosome remodelling factor 55 kDa subunit (Drosophila, PRC2)
Pc Polycomb (Drosophila, PRC1)
PcG Polycomb group of proteins
PRC1 Polycomb repressive complex 1
PRC2 Polycomb repressive complex 2
SDG8 SET-DOMAIN GROUP 8 / EARLY FLOWERING IN SHORT DAYS (At1g77300, EFS)
SEP3 SEPALLATA3 (At1g24260, LHP1 and Polycomb target gene)
STM SHOOT MERISTEMLESS (At1g62360, Polycomb target gene)
Su(z)12 Supressor of Zeste 12 (Drosophila)
Swi6 HP1 homologue (*S. pombe*)
SWN SWINGER (At4g02020, E(z) homologue, PRC2)
Trx Trithorax (Drosophila)
TrxG Trithorax group of proteins
UBC UBIQUITIN-CONJUGATING ENZYME 21 (At5g25760, RT-qPCR control gene)
VEL1 VRN5/VIN3-like protein (At4g30200)
VIN3 VERNALISATION INSENSITIVE 3 (At5g57380, PHD-PRC2)

VRN1 VERNALISATION 1 (At3g18990, B3-domain containing protein)

VRN2 VERNALISATION 2 (At4g16845, Su(z)12 homologue, PRC2)

VRN5 VERNALISATION 5 (At3g24440, PHD-PRC2)

WUS WUSCHEL (At2g17950, LHP1 and Polycomb target gene)

Miscellaneous

ChIP Chromatin immunoprecipitation

EMSA Electrophoretic mobility shift assay

FP Fluorescence polarisation

NLS Nuclear localisation signal

PAGE Polyacrylamide gel electrophoresis

PCR Polymerase chain reaction

SILAC Stable isotope labelling by amino acids in cell culture

TES Transcription end site

TSS Transcription start site

BIBLIOGRAPHY

[1] A. Kicheva, M. Cohen and J. Briscoe. **Developmental pattern formation: insights from physics and biology.** *Science* 338.6104 (2012), 210–212 (see p. 17).

[2] T. Tabata. **Genetics of morphogen gradients.** *Nat. Rev. Genet.* 2.8 (2001), 620–630 (see p. 17).

[3] Y. Lavin, D. Winter, R. Blecher-Gonen, E. David, H. Keren-Shaul, M. Merad, S. Jung and I. Amit. **Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment.** *Cell* 159.6 (2014), 1312–1326 (see p. 17).

[4] M. Frechin, T. Stoeger, S. Daetwyler, C. Gehin, N. Battich, E.-M. Damm, L. Stergiou, H. Riezman and L. Pelkmans. **Cell-intrinsic adaptation of lipid composition to local crowding drives social behaviour.** *Nature* 523.7558 (2015), 88–91 (see p. 17).

[5] B. W. Benham-Pyle, B. L. Pruitt and W. J. Nelson. **Mechanical strain induces E-cadherin-dependent Yap1 and β-catenin activation to drive cell cycle entry**. *Science* 348.6238 (2015), 1024–1027 (see p. 17).

[6] R. Bonasio, S. Tu and D. Reinberg. **Molecular signals of epigenetic states.** *Science* 330.6004 (2010), 612–616 (see pp. 17, 18, 20, 62, 218).

[7] S. N. Ho, S. R. Biggar, D. M. Spencer, S. L. Schreiber and G. R. Crabtree. **Dimeric ligands define a role for transcriptional activation domains in reinitiation.** *Nature* 382.6594 (1996), 822–826 (see p. 17).

[8] T. H. Cheng and M. R. Gartenberg. **Yeast heterochromatin is a dynamic structure that requires silencers continuously.** *Genes Dev.* 14.4 (2000), 452–463 (see p. 17).

[9] A. B. Oppenheim, O. Kobiler, J. Stavans, D. L. Court and S. Adhya. **Switches in bacteriophage lambda development.** *Annu. Rev. Genet.* 39 (2005), 409–429 (see p. 18).

[10] J. M. G. Vilar, C. C. Guet and S. Leibler. **Modeling network dynamics: the lac operon, a case study.** *J. Cell Biol.* 161.3 (2003), 471–476 (see p. 18).

[11] J.-W. Veening, W. K. Smits and O. P. Kuipers. **Bistability, epigenetics, and bet-hedging in bacteria.** *Annu. Rev. Microbiol.* 62.1 (2008), 193–210 (see p. 18).

[12] J. E. Ferrell. **Bistability, bifurcations, and Waddington's epigenetic landscape.** *Curr. Biol.* 22.11 (2012), R458–66 (see pp. 18, 35, 59).

[13] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent and T. J. Richmond. **Crystal structure of the nucleosome core particle at 2.8 A resolution.** *Nature* 389.6648 (1997), 251–260 (see p. 20).

[14] R. Kornberg and Y. Lorch. **Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome.** *Cell* 98.3 (1999), 285–294 (see p. 20).

[15] R. Holliday. **The inheritance of epigenetic defects.** *Science* 238.4824 (1987), 163–170 (see p. 20).

[16] D. Schübeler. **Function and information content of DNA methylation**. *Nature* 517.7534 (2015), 321–326 (see pp. 20, 33, 216).

[17] V. W. Zhou, A. Goren and B. E. Bernstein. **Charting histone modifications and the functional organization of mammalian genomes.** *Nat. Rev. Genet.* 12.1 (2011), 7–18 (see p. 20).

[18] G. J. Narlikar, R. Sundaramoorthy and T. Owen-Hughes. **Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes.** *Cell* 154.3 (2013), 490–503 (see p. 20).

[19] G. E. Zentner and S. Henikoff. **Regulation of nucleosome dynamics by histone modifications.** *Nat. Struct. Mol. Biol.* 20.3 (2013), 259–266 (see pp. 20, 33, 135, 136).

[20] A. C. Ferguson-Smith. **Genomic imprinting: the emergence of an epigenetic paradigm.** *Nat. Rev. Genet.* 12.8 (2011), 565–575 (see pp. 21, 216).

[21] S. I. S. Grewal and S. C. R. Elgin. **Heterochromatin: new possibilities for the inheritance of structure.** *Curr. Opin. Genet. Dev.* 12.2 (2002), 178–187 (see p. 21).

[22] J. Sharif, M. Muto, S.-i. Takebayashi, I. Suetake, A. Iwamatsu, T. A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura, S. Tajima, K. Mitsuya, M. Okano and H. Koseki. **The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA**. *Nature* 450.7171 (2007), 908–912 (see pp. 21, 216).

[23] D. Moazed. **Mechanisms for the inheritance of chromatin states.** *Cell* 146.4 (2011), 510–518 (see pp. 21, 85, 216).

[24] P. A. Steffen and L. Ringrose. **What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory.** *Nat. Rev. Mol. Cell Biol.* 15.5 (2014), 340–356 (see pp. 21, 25, 27, 36, 85, 89, 222).

[25] A. T. Annunziato. **Split decision: what happens to nucleosomes during DNA replication?** *J. Biol. Chem.* 280.13 (2005), 12065–12068 (see pp. 22, 34, 144, 216).

[26] C. Alabert, T. K. Barth, N. Reverón-Gómez, S. Sidoli, A. Schmidt, O. N. Jensen, A. Imhof and A. Groth. **Two distinct modes for propagation of histone PTMs across the cell cycle.** *Genes Dev.* 29.6 (2015), 585–590 (see pp. 22, 34, 131, 148–151, 170, 173, 174, 216, 234).

[27] M. Radman-Livaja, K. F. Verzijlbergen, A. Weiner, T. van Welsem, N. Friedman, O. J. Rando and F. van Leeuwen. **Patterns and Mechanisms of Ancestral Histone Protein Inheritance in Budding Yeast**. *PLoS Biol.* 9.6 (2011), e1001075 (see pp. 22, 34, 167, 216).

[28] I. B. Dodd, M. A. Micheelsen, K. Sneppen and G. Thon. **Theoretical analysis of epigenetic cell memory by nucleosome modification.** *Cell* 129.4 (2007), 813–822 (see pp. 22, 33, 35, 85, 86, 100–103, 106, 110, 128, 130, 135, 136, 139, 144, 148, 166, 167, 216, 231).

[29] M. Ptashne. **On the use of the word 'epigenetic'**. *Curr. Biol.* 17.7 (2007), R233–R236 (see pp. 22, 32, 216).

[30] P. D. Kaufman and O. J. Rando. **Chromatin as a potential carrier of heritable information.** *Curr. Opin. Cell Biol.* 22.3 (2010), 284–290 (see pp. 22, 32, 85, 216).

[31] S. Henikoff and A. Shilatifard. **Histone modification: cause or cog?** *Trends Genet.* 27.10 (2011), 389–396 (see pp. 22, 32, 216).

[32] M. Ptashne. **Epigenetics: core misconcept.** *Proc. Natl. Acad. Sci. U. S. A.* 110.18 (2013), 7101–7103 (see pp. 22, 32, 216).

[33] P. Chouard. **Vernalization and its relations to dormancy**. *Annu. Rev. Plant. Physiol.* 11.1 (1960), 191–238 (see p. 22).

[34] R. Amasino. **Vernalization, competence, and the epigenetic memory of winter.** *Plant Cell* 16.10 (2004), 2553–2559 (see p. 22).

[35] C. C. Sheldon, J. E. Burn, P. P. Perez, J. Metzger, J. A. Edwards, W. J. Peacock and E. S. Dennis. **The FLF MADS box gene: A repressor of flowering in Arabidopsis regulated by vernalization and methylation**. *Plant Cell* 11.3 (1999), 445–458 (see p. 22).

[36] S. D. Michaels and R. M. Amasino. **FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering**. *Plant Cell* 11.5 (1999), 949–956 (see pp. 22, 23, 60, 71).

[37] I. Searle, Y. He, F. Turck, C. Vincent, F. Fornara, S. Kröber, R. A. Amasino and G. Coupland. **The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in Arabidopsis.** *Genes Dev.* 20.7 (2006), 898–912 (see pp. 22, 30).

[38] C. C. Sheldon, D. T. Rouse, E. J. Finnegan, W. J. Peacock and E. S. Dennis. **The molecular basis of vernalization: the central role of FLOWERING LOCUS C (FLC).** *Proc. Natl. Acad. Sci. U. S. A.* 97.7 (2000), 3753–3758 (see p. 23).

[39] O. Leyser and S. Day. **Mechanisms in Plant Development**. John Wiley & Sons, 2009 (see pp. 23, 40–43).

[40] I. Lee, S. D. Michaels, A. S. Masshardt and R. M. Amasino. **The late-flowering phenotype of FRIGIDA and mutations in LUMINIDEPENDENS is suppressed in the Landsberg erecta strain of Arabidopsis**. *Plant J.* 6.6 (1994), 903–909 (see pp. 23, 25, 49, 50, 59, 71).

[41] U. Johanson. **Molecular Analysis of FRIGIDA, a Major Determinant of Natural Variation in Arabidopsis Flowering Time**. *Science* 290.5490 (2000), 344–347 (see p. 23).

[42] M. Koornneef, C. J. Hanhart and J. H. van der Veen. **A genetic and physiological analysis of late flowering mutants in Arabidopsis thaliana**. *Molec. Gen. Genet.* 229.1 (1991), 57–66 (see p. 25).

[43] S. Berry and C. Dean. **Environmental perception and epigenetic memory: mechanistic insight through FLC.** *Plant J.* 83.1 (2015), 133–148 (see pp. 25, 30, 216, 235).

[44] J. Chandler, A. Wilson and C. Dean. **Arabidopsis mutants showing an altered response to vernalization.** *Plant J.* 10.4 (1996), 637–644 (see p. 25).

[45] A. R. Gendall, Y. Y. Levy, A. Wilson and C. Dean. **The VERNALIZATION 2 gene mediates the epigenetic regulation of vernalization in Arabidopsis.** *Cell* 107.4 (2001), 525–535 (see pp. 25, 58, 59, 70, 227).

[46] Y. Y. Levy, S. Mesnage, J. S. Mylne, A. R. Gendall and C. Dean. **Multiple roles of Arabidopsis VRN1 in vernalization and flowering time control.** *Science* 297.5579 (2002), 243–246 (see pp. 25, 58).

[47] T. Greb, J. S. Mylne, P. Crevillen, N. Geraldo, H. An, A. R. Gendall and C. Dean. **The PHD finger protein VRN5 functions in the epigenetic silencing of Arabidopsis FLC.** *Curr. Biol.* 17.1 (2007), 73–78 (see pp. 25, 28, 58, 59, 104, 227).

[48] R. Bastow, J. S. Mylne, C. Lister, Z. Lippman, R. A. Martienssen and C. Dean. **Vernalization requires epigenetic silencing of FLC by histone methylation.** *Nature* 427.6970 (2004), 164–167 (see pp. 25, 28, 48, 70, 71, 96, 104, 227).

[49] F. De Lucia, P. Crevillen, A. M. E. Jones, T. Greb and C. Dean. **A PHD-polycomb repressive complex 2 triggers the epigenetic silencing of FLC during vernalization.** *Proc. Natl. Acad. Sci. U. S. A.* 105.44 (2008), 16831–16836 (see pp. 25, 28, 59, 70, 76, 87, 94, 225, 227).

[50] Y. B. Schwartz and V. Pirrotta. **Polycomb silencing mechanisms and the management of genomic programmes**. *Nat. Rev. Genet.* 8.1 (2007), 9–22 (see p. 25).

[51] R. Margueron and D. Reinberg. **The Polycomb complex PRC2 and its mark in life.** *Nature* 469.7330 (2011), 343–349 (see p. 25).

[52] A. Kuzmichev, K. Nishioka, H. Erdjument-Bromage, P. Tempst and D. Reinberg. **Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein.** *Genes Dev.* 16.22 (2002), 2893–2905 (see p. 25).

[53] R. Cao, L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R. S. Jones and Y. Zhang. **Role of histone H3 lysine 27 methylation in Polycomb-group silencing.** *Science* 298.5595 (2002), 1039–1043 (see pp. 25, 216).

[54] B. Czermin, R. Melfi, D. McCabe, V. Seitz, A. Imhof and V. Pirrotta. **Drosophila Enhancer of Zeste/Esc complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites**. *Cell* 111.2 (2002), 185–196 (see p. 25).

[55] J. Müller, C. M. Hart, N. J. Francis, M. L. Vargas, A. Sengupta, B. Wild, E. L. Miller, M. B. O'Connor, R. E. Kingston and J. A. Simon. **Histone methyltransferase activity of a Drosophila Polycomb Group repressor complex**. *Cell* 111.2 (2002), 197–208 (see p. 25).

[56] R. Cao and Y. Zhang. **SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex.** *Mol. Cell* 15.1 (2004), 57–67 (see p. 26).

[57] A. N. Rai, M. L. Vargas, L. Wang, E. F. Andersen, E. L. Miller and J. A. Simon. **Elements of the polycomb repressor SU(Z)12 needed for histone H3-K27 methylation, the interface with E(Z), and in vivo function.** *Mol. Cell. Biol.* 33.24 (2013), 4844–4856 (see p. 26).

[58] C. S. Ketel, E. F. Andersen, M. L. Vargas, J. Suh, S. Strome and J. A. Simon. **Subunit contributions to histone methyltransferase activities of fly and worm polycomb group complexes.** *Mol. Cell. Biol.* 25.16 (2005), 6857–6868 (see p. 26).

[59] M. Nekrasov, B. Wild and J. Müller. **Nucleosome binding and histone methyltransferase activity of Drosophila PRC2.** *EMBO Rep.* 6.4 (2005), 348–353 (see p. 26).

[60] R. Margueron, N. Justin, K. Ohno, M. L. Sharpe, J. Son, W. J. I. Drury, P. Voigt, S. R. Martin, W. R. Taylor, V. De Marco, V. Pirrotta, D. Reinberg and S. J. Gamblin. **Role of the polycomb protein EED in the propagation of repressive histone marks**. *Nature* 461.7265 (2009), 762–767 (see pp. 26, 28, 37, 86, 133, 141, 169, 216, 227).

[61] F. W. Schmitges et al. **Histone methylation by PRC2 is inhibited by active chromatin marks.** *Mol. Cell* 42.3 (2011), 330–341 (see pp. 26, 37, 93, 169, 223, 227).

[62] W. Yuan, M. Xu, C. Huang, N. Liu, S. Chen and B. Zhu. **H3K36 methylation antagonizes PRC2-mediated H3K27 methylation.** *J. Biol. Chem.* 286.10 (2011), 7983–7989 (see pp. 26, 37, 93, 169, 223, 227).

[63] J. A. Simon and R. E. Kingston. **Mechanisms of Polycomb gene silencing: knowns and unknowns**. *Nat. Rev. Mol. Cell Biol.* 10.10 (2009), 697–708 (see pp. 26, 136).

[64] A. R. Pengelly, Ö. Copur, H. Jäckle, A. Herzig and J. Müller. **A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb.** *Science* 339.6120 (2013), 698–699 (see pp. 27, 131, 216).

[65] N. P. Blackledge, N. R. Rose and R. J. Klose. **Targeting Polycomb systems to regulate gene expression: modifications to a complex story.** *Nat. Rev. Mol. Cell Biol.* 16.11 (2015), 643–649 (see p. 27).

[66] G. Buchwald, P. van der Stoop, O. Weichenrieder, A. Perrakis, M. van Lohuizen and T. K. Sixma. **Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b.** *EMBO J.* 25.11 (2006), 2465–2474 (see p. 27).

[67] M. de Napoles, J. E. Mermoud, R. Wakao, Y. A. Tang, M. Endoh, R. Appanah, T. B. Nesterova, J. Silva, A. P. Otte, M. Vidal, H. Koseki and N. Brockdorff. **Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to**

heritable gene silencing and X inactivation. *Dev. Cell* 7.5 (2004), 663–676 (see p. 27).

[68] N. J. Francis, R. E. Kingston and C. L. Woodcock. **Chromatin compaction by a polycomb group protein complex.** *Science* 306.5701 (2004), 1574–1577 (see p. 27).

[69] R. Eskeland, M. Leeb, G. R. Grimes, C. Kress, S. Boyle, D. Sproul, N. Gilbert, Y. Fan, A. I. Skoultchi, A. Wutz and W. A. Bickmore. **Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination.** *Mol. Cell* 38.3 (2010), 452–464 (see pp. 27, 136).

[70] R. S. Illingworth, M. Moffat, A. R. Mann, D. Read, C. J. Hunter, M. M. Pradeepa, I. R. Adams and W. A. Bickmore. **The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development.** *Genes Dev.* 29.18 (2015), 1897–1902 (see p. 27).

[71] J. Min, Y. Zhang and R.-M. Xu. **Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27.** *Genes Dev.* 17.15 (2003), 1823–1828 (see p. 27).

[72] L. Wang, J. L. Brown, R. Cao, Y. Zhang, J. A. Kassis and R. S. Jones. **Hierarchical recruitment of polycomb group silencing complexes.** *Mol. Cell* 14.5 (2004), 637–646 (see p. 27).

[73] N. P. Blackledge, A. M. Farcas, T. Kondo, H. W. King, J. F. McGouran, L. L. P. Hanssen, S. Ito, S. Cooper, K. Kondo, Y. Koseki, T. Ishikura, H. K. Long, T. W. Sheahan, N. Brockdorff, B. M. Kessler, H. Koseki and R. J. Klose. **Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation.** *Cell* 157.6 (2014), 1445–1459 (see p. 27).

[74] S. J. Whitcomb, A. Basu, C. D. Allis and E. Bernstein. **Polycomb Group proteins: an evolutionary perspective.** *Trends Genet.* 23.10 (2007), 494–502 (see p. 27).

[75] K. Zhang, V. V. Sridhar, J. Zhu, A. Kapoor and J.-K. Zhu. **Distinctive core histone post-translational modification patterns in Arabidopsis thaliana.** *PLoS ONE* 2.11 (2007), e1210 (see p. 27).

[76] L. Sanchez-Pulido, D. Devos, Z. R. Sung and M. Calonje. **RAWUL: a new ubiquitin-like domain in PRC1 ring finger proteins that unveils putative plant and worm PRC1 orthologs.** *BMC Genomics* 9 (2008), 308 (see p. 27).

[77] L. Xu and W.-H. Shen. **Polycomb silencing of KNOX genes confines shoot stem cell niches in Arabidopsis.** *Curr. Biol.* 18.24 (2008), 1966–1971 (see p. 27).

[78] F. Bratzel, G. López-Torrejón, M. Koch, J. C. Del Pozo and M. Calonje. **Keeping cell identity in Arabidopsis requires PRC1 RING-finger homologs that catalyze H2A monoubiquitination.** *Curr. Biol.* 20.20 (2010), 1853–1859 (see p. 27).

[79] D. Chen, A. Molitor, C. Liu and W.-H. Shen. **The Arabidopsis PRC1-like ring-finger proteins are necessary for repression of embryonic traits during vegetative growth.** *Cell Res.* 20.12 (2010), 1332–1344 (see p. 27).

[80] S. Sung, R. J. Schmitz and R. M. Amasino. **A PHD finger protein involved in both the vernalization and photoperiod pathways in Arabidopsis.** *Genes Dev.* 20.23 (2006), 3244–3248 (see pp. 28, 59, 87, 227).

[81] S. Sung and R. M. Amasino. **Vernalization in Arabidopsis thaliana is mediated by the PHD finger protein VIN3.** *Nature* 427.6970 (2004), 159–164 (see pp. 28, 59, 87, 94–96, 104, 225, 227).

[82] E. J. Finnegan and E. S. Dennis. **Vernalization-induced trimethylation of histone H3 lysine 27 at FLC is not maintained in mitotically quiescent cells.** *Curr. Biol.* 17.22 (2007), 1978–1983 (see pp. 28, 89, 122).

[83] A. Angel, J. Song, C. Dean and M. Howard. **A Polycomb-based switch underlying quantitative epigenetic memory.** *Nature* 476.7358 (2011), 105–108 (see pp. 28, 30, 35, 54, 78, 85, 86, 88, 89, 94, 110, 117, 135, 148, 169, 219, 227, 232).

[84] S. Swiezewski, F. Q. Liu, A. Magusin and C. Dean. **Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target.** *Nature* 462.7274 (2009), 799–802 (see pp. 28, 37, 48, 89, 225, 235).

[85] T. Csorba, J. I. Questa, Q. Sun and C. Dean. **Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization.** *Proc. Natl. Acad. Sci. U. S. A.* 111.45 (2014), 16160–16165 (see pp. 28, 50, 60, 78, 89, 235).

[86] P. A. Wigge. **FT, a mobile developmental signal in plants.** *Curr. Biol.* 21.9 (2011), R374–8 (see p. 31).

[87] S. L. Berger, T. Kouzarides, R. Shiekhattar and A. Shilatifard. **An operational definition of epigenetics.** *Genes Dev.* 23.7 (2009), 781–783 (see p. 32).

[88] M. Ptashne. **A genetic switch**. Cold Spring Harbor, N.Y.: CSHL Press, 2004 (see pp. 33, 59).

[89] T. S. Gardner, C. R. Cantor and J. J. Collins. **Construction of a genetic toggle switch in Escherichia coli.** *Nature* 403.6767 (2000), 339–342 (see p. 33).

[90] S. I. Grewal and A. J. Klar. **Chromosomal inheritance of epigenetic states in fission yeast during mitosis and meiosis.** *Cell* 86.1 (1996), 95–101 (see p. 33).

[91] J. L. Cherry and F. R. Adler. **How to make a biological switch.** *J. Theor. Biol.* 203.2 (2000), 117–133 (see p. 35).

[92] G. Craciun, Y. Tang and M. Feinberg. **Understanding bistability in complex enzyme-driven reaction networks.** *Proc. Natl. Acad. Sci. U. S. A.* 103.23 (2006), 8697–8702 (see p. 35).

[93] T. Wilhelm. **The smallest chemical reaction system with bistability**. *BMC Syst. Biol.* 3.1 (2009), 90–9 (see p. 35).

[94] M. Sedighi and A. M. Sengupta. **Epigenetic chromatin silencing: bistability and front propagation.** *Phys. Biol.* 4.4 (2007), 246–255 (see pp. 35, 93, 135, 136, 139).

[95] D. David-Rus, S. Mukhopadhyay, J. L. Lebowitz and A. M. Sengupta. **Inheritance of epigenetic chromatin silencing.** *J. Theor. Biol.* 258.1 (2009), 112–120 (see pp. 35, 93, 128, 139).

[96] M. A. Micheelsen, N. Mitarai, K. Sneppen and I. B. Dodd. **Theory for the stability and regulation of epigenetic landscapes.** *Phys. Biol.* 7.2 (2010), 026010 (see pp. 35, 93, 128, 139).

[97] S. Mukhopadhyay, V. H. Nagaraj and A. M. Sengupta. **Locus dependence in epigenetic chromatin silencing**. *Biosystems* 102.1 (2010), 49–54 (see pp. 35, 93, 135).

[98] I. B. Dodd and K. Sneppen. **Barriers and silencers: a theoretical toolkit for control and containment of nucleosome-based epigenetic states.** *J. Mol. Biol.* 414.4 (2011), 624–637 (see pp. 35, 93, 128, 135, 139, 148, 167, 168).

[99] A. K. Alsing and K. Sneppen. **Differentiation of developing olfactory neurons analysed in terms of coupled epigenetic landscapes.** *Nucl. Acids Res.* 41.9 (2013), 4755–4764 (see pp. 35, 128, 135, 148).

[100] D. Jost. **Bifurcation in epigenetics: Implications in development, proliferation, and diseases**. *Phys. Rev. E* 89.1 (2014), 010701 (see pp. 35, 135).

[101] A. Dayarian and A. M. Sengupta. **Titration and hysteresis in epigenetic chromatin silencing**. *Phys. Biol.* 10.3 (2013), 036005 (see pp. 35, 135).

[102] S. Mukhopadhyay and A. M. Sengupta. **The role of multiple marks in epigenetic silencing and the emergence of a stable bivalent chromatin state.** *PLoS Comput. Biol.* 9.7 (2013), e1003121 (see pp. 35, 129, 135).

[103] K. Sneppen and I. B. Dodd. **Cooperative stabilization of the SIR complex provides robust epigenetic memory in a model of SIR silencing in Saccharomyces cerevisiae**. *Epigenetics* 10.4 (2015), 293–302 (see pp. 35, 129, 135, 139, 148, 232).

[104] K. Muller-Ott, F. Erdel, A. Matveeva, J. P. Mallm, A. Rademacher, M. Hahn, C. Bauer, Q. Zhang, S. Kaltofen, G. Schotta, T. Hofer and K. Rippe. **Specificity, propagation, and memory of pericentric heterochromatin**. *Mol. Syst. Biol.* 10.8 (2014), 746–746 (see pp. 35, 129, 182, 203).

[105] A. Satake and Y. Iwasa. **A stochastic model of chromatin modification: cell population coding of winter memory in plants.** *J. Theor. Biol.* 302 (2012), 6–17 (see p. 35).

[106] A. Angel, J. Song, H. Yang, J. I. Questa, C. Dean and M. Howard. **Vernalizing cold is registered digitally at FLC.** *Proc. Natl. Acad. Sci. U. S. A.* 112.13 (2015), 4146–4151 (see pp. 35, 37, 135, 232).

[107] K. H. Hansen, A. P. Bracken, D. Pasini, N. Dietrich, S. S. Gehani, A. Monrad, J. Rappsilber, M. Lerdrup and K. Helin. **A model for transmission of the H3K27me3 epigenetic mark.** *Nat. Cell Biol.* 10.11 (2008), 1291–1300 (see pp. 37, 131, 169, 216).

[108] G. Struhl and M. Akam. **Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of Drosophila.** *EMBO J.* 4.12 (1985), 3259–3264 (see pp. 37, 131).

[109] K. Harding and M. Levine. **Gap genes define the limits of antennapedia and bithorax gene expression during early development in Drosophila.** *EMBO J.* 7.1 (1988), 205–214 (see pp. 37, 131).

[110] R. K. Maeda and F. Karch. **The ABC of the BX-C: the bithorax complex explained.** *Development* 133.8 (2006), 1413–1422 (see pp. 37, 131).

[111] E. M. Riising, I. Comet, B. Leblanc, X. Wu, J. V. Johansen and K. Helin. **Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide.** *Mol. Cell* 55.3 (2014), 347–360 (see pp. 37, 131, 162, 163).

[112] M. G. Lee, R. Villa, P. Trojer, J. Norman, K.-P. Yan, D. Reinberg, L. Di Croce and R. Shiekhattar. **Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination.** *Science* 318.5849 (2007), 447–450 (see pp. 37, 134).

[113] S. Chen, J. Ma, F. Wu, L.-J. Xiong, H. Ma, W. Xu, R. Lv, X. Li, J. Villén, S. P. Gygi, X. S. Liu and Y. Shi. **The histone H3 Lys 27 demethylase JMJD3 regulates gene expression by impacting transcriptional elongation.** *Genes Dev.* 26.12 (2012), 1364–1375 (see pp. 37, 134).

[114] E. R. Smith, M. G. Lee, B. Winter, N. M. Droz, J. C. Eissenberg, R. Shiekhattar and A. Shilatifard. **Drosophila UTX is a histone H3 Lys27 demethylase that colocalizes with the elongating form of RNA polymerase II.** *Mol. Cell. Biol.* 28.3 (2008), 1041–1046 (see pp. 37, 134).

[115] M. F. Dion, T. Kaplan, M. Kim, S. Buratowski, N. Friedman and O. J. Rando. **Dynamics of replication-independent histone turnover in budding yeast.** *Science* 315.5817 (2007), 1405–1408 (see pp. 37, 134, 143).

[116] A. Rufiange, P.-É. Jacques, W. Bhat, F. Robert and A. Nourani. **Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1.** *Mol. Cell* 27.3 (2007), 393–405 (see pp. 37, 134, 143).

[117] A. Jamai, R. M. Imoberdorf and M. Strubin. **Continuous Histone H2B and Transcription-Dependent Histone H3 Exchange in Yeast Cells outside of Replication.** *Mol. Cell* 25.3 (2007), 345–355 (see pp. 37, 134, 143).

[118] R. B. Deal, J. G. Henikoff and S. Henikoff. **Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones.** *Science* 328.5982 (2010), 1161–1164 (see pp. 37, 134, 144).

[119] M. R. Doyle and R. M. Amasino. **A single amino acid change in the enhancer of zeste ortholog CURLY LEAF results in vernalization-independent, rapid flowering in Arabidopsis.** *Plant Physiol.* 151.3 (2009), 1688–1697 (see p. 37).

[120] X. Yu and S. D. Michaels. **The Arabidopsis Paf1c complex component CDC73 participates in the modification of FLOWERING LOCUS C chromatin.** *Plant Physiol.* 153.3 (2010), 1074–1084 (see p. 37).

[121] D. M. Buzas, M. Robertson, E. J. Finnegan and C. A. Helliwell. **Transcription-dependence of histone H3 lysine 27 trimethylation at the**

Arabidopsis polycomb target gene FLC. *Plant J.* 65.6 (2011), 872–881 (see pp. 37, 60, 131, 162, 227).

[122] V. Coustham, P. Li, A. Strange, C. Lister, J. Song and C. Dean. **Quantitative modulation of polycomb silencing underlies natural variation in vernalization.** *Science* 337.6094 (2012), 584–587 (see pp. 37, 47, 50, 63).

[123] H. Yang, M. Howard and C. Dean. **Antagonistic Roles for H3K36me3 and H3K27me3 in the Cold-Induced Epigenetic Switch at Arabidopsis FLC.** *Curr. Biol.* 24.15 (2014), 1793–1797 (see pp. 37, 38, 60, 78, 85, 88–90, 93, 99, 117, 169, 223, 227).

[124] K. Yamada et al. **Empirical Analysis of Transcriptional Activity in the Arabidopsis Genome**. *Science* 302.5646 (2003), 842–846 (see p. 38).

[125] P. Bertone, V. Stolc, T. E. Royce, J. S. Rozowsky and A. E. Urban. **Global identification of human transcribed sequences with genome tiling arrays**. *Science* (2004) (see p. 38).

[126] P. Carninci et al. **The transcriptional landscape of the mammalian genome.** *Science* 309.5740 (2005), 1559–1563 (see p. 38).

[127] M. Guttman, I. Amit, M. Garber, C. French, M. F. Lin, D. Feldser, M. Huarte, O. Zuk, B. W. Carey, J. P. Cassady, M. N. Cabili, R. Jaenisch, T. S. Mikkelsen, T. Jacks, N. Hacohen, B. E. Bernstein, M. Kellis, A. Regev, J. L. Rinn and E. S. Lander. **Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals.** *Nature* 458.7235 (2009), 223–227 (see pp. 38, 39).

[128] K. Struhl. **Transcriptional noise and the fidelity of initiation by RNA polymerase II.** *Nat. Struct. Mol. Biol.* 14.2 (2007), 103–105 (see p. 38).

[129] V. Pelechano, W. Wei and L. M. Steinmetz. **Extensive transcriptional heterogeneity revealed by isoform profiling.** *Nature* 497.7447 (2013), 127–131 (see p. 38).

[130] M. Tisseur, M. Kwapisz and A. Morillon. **Pervasive transcription - Lessons from yeast**. *Biochimie* 93.11 (2011), 1889–1896 (see p. 38).

[131] T. R. Cech and J. A. Steitz. **The Noncoding RNA Revolution— Trashing Old Rules to Forge New Ones**. *Cell* 157.1 (2014), 77–94 (see p. 38).

[132] D. Mancini-Dinardo, S. J. S. Steele, J. M. Levorse, R. S. Ingram and S. M. Tilghman. **Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes.** *Genes Dev.* 20.10 (2006), 1268–1282 (see p. 38).

[133] F. Lai, U. A. Orom, M. Cesaroni, M. Beringer, D. J. Taatjes, G. A. Blobel and R. Shiekhattar. **Activating RNAs associate with Mediator to enhance chromatin architecture and transcription.** *Nature* 494.7438 (2013), 497–501 (see p. 38).

[134] J. L. Rinn, M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H. Goodnough, J. A. Helms, P. J. Farnham, E. Segal and H. Y. Chang. **Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.** *Cell* 129.7 (2007), 1311–1323 (see p. 38).

[135] M.-C. Tsai, O. Manor, Y. Wan, N. Mosammaparast, J. K. Wang, F. Lan, Y. Shi, E. Segal and H. Y. Chang. **Long noncoding RNA as modular scaffold of histone modification complexes.** *Science* 329.5992 (2010), 689–693 (see p. 38).

[136] C. Chu, Q. C. Zhang, S. T. da Rocha, R. A. Flynn, M. Bharadwaj, J. M. Calabrese, T. Magnuson, E. Heard and H. Y. Chang. **Systematic discovery of Xist RNA binding proteins.** *Cell* 161.2 (2015), 404–416 (see p. 38).

[137] C. A. McHugh, C.-K. Chen, A. Chow, C. F. Surka, C. Tran, P. McDonel, A. Pandya-Jones, M. Blanco, C. Burghard, A. Moradian, M. J. Sweredoski, A. A. Shishkin, J. Su, E. S. Lander, S. Hess, K. Plath and M. Guttman. **The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3.** *Nature* 521.7551 (2015), 232–236 (see p. 38).

[138] C. Keller, R. Kulasegaran-Shylini, Y. Shimada, H.-R. Hotz and M. Bühler. **Noncoding RNAs prevent spreading of a repressive histone mark**. *Nat. Struct. Mol. Biol.* 20.8 (2013), 994–1000 (see pp. 38, 179).

[139] V. A. Herzog, A. Lempradl, J. Trupke, H. Okulski, C. Altmutter, F. Ruge, B. Boidol, S. Kubicek, G. Schmauss, K. Aumayr, M. Ruf, A. Pospisilik, A. Dimond, H. B. Senergin, M. L. Vargas, J. A. Simon and L. Ringrose. **A strand-specific switch in noncoding transcription switches the function of a Polycomb/Trithorax response element.** *Nat. Genet.* 46.9 (2014), 973–981 (see pp. 38, 39, 131, 177, 179).

[140] R. A. Gupta, N. Shah, K. C. Wang, J. Kim, H. M. Horlings, D. J. Wong, M.-C. Tsai, T. Hung, P. Argani, J. L. Rinn, Y. Wang, P. Brzoska, B. Kong, R. Li, R. B. West, M. J. van de Vijver, S. Sukumar and H. Y. Chang. **Long non-coding RNA**

HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464.7291 (2010), 1071–1076 (see p. 38).

[141] P. J. Batista and H. Y. Chang. **Long noncoding RNAs: cellular address codes in development and disease.** *Cell* 152.6 (2013), 1298–1307 (see pp. 39, 177).

[142] S. Kaneko, R. Bonasio, R. Saldaña-Meyer, T. Yoshida, J. Son, K. Nishino, A. Umezawa and D. Reinberg. **Interactions between JARID2 and Noncoding RNAs Regulate PRC2 Recruitment to Chromatin**. *Mol. Cell* (2013), 1–11 (see p. 39).

[143] S. Kaneko, J. Son, S. S. Shen, D. Reinberg and R. Bonasio. **PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells.** *Nat. Struct. Mol. Biol.* 20.11 (2013), 1258–1264 (see pp. 39, 177, 179).

[144] S. Kaneko, J. Son, R. Bonasio, S. S. Shen and D. Reinberg. **Nascent RNA interaction keeps PRC2 activity poised and in check.** *Genes Dev.* 28.18 (2014), 1983–1988 (see pp. 39, 177, 179).

[145] C. Davidovich, L. Zheng, K. J. Goodrich and T. R. Cech. **Promiscuous RNA binding by Polycomb repressive complex 2**. *Nat. Struct. Mol. Biol.* 20.11 (2013), 1250–1257 (see pp. 39, 179).

[146] C. Muchardt, M. Guilleme, J.-S. Seeler, D. Trouche, A. Dejean and M. Yaniv. **Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1alpha.** *EMBO Rep.* 3.10 (2002), 975–981 (see pp. 39, 40, 177, 180, 181).

[147] C. Keller, R. Adaixo, R. Stunnenberg, K. J. Woolcock, S. Hiller and M. Bühler. **HP1(Swi6) Mediates the Recognition and Destruction of Heterochromatic RNA Transcripts.** *Mol. Cell* 47.2 (2012), 215–227 (see pp. 39, 40, 177–181, 184, 186, 202).

[148] J. S. Mylne, L. Barrett, F. Tessadori, S. Mesnage, L. Johnson, Y. V. Bernatavichute, S. E. Jacobsen, P. Fransz and C. Dean. **LHP1, the Arabidopsis homologue of HETEROCHROMATIN PROTEIN1, is required for epigenetic silencing of FLC**. *Proc. Natl. Acad. Sci. U. S. A.* 103.13 (2006), 5012–5017 (see pp. 39, 96, 177).

[149] M. Libault, F. Tessadori, S. Germann, B. Snijder, P. Fransz and V. Gaudin. **The Arabidopsis LHP1 protein is a component of euchromatin**. *Planta* 222.5 (2005), 910–925 (see pp. 39, 179, 191).

[150] X. Zhang, S. Germann, B. J. Blus, S. Khorasanizadeh, V. Gaudin and S. E. Jacobsen. **The Arabidopsis LHP1 protein colocalizes with histone H3 Lys27 trimethylation.** *Nat. Struct. Mol. Biol.* 14.9 (2007), 869–871 (see pp. 39, 177, 179, 180, 188, 194, 202, 224).

[151] F. Turck, F. Roudier, S. Farrona, M.-L. Martin-Magniette, E. Guillaume, N. Buisine, S. Gagnot, R. A. Martienssen, G. Coupland and V. Colot. **Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27.** *PLoS Genet.* 3.6 (2007), e86 (see pp. 39, 97, 177, 179, 180, 194).

[152] C. Maison, D. Bailly, A. H. F. M. Peters, J.-P. Quivy, D. Roche, A. Taddei, M. Lachner, T. Jenuwein and G. Almouzni. **Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component.** *Nat. Genet.* 30.3 (2002), 329–334 (see p. 40).

[153] M. Derkacheva, Y. Steinbach, T. Wildhaber, I. Mozgová, W. Mahrez, P. Nanni, S. Bischof, W. Gruissem and L. Hennig. **Arabidopsis MSI1 connects LHP1 to PRC2 complexes.** *EMBO J.* 32.14 (2013), 2073–2085 (see pp. 40, 181, 203, 224).

[154] R. J. Abbott and M. F. Gomes. **Population genetic structure and outcrossing rate of Arabidopsis thaliana (L.) Heynh.** *Heredity* 62.3 (1989), 411–418 (see p. 40).

[155] A. H. K. Roeder and M. F. Yanofsky. **Fruit Development in Arabidopsis**. *The Arabidopsis Book* 4 (2006), e0075 (see p. 40).

[156] W. Nagl. **DNA endoreduplication and polyteny understood as evolutionary strategies**. *Nature* 261.5561 (1976), 614–615 (see p. 41).

[157] J. Joubès and C. Chevalier. **Endoreduplication in higher plants.** *Plant Mol. Biol.* 43.5-6 (2000), 735–745 (see p. 41).

[158] E. Gendreau, H. Höfte, O. Grandjean, S. Brown and J. Traas. **Phytochrome controls the number of endoreduplication cycles in the Arabidopsis thaliana hypocotyl.** *Plant J.* 13.2 (1998), 221–230 (see p. 41).

[159] J. E. Melaragno, B. Mehrotra and A. W. Coleman. **Relationship between endopolyploidy and cell size in epidermal tissue of Arabidopsis.** *Plant Cell* 5.11 (1993), 1661–1668 (see p. 41).

[160] J. J. Petricka, C. M. Winter and P. N. Benfey. **Control of Arabidopsis root development.** *Annu. Rev. Plant Biol.* 63 (2012), 563–590 (see pp. 41–43).

[161] G. T. S. Beemster, L. De Veylder, S. Vercruysse, G. West, D. Rombaut, P. Van Hummelen, A. Galichet, W. Gruissem, D. Inzé and M. Vuylsteke. **Genome-wide analysis of gene expression profiles associated with cell cycle transitions in growing organs of Arabidopsis.** *Plant Physiol.* 138.2 (2005), 734–743 (see pp. 42, 117).

[162] J. I. Medford. **Vegetative Apical Meristems.** *Plant Cell* 4.9 (1992), 1029–1039 (see p. 42).

[163] A. Jacqmard, L. De Veylder, G. Segers, J. de Almeida Engler, G. Bernier, M. Van Montagu and D. Inzé. **Expression of CKS1At in Arabidopsis thaliana indicates a role for the protein in both the mitotic and the endoreduplication cycle.** *Planta* 207.4 (1999), 496–504 (see p. 42).

[164] L. Dolan, K. Janmaat, V. Willemsen, P. Linstead, S. Poethig, K. Roberts and B. Scheres. **Cellular organisation of the Arabidopsis thaliana root.** *Development* 119.1 (1993), 71–84 (see pp. 42, 55).

[165] G. T. Beemster and T. I. Baskin. **Analysis of cell division and elongation underlying the developmental acceleration of root growth in Arabidopsis thaliana.** *Plant Physiol.* 116.4 (1998), 1515–1526 (see p. 43).

[166] K. Hayashi, J. Hasegawa and S. Matsunaga. **The boundary of the meristematic and elongation zones in roots: endoreduplication precedes rapid cell expansion.** *Sci. Rep.* 3 (2013), 2723 (see pp. 43, 118).

[167] L. Dolan, C. M. Duckett, C. Grierson, P. Linstead, K. Schneider, E. Lawson, C. Dean, S. Poethig and K. Roberts. **Clonal relationships and cell patterning in the root epidermis of *Arabidopsis*.** *Development* 120 (1994), 2465–2474 (see p. 43).

[168] S. J. Clough and A. F. Bent. **Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana.** *Plant J.* 16.6 (1998), 735–743 (see pp. 44, 73).

[169] B. M. Turner. **Reading signals on the nucleosome with a new nomenclature for modified histones.** *Nat. Struct. Mol. Biol.* 12.2 (2005), 110–112 (see p. 44).

[170] P. B. Talbert et al. **A unified phylogeny-based nomenclature for histone variants.** *Epigenetics & Chromatin* 5.1 (2012), 7 (see p. 44).

[171] C. C. Sheldon, A. B. Conn, E. S. Dennis and W. J. Peacock. **Different regulatory regions are required for the vernalization-induced repression of FLOWERING LOCUS C and for the epigenetic maintenance of repression.** *Plant Cell* 14.10 (2002), 2527–2537 (see p. 47).

[172] S. Sung, Y. He, T. W. Eshoo, Y. Tamada, L. Johnson, K. Nakahigashi, K. Goto, S. E. Jacobsen and R. M. Amasino. **Epigenetic maintenance of the vernalized state in Arabidopsis thaliana requires LIKE HETEROCHROMATIN PROTEIN 1.** *Nat. Genet.* 38.6 (2006), 706–710 (see pp. 47, 96, 219).

[173] Q. Sun, T. Csorba, K. Skourti-Stathaki, N. J. Proudfoot and C. Dean. **R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus.** *Science* 340.6132 (2013), 619–621 (see pp. 48, 235).

[174] J. S. Mylne, T. Greb, C. Lister and C. Dean. **Epigenetic regulation in the control of flowering.** *Cold Spring Harb. Symp. Quant. Biol.* 69 (2004), 457–464 (see p. 48).

[175] T. Nagai, K. Ibata, E. Park, M. Kubota, K. Mikoshiba and A. Miyawaki. **A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications**. *Nat. Biotechnol.* 20.1 (2002), 87–90 (see pp. 49, 72).

[176] P. Li, Z. Tao and C. Dean. **Phenotypic evolution through variation in splicing of the noncoding RNA COOLAIR.** *Genes Dev.* 29.7 (2015), 696–701 (see p. 50).

[177] M. Koornneef, H. Vries, C. Hanhart, W. Soppe and T. Peeters. **The phenotype of some late-flowering mutants is enhanced by a locus on chromosome 5 that is not effective in the Landsberg erecta wild-type**. *Plant J.* 6.6 (1994), 911–919 (see pp. 50, 59).

[178] D. Li, C. Liu, L. Shen, Y. Wu, H. Chen, M. Robertson, C. A. Helliwell, T. Ito, E. Meyerowitz and H. Yu. **A repressor complex governs the integration of flowering signals in Arabidopsis.** *Dev. Cell* 15.1 (2008), 110–120 (see pp. 52, 54).

[179] A. Campilho, B. Garcia, H. v. d. Toorn, H. v. Wijk, A. Campilho and B. Scheres. **Time-lapse analysis of stem-cell divisions in the Arabidopsis thaliana root meristem**. *Plant J.* 48.4 (2006), 619–627 (see p. 55).

[180] S. D. Michaels and R. M. Amasino. **Loss of FLOWERING LOCUS C activity eliminates the late-flowering phenotype of FRIGIDA and autonomous pathway mutations but not responsiveness to vernalization.** *Plant Cell* 13.4 (2001), 935–941 (see p. 59).

[181] G. G. Simpson, P. P. Dijkwel, V. Quesada, I. R. Henderson and C. Dean. **FY Is an RNA 3′ End-Processing Factor that Interacts with FCA to Control the Arabidopsis Floral Transition**. *Cell* 113.6 (2003), 777–787 (see p. 59).

[182] Y. He, S. D. Michaels and R. M. Amasino. **Regulation of flowering time by histone acetylation in Arabidopsis.** *Science* 302.5651 (2003), 1751–1754 (see p. 59).

[183] I. Ausín, C. Alonso-Blanco, J. A. Jarillo, L. Ruiz-García and J. M. Martínez-Zapater. **Regulation of flowering time by FVE, a retinoblastoma-associated protein.** *Nat. Genet.* 36.2 (2004), 162–166 (see p. 59).

[184] F. Liu, S. Marquardt, C. Lister, S. Swiezewski and C. Dean. **Targeted 3' processing of antisense transcripts triggers Arabidopsis FLC chromatin silencing.** *Science* 327.5961 (2010), 94–97 (see pp. 59, 60).

[185] F. Liu, V. Quesada, P. Crevillen, I. Bäurle, S. Swiezewski and C. Dean. **The Arabidopsis RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate FLC.** *Mol. Cell* 28.3 (2007), 398–407 (see p. 60).

[186] C. A. Helliwell, M. Robertson, E. J. Finnegan, D. M. Buzas and E. S. Dennis. **Vernalization-repression of Arabidopsis FLC requires promoter sequences but not antisense transcripts.** *PLoS ONE* 6.6 (2011), e21513 (see p. 60).

[187] S. Marquardt, O. Raitskin, Z. Wu, F. Liu, Q. Sun and C. Dean. **Functional consequences of splicing of the antisense transcript COOLAIR on FLC transcription.** *Mol. Cell* 54.1 (2014), 156–165 (see p. 60).

[188] Z.-W. Wang, Z. Wu, O. Raitskin, Q. Sun and C. Dean. **Antisense-mediated FLC transcriptional repression requires the P-TEFb transcription elongation factor.** *Proc. Natl. Acad. Sci. U. S. A.* 111.20 (2014), 7468–7473 (see p. 60).

[189] C. Shindo, C. Lister, P. Crevillen, M. Nordborg and C. Dean. **Variation in the epigenetic silencing of FLC contributes to natural variation in Arabidopsis vernalization response.** *Genes Dev.* 20.22 (2006), 3079–3083 (see p. 62).

[190] M. A. Rizzo, G. H. Springer, B. Granada and D. W. Piston. **An improved cyan fluorescent protein variant useful for FRET.** *Nat. Biotechnol.* 22.4 (2004), 445–449 (see p. 72).

[191] N. C. Shaner, R. E. Campbell, P. A. Steinbach, B. N. G. Giepmans, A. E. Palmer and R. Y. Tsien. **Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein.** *Nat. Biotechnol.* 22.12 (2004), 1567–1572 (see p. 72).

[192] J. D. G. Jones, L. Shlumukov, F. Carland, J. English, S. R. Scofield, G. J. Bishop and K. Harrison. **Effective vectors for transformation, expression of heterologous genes, and assaying transposon excision in transgenic plants**. *Transgen. Res.* 1.6 (1992), 285–297 (see pp. 72, 74, 238).

[193] H. Inoue, H. Nojima and H. Okayama. **High efficiency transformation of Escherichia coli with plasmids.** *Gene* 96.1 (1990), 23–28 (see p. 72).

[194] A. A. Wise, Z. Liu and A. N. Binns. **Three methods for the introduction of foreign DNA into Agrobacterium.** *Methods Mol. Biol.* 343 (2006), 43–53 (see p. 73).

[195] R. Hellens, P. Mullineaux and H. Klee. **Technical Focus:a guide to Agrobacterium binary Ti vectors.** *Trends Plant Sci.* 5.10 (2000), 446–451 (see p. 73).

[196] C. J. Thompson, N. R. Movva, R. Tizard, R. Crameri, J. E. Davies, M. Lauwereys and J. Botterman. **Characterization of the herbicide-resistance gene bar from Streptomyces hygroscopicus.** *EMBO J.* 6.9 (1987), 2519–2523 (see p. 74).

[197] M. D. Block, J. Botterman, M. Vandewiele, J. Dockx, C. Thoen, V. Gosselé, N. R. Movva, C. Thompson, M. V. Montagu and J. Leemans. **Engineering herbicide resistance in plants by expression of a detoxifying enzyme.** *EMBO J.* 6.9 (1987), 2513–2518 (see p. 74).

[198] J. G. Bartlett, S. C. Alves, M. Smedley, J. W. Snape and W. A. Harwood. **High-throughput Agrobacterium-mediated barley transformation.** *Plant Methods* 4.1 (2008), 22 (see p. 74).

[199] N. Etheridge, Y. Trusov, J. P. Verbelen and J. R. Botella. **Characterization of ATDRG1, a member of a new class of GTP-binding proteins in plants.** *Plant Mol. Biol.* 39.6 (1999), 1113–1126 (see p. 74).

[200] T. Czechowski, M. Stitt, T. Altmann, M. K. Udvardi and W.-R. Scheible. **Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis.** *Plant Physiol.* 139.1 (2005), 5–17 (see pp. 75, 78).

[201] K. J. Livak and T. D. Schmittgen. **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.** *Methods* 25.4 (2001), 402–408 (see p. 75).

[202] A.-V. Gendrel, Z. Lippman, C. Yordan, V. Colot and R. A. Martienssen. **Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1.** *Science* 297.5588 (2002), 1871–1873 (see p. 77).

[203] A. T. Wierzbicki, J. R. Haag and C. S. Pikaard. **Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes.** *Cell* 135.4 (2008), 635–648 (see p. 77).

[204] K. Lu, W. Ye, L. Zhou, L. B. Collins, X. Chen, A. Gold, L. M. Ball and J. A. Swenberg. **Structural characterization of formaldehyde-induced cross-links between amino acids and deoxynucleosides and their oligomers.** *J. Am. Chem. Soc.* 132.10 (2010), 3388–3399 (see p. 77).

[205] K. Poorey, R. Viswanathan, M. N. Carver, T. S. Karpova, S. M. Cirimotich, J. G. McNally, S. Bekiranov and D. T. Auble. **Measuring chromatin interaction dynamics on the second time scale at single-copy genes.** *Science* 342.6156 (2013), 369–372 (see p. 77).

[206] X. Zhang, O. Clarenz, S. Cokus, Y. V. Bernatavichute, M. Pellegrini, J. Goodrich and S. E. Jacobsen. **Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis**. *PLoS Biol.* 5.5 (2007), 1026–1035 (see p. 78).

[207] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak and A. Cardona. **Fiji: an open-source platform for biological-image analysis.** *Nat. Meth.* 9.7 (2012), 676–682 (see pp. 80, 82).

[208] S. Berry, M. Hartley, T. S. G. Olsson, C. Dean and M. Howard. **Local chromatin environment of a Polycomb target gene instructs its own epigenetic inheritance.** *eLife* 4 (2015), e07205 (see pp. 80, 88, 89, 95, 131, 217).

[209] L. Johnson, S. Mollah, B. A. Garcia, T. L. Muratore, J. Shabanowitz, D. F. Hunt and S. E. Jacobsen. **Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications.** *Nucl. Acids Res.* 32.22 (2004), 6511–6518 (see p. 93).

[210] J. J. Infante, G. L. Law and E. T. Young. **Analysis of Nucleosome Positioning Using a Nucleosome-Scanning Assay**. In: *Chromatin Remodeling*. 2012, 63–87 (see pp. 94, 127).

[211] A. D. Rudner, B. E. Hall, T. Ellenberger and D. Moazed. **A nonhistone protein-protein interaction required for assembly of the SIR complex and silent chromatin.** *Mol. Cell. Biol.* 25.11 (2005), 4514–4528 (see p. 104).

[212] D. T. Gillespie. **Exact stochastic simulation of coupled chemical reactions**. *J. Phys. Chem.* 81.25 (1977), 2340–2361 (see pp. 107, 129, 137, 171).

[213] K. Sneppen and I. B. Dodd. **A simple histone code opens many paths to epigenetics.** *PLoS Comput. Biol.* 8.8 (2012), e1002643 (see pp. 110, 128, 135, 137, 139, 148, 172, 173).

[214] G. V. Reddy, M. G. Heisler, D. W. Ehrhardt and E. M. Meyerowitz. **Real-time lineage analysis reveals oriented cell divisions associated with morphogenesis at the shoot apex of Arabidopsis thaliana.** *Development* 131.17 (2004), 4225–4237 (see p. 118).

[215] S. Rosa, F. De Lucia, J. S. Mylne, D. Zhu, N. Ohmido, A. Pendle, N. Kato, P. Shaw and C. Dean. **Physical clustering of FLC alleles during Polycomb-mediated epigenetic silencing in vernalization.** *Genes Dev.* 27.17 (2013), 1845–1850 (see p. 120).

[216] W. W. Schwabe. **Factors controlling Flowering in the Chrysanthemum IV. THE SITE OF VERNALIZATION AND TRANSLOCATION OF THE STIMULUS**. *J. Exp. Bot.* 5.3 (1954), 389–400 (see p. 122).

[217] S. J. Wellensiek. **Dividing cells as the locus for vernalization**. *Nature* 195.4838 (1962), 307–308 (see p. 122).

[218] S. J. Wellensiek. **Dividing Cells as the Prerequisite for Vernalization.** *Plant Physiol.* 39.5 (1964), 832–835 (see p. 122).

[219] S. Henikoff, T. Furuyama and K. Ahmad. **Histone variants, nucleosome assembly and epigenetic inheritance**. *Trends Genet.* 20.7 (2004), 320–326 (see p. 123).

[220] M. Ingouff and F. Berger. **Histone3 variants in plants**. *Chromosoma* 119.1 (2009), 27–33 (see p. 123).

[221] Y. Jacob, E. Bergamin, M. T. A. Donoghue, V. Mongeon, C. LeBlanc, P. Voigt, C. J. Underwood, J. S. Brunzelle, S. D. Michaels, D. Reinberg, J.-F. Couture and R. A. Martienssen. **Selective methylation of histone H3 variant H3.1 regulates heterochromatin replication.** *Science* 343.6176 (2014), 1249–1253 (see p. 123).

[222] A. S. Larsson, K. Landberg and D. R. Meeks-Wagner. **The TERMINAL FLOWER2 (TFL2) Gene Controls the Reproductive Transition and Meristem Identity in Arabidopsis thaliana**. *Genetics* 149.2 (1998), 597–605 (see pp. 127, 179, 194).

[223] T. Kotake, S. Takada, K. Nakahigashi, M. Ohto and K. Goto. **Arabidopsis TERMINAL FLOWER 2 gene encodes a heterochromatin protein 1 homolog and represses both FLOWERING LOCUS T to regulate flowering time and several floral homeotic genes.** *Plant Cell Physiol.* 44.6 (2003), 555–564 (see pp. 127, 179, 194, 237).

[224] M. Matsumoto and T. Nishimura. **Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator**. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 8.1 (1998), 3–30 (see p. 128).

[225] R Core Team. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing. Vienna, Austria, 2015 (see p. 128).

[226] H. Wickham. **ggplot2: elegant graphics for data analysis**. New York: Springer, 2009 (see p. 128).

[227] H. Wickham. **Reshaping data with the reshape package**. *Journal of Statistical Software* 21.12 (2007), 1–20 (see p. 128).

[228] H. Wickham. **The split-apply-combine strategy for data analysis**. *Journal of Statistical Software* 40.1 (2011), 1–29 (see p. 128).

[229] H. Wickham. **scales: Scale functions for graphics.** R package version 0.2.4. 2014 (see p. 128).

[230] K. Sneppen, M. A. Micheelsen and I. B. Dodd. **Ultrasensitive gene regulation by positive feedback loops in nucleosome modification**. *Mol. Syst. Biol.* 4 (2008) (see pp. 128, 148).

[231] D. J. Wilkinson. **Stochastic modelling for quantitative description of heterogeneous biological systems**. *Nat. Rev. Genet.* 10.2 (2009), 122–133 (see p. 129).

[232] L. C. M. Anink-Groenen, T. R. Maarleveld, P. J. Verschure and F. J. Bruggeman. **Mechanistic stochastic model of histone modification pattern formation.** *Epigenetics & Chromatin* 7.1 (2014), 30–16 (see pp. 129, 135, 148).

[233] H. Zhang, X. J. Tian, A. Mukhopadhyay, K. S. Kim and J. Xing. **Statistical mechanics model for the dynamics of collective epigenetic histone modification**. *Phys. Rev. Lett.* (2014) (see p. 129).

[234] J. O. Haerter, C. Lövkvist, I. B. Dodd and K. Sneppen. **Collaboration between CpG sites is needed for stable somatic inheritance of DNA methylation states.** *Nucl. Acids Res.* 42.4 (2014), 2235–2244 (see p. 129).

[235] A. Slepoy, A. P. Thompson and S. J. Plimpton. **A constant-time kinetic Monte Carlo algorithm for simulation of large biochemical reaction networks.** *J. Chem. Phys.* 128.20 (2008), 205101 (see p. 129).

[236] L. J. Gaydos, W. Wang and S. Strome. **H3K27me and PRC2 transmit a memory of repression across generations and during development.** *Science* 345.6203 (2014), 1515–1518 (see pp. 131, 144, 167, 216, 217).

[237] J. van der Vlag and A. P. Otte. **Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation.** *Nat. Genet.* 23.4 (1999), 474–478 (see p. 131).

[238] R. R. Roseman, K. Morgan, D. R. Mallin, R. Roberson, T. J. Parnell, D. J. Bornemann, J. A. Simon and P. K. Geyer. **Long-range repression by multiple polycomb group (PcG) proteins targeted by fusion to a defined DNA-binding domain in Drosophila.** *Genetics* 158.1 (2001), 291–307 (see p. 131).

[239] G. Cavalli and R. Paro. **The Drosophila Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis.** *Cell* 93.4 (1998), 505–518 (see p. 132).

[240] S. Schmitt, M. Prestel and R. Paro. **Intergenic transcription through a polycomb group response element counteracts silencing.** *Genes Dev.* 19.6 (2005), 697–708 (see p. 132).

[241] W. L. Ku, M. Girvan, G.-C. Yuan, F. Sorrentino and E. Ott. **Modeling the dynamics of bivalent histone modifications.** *PLoS ONE* 8.11 (2013), e77944 (see p. 135).

[242] V. M. Studitsky, W. Walter, M. Kireeva, M. Kashlev and G. Felsenfeld. **Chromatin remodeling by RNA polymerases.** *Trends Biochem. Sci.* 29.3 (2004), 127–135 (see p. 135).

[243] O. I. Kulaeva, F.-K. Hsieh, H.-W. Chang, D. S. Luse and V. M. Studitsky. **Mechanism of transcription through a nucleosome by RNA polymerase II.** *Biochim. Biophys. Acta* 1829.1 (2013), 76–83 (see p. 135).

[244] V. A. Bondarenko, L. M. Steele, A. Ujvári, D. A. Gaykalova, O. I. Kulaeva, Y. S. Polikanov, D. S. Luse and V. M. Studitsky. **Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II.** *Mol. Cell* 24.3 (2006), 469–479 (see p. 135).

[245] L. Bintu, T. Ishibashi, M. Dangkulwanich, Y.-Y. Wu, L. Lubkowska, M. Kashlev and C. Bustamante. **Nucleosomal Elements that Control the Topography of the Barrier to Transcription**. *Cell* 151.4 (2012), 738–749 (see p. 135).

[246] S. Venkatesh and J. L. Workman. **Histone exchange, chromatin structure and the regulation of transcription**. *Nat. Rev. Mol. Cell Biol.* 16.3 (2015), 178–189 (see pp. 135, 143).

[247] A. J. Gossett and J. D. Lieb. **In vivo effects of histone H3 depletion on nucleosome occupancy and position in Saccharomyces cerevisiae.** *PLoS Genet.* 8.6 (2012), e1002771 (see p. 135).

[248] R. K. Chodavarapu, S. Feng, Y. V. Bernatavichute, P.-Y. Chen, H. Stroud, Y. Yu, J. A. Hetzel, F. Kuo, J. Kim, S. J. Cokus, D. Casero, M. Bernal, P. Huijser, A. T. Clark, U. Krämer, S. S. Merchant, X. Zhang, S. E. Jacobsen and M. Pellegrini. **Relationship between nucleosome positioning and DNA methylation.** *Nature* 466.7304 (2010), 388–392 (see p. 136).

[249] L. Hong, G. P. Schroth, H. R. Matthews, P. Yau and E. M. Bradbury. **Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA.** *J. Biol. Chem.* 268.1 (1993), 305–314 (see p. 136).

[250] V. G. Allfrey, R. Faulkner and A. E. Mirsky. **Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis.** *Proc. Natl. Acad. Sci. U. S. A.* 51 (1964), 786–794 (see p. 136).

[251] D. Pasini, M. Malatesta, H. R. Jung, J. Walfridsson, A. Willer, L. Olsson, J. Skotte, A. Wutz, B. Porse, O. N. Jensen and K. Helin. **Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes.** *Nucl. Acids Res.* 38.15 (2010), 4958–4969 (see p. 136).

[252] X. Shen, Y. Liu, Y.-J. Hsu, Y. Fujiwara, J. Kim, X. Mao, G.-C. Yuan and S. H. Orkin. **EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in**

maintaining stem cell identity and executing pluripotency. *Mol. Cell* 32.4 (2008), 491–502 (see p. 136).

[253] R. Margueron, G. Li, K. Sarma, A. Blais, J. Zavadil, C. L. Woodcock, B. D. Dynlacht and D. Reinberg. **Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms.** *Mol. Cell* 32.4 (2008), 503–518 (see p. 136).

[254] J. M. Berg, J. L. Tymoczko and L. Stryer. **Biochemistry, Fifth Edition**. W. H. Freeman, 2002 (see p. 139).

[255] D. Patnaik, H. G. Chin, P.-O. Estève, J. Benner, S. E. Jacobsen and S. Pradhan. **Substrate specificity and kinetic mechanism of mammalian G9a histone H3 methyltransferase.** *J. Biol. Chem.* 279.51 (2004), 53248–53258 (see p. 140).

[256] H. G. Chin, D. Patnaik, P.-O. Estève, S. E. Jacobsen and S. Pradhan. **Catalytic properties and kinetic mechanism of human recombinant Lys-9 histone H3 methyltransferase suv39h1: participation of the chromodomain in enzymatic catalysis**. *Biochemistry* 45.10 (2006), 3272–3284 (see p. 140).

[257] C. J. Sneeringer, M. P. Scott, K. W. Kuntz, S. K. Knutson, R. M. Pollock, V. M. Richon and R. A. Copeland. **Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas.** *Proc. Natl. Acad. Sci. U. S. A.* 107.49 (2010), 20980–20985 (see p. 140).

[258] M. T. McCabe et al. **Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27).** *Proc. Natl. Acad. Sci. U. S. A.* 109.8 (2012), 2989–2994 (see pp. 140, 141).

[259] T. J. Wigle, S. K. Knutson, L. Jin, K. W. Kuntz, R. M. Pollock, V. M. Richon, R. A. Copeland and M. P. Scott. **The Y641C mutation of EZH2 alters substrate specificity for histone H3 lysine 27 methylation states**. *FEBS Letters* 585.19 (2011), 3011–3014 (see p. 140).

[260] S. Schoeftner, A. K. Sengupta, S. Kubicek, K. Mechtler, L. Spahn, H. Koseki, T. Jenuwein and A. Wutz. **Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing.** *EMBO J.* 25.13 (2006), 3110–3122 (see p. 140).

[261] D. Pasini, A. P. Bracken, J. B. Hansen, M. Capillo and K. Helin. **The polycomb group protein Suz12 is required for embryonic stem cell differentiation.** *Mol. Cell. Biol.* 27.10 (2007), 3769–3779 (see p. 140).

[262] H. R. Jung, D. Pasini, K. Helin and O. N. Jensen. **Quantitative mass spectrometry of histones H3.2 and H3.3 in Suz12-deficient mouse embryonic stem cells reveals distinct, dynamic post-translational modifications at Lys-27 and Lys-36.** *Mol. Cell. Proteomics* 9.5 (2010), 838–850 (see p. 140).

[263] K. J. Ferrari, A. Scelfo, S. Jammula, A. Cuomo, I. Barozzi, A. Stützer, W. Fischle, T. Bonaldi and D. Pasini. **Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity.** *Mol. Cell* 53.1 (2014), 49–62 (see p. 140).

[264] K. Cui, C. Zang, T.-Y. Roh, D. E. Schones, R. W. Childs, W. Peng and K. Zhao. **Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation.** *Cell Stem Cell* 4.1 (2009), 80–93 (see p. 140).

[265] L. A. Steiner, V. P. Schulz, Y. Maksimova, C. Wong and P. G. Gallagher. **Patterns of histone H3 lysine 27 monomethylation and erythroid cell type-specific gene expression.** *J. Biol. Chem.* 286.45 (2011), 39457–39465 (see p. 140).

[266] B. M. Zee, L.-M. P. Britton, D. Wolle, D. M. Haberman and B. A. Garcia. **Origins and formation of histone methylation across the human cell cycle.** *Mol. Cell. Biol.* 32.13 (2012), 2503–2514 (see pp. 140, 144).

[267] M. Xu, W. Wang, S. Chen and B. Zhu. **A model for mitotic inheritance of histone lysine methylation.** *EMBO Rep.* 13.1 (2012), 60–67 (see pp. 140, 144, 148).

[268] S. M. Kooistra and K. Helin. **Molecular mechanisms and potential functions of histone demethylases.** *Nat. Rev. Mol. Cell Biol.* 13.5 (2012), 297–311 (see p. 140).

[269] S. Hong, Y.-W. Cho, L.-R. Yu, H. Yu, T. D. Veenstra and K. Ge. **Identification of JmjC domain-containing UTX and JMJD3 as histone H3 lysine 27 demethylases.** *Proc. Natl. Acad. Sci. U. S. A.* 104.47 (2007), 18439–18444 (see p. 140).

[270] H. Tagami, D. Ray-Gallet, G. Almouzni and Y. Nakatani. **Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis.** *Cell* 116.1 (2004), 51–61 (see pp. 144, 146).

[271] Y. Mito, J. G. Henikoff and S. Henikoff. **Histone replacement marks the boundaries of cis-regulatory domains.** *Science* 315.5817 (2007), 1408–1411 (see p. 144).

[272] Y. Mito, J. G. Henikoff and S. Henikoff. **Genome-scale profiling of histone H3.3 replacement patterns.** *Nat. Genet.* 37.10 (2005), 1090–1097 (see p. 144).

[273] C. Jin, C. Zang, G. Wei, K. Cui, W. Peng, K. Zhao and G. Felsenfeld. **H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions.** *Nat. Genet.* 41.8 (2009), 941–945 (see p. 144).

[274] S.-L. Ooi, J. G. Henikoff and S. Henikoff. **A native chromatin purification system for epigenomic profiling in Caenorhabditis elegans.** *Nucl. Acids Res.* 38.4 (2010), e26–e26 (see p. 144).

[275] H. Shu, M. Nakamura, A. Siretskiy, L. Borghi, I. Moraes, T. Wildhaber, W. Gruissem and L. Hennig. **Arabidopsis replacement histone variant H3.3 occupies promoters of regulated genes.** *Genome Biol.* 15.4 (2014), R62 (see p. 144).

[276] B. M. Zee, R. S. Levin, B. Xu, G. LeRoy, N. S. Wingreen and B. A. Garcia. **In vivo residue-specific histone methylation dynamics.** *J. Biol. Chem.* 285.5 (2010), 3341–3350 (see p. 144).

[277] R. Bermejo, M. S. Lai and M. Foiani. **Preventing Replication Stress to Maintain Genome Stability: Resolving Conflicts between Replication and Transcription.** *Mol. Cell* 45.6 (2012), 710–718 (see p. 145).

[278] A. N. D. Scharf, T. K. Barth and A. Imhof. **Establishment of histone modifications after chromatin assembly.** *Nucl. Acids Res.* 37.15 (2009), 5032–5040 (see p. 148).

[279] C. Alabert, J.-C. Bukowski-Wills, S.-B. Lee, G. Kustatscher, K. Nakamura, F. de Lima Alves, P. Menard, J. Mejlvang, J. Rappsilber and A. Groth. **Nascent chromatin capture proteomics determines chromatin dynamics during DNA replication and identifies unknown fork components.** *Nat. Cell Biol.* 16.3 (2014), 281–293 (see pp. 148, 167).

[280] M. Thattai and A. van Oudenaarden. **Intrinsic noise in gene regulatory networks.** *Proc. Natl. Acad. Sci. U. S. A.* 98.15 (2001), 8614–8619 (see p. 155).

[281] J. Paulsson. **Summing up the noise in gene networks.** *Nature* 427.6973 (2004), 415–418 (see p. 155).

[282] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas and S. Tyagi. **Stochastic mRNA synthesis in mammalian cells.** *PLoS Biol.* 4.10 (2006), e309 (see p. 155).

[283] P. Hammar, P. Leroy, A. Mahmutovic, E. G. Marklund, O. G. Berg and J. Elf. **The lac repressor displays facilitated diffusion in living cells.** *Science* 336.6088 (2012), 1595–1598 (see p. 155).

[284] A. Grönlund, P. Lötstedt and J. Elf. **Transcription factor binding kinetics constrain noise suppression via negative feedback.** *Nat. Commun.* 4 (2013), 1864 (see p. 155).

[285] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman and A. van Oudenaarden. **Regulation of noise in the expression of a single gene.** *Nat. Genet.* 31.1 (2002), 69–73 (see pp. 155, 174, 175).

[286] A. Mohd-Sarip, A. Lagarou, C. M. Doyen, J. A. van der Knaap, U. Aslan, K. Bezstarosti, Y. Yassin, H. W. Brock, J. A. A. Demmers and C. P. Verrijzer. **Transcription-independent function of Polycomb group protein PSC in cell cycle control.** *Science* 336.6082 (2012), 744–747 (see p. 162).

[287] M. Hosogane, R. Funayama, Y. Nishida, T. Nagashima and K. Nakayama. **Ras-Induced Changes in H3K27me3 Occur after Those in Transcriptional Activity**. *PLoS Genet.* 9.8 (2013), e1003698–16 (see p. 162).

[288] G. Cavalli and T. Misteli. **Functional implications of genome topology**. *Nat. Struct. Mol. Biol.* 20.3 (2013), 290–299 (see p. 166).

[289] P. Therizols, R. S. Illingworth, C. Courilleau, S. Boyle, A. J. Wood and W. A. Bickmore. **Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells.** *Science* 346.6214 (2014), 1238–1242 (see p. 166).

[290] M.-C. Keogh, S. K. Kurdistani, S. A. Morris, S. H. Ahn, V. Podolny, S. R. Collins, M. Schuldiner, K. Chin, T. Punna, N. J. Thompson, C. Boone, A. Emili, J. S. Weissman, T. R. Hughes, B. D. Strahl, M. Grunstein, J. F. Greenblatt, S. Buratowski and N. J. Krogan. **Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex.** *Cell* 123.4 (2005), 593–605 (see p. 166).

[291] C. A. Spencer, M. J. Kruhlak, H. L. Jenkins, X. Sun and D. P. Bazett-Jones. **Mitotic transcription repression in vivo in the absence of nucleosomal chromatin condensation.** *J. Cell Biol.* 150.1 (2000), 13–26 (see p. 167).

[292] J. P. Fonseca, P. A. Steffen, S. Müller, J. Lu, A. Sawicka, C. Seiser and L. Ringrose. **In vivo Polycomb kinetics and mitotic chromatin binding distinguish stem cells from differentiated cells.** *Genes Dev.* 26.8 (2012), 857–871 (see p. 167).

[293] X. Zhang, Y. V. Bernatavichute, S. Cokus, M. Pellegrini and S. E. Jacobsen. **Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana.** *Genome Biol.* 10.6 (2009), R62 (see pp. 169, 203).

[294] E. Brookes, I. de Santiago, D. Hebenstreit, K. J. Morris, T. Carroll, S. Q. Xie, J. K. Stock, M. Heidemann, D. Eick, N. Nozaki, H. Kimura, J. Ragoussis, S. A. Teichmann and A. Pombo. **Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs.** *Cell Stem Cell* 10.2 (2012), 157–170 (see p. 169).

[295] S. Poux, B. Horard, C. J. A. Sigrist and V. Pirrotta. **The Drosophila trithorax protein is a coactivator required to prevent re-establishment of polycomb silencing.** *Development* 129.10 (2002), 2483–2493 (see p. 169).

[296] T. Klymenko and J. Müller. **The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins.** *EMBO Rep.* 5.4 (2004), 373–377 (see p. 169).

[297] D. Bratsun, D. Volfson, L. S. Tsimring and J. Hasty. **Delay-induced stochastic oscillations in gene regulation.** *Proc. Natl. Acad. Sci. U. S. A.* 102.41 (2005), 14593–14598 (see p. 171).

[298] J. L. Rinn and H. Y. Chang. **Genome Regulation by Long Noncoding RNAs.** *Annu. Rev. Biochem.* 81.1 (2012), 145–166 (see p. 177).

[299] L. Yang, J. E. Froberg and J. T. Lee. **Long noncoding RNAs: fresh perspectives into the RNA world.** *Trends Biochem. Sci.* 39.1 (2014), 35–43 (see p. 177).

[300] E. J. Steinmetz, C. L. Warren, J. N. Kuehner, B. Panbehi, A. Z. Ansari and D. A. Brow. **Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase.** *Mol. Cell* 24.5 (2006), 735–746 (see p. 178).

[301] M. G. Guenther, S. S. Levine, L. A. Boyer, R. Jaenisch and R. A. Young. **A chromatin landmark and transcription initiation at most promoters in human cells.** *Cell* 130.1 (2007), 77–88 (see p. 178).

[302] M. Bühler, A. Verdel and D. Moazed. **Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing.** *Cell* 125.5 (2006), 873–886 (see p. 178).

[303] D. L. Bentley. **Coupling mRNA processing with transcription in time and space**. *Nat. Rev. Genet.* 15.3 (2014), 163–175 (see pp. 178, 230).

[304] J. Berretta and A. Morillon. **Pervasive transcription constitutes a new level of eukaryotic genome regulation.** *EMBO Rep.* 10.9 (2009), 973–982 (see p. 178).

[305] R. K. Gudipati, Z. Xu, A. Lebreton, B. Séraphin, L. M. Steinmetz, A. Jacquier and D. Libri. **Extensive degradation of RNA precursors by the exosome in wild-type cells.** *Mol. Cell* 48.3 (2012), 409–421 (see p. 178).

[306] T. Trcek and R. H. Singer. **The cytoplasmic fate of an mRNP is determined cotranscriptionally: exception or rule?** *Genes Dev.* 24.17 (2010), 1827–1831 (see p. 178).

[307] D. Moazed. **Small RNAs in transcriptional gene silencing and genome defence**. *Nature* 457.7228 (2009), 413–420 (see p. 178).

[308] J. Zhao, T. K. Ohsumi, J. T. Kung, Y. Ogawa, D. J. Grau, K. Sarma, J.-J. Song, R. E. Kingston, M. Borowsky and J. T. Lee. **Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq**. *Mol. Cell* 40.6 (2010), 939–953 (see p. 179).

[309] A. Kanhere, K. Viiri, C. C. Araújo, J. Rasaiyaah, R. D. Bouwman, W. A. Whyte, C. F. Pereira, E. Brookes, K. Walker, G. W. Bell, A. Pombo, A. G. Fisher, R. A. Young and R. G. Jenner. **Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2.** *Mol. Cell* 38.5 (2010), 675–688 (see p. 179).

[310] C. Maison and G. Almouzni. **HP1 and the dynamics of heterochromatin maintenance.** *Nat. Rev. Mol. Cell Biol.* 5.4 (2004), 296–304 (see p. 179).

[311] K. Nakahigashi, Z. Jasencakova, I. Schubert and K. Goto. **The Arabidopsis heterochromatin protein1 homolog (TERMINAL FLOWER2) silences genes within the euchromatic region but not genes positioned in heterochromatin.** *Plant Cell Physiol.* 46.11 (2005), 1747–1756 (see p. 179).

[312] V. Exner, E. Aichinger, H. Shu, T. Wildhaber, P. Alfarano, A. Caflisch, W. Gruissem, C. Köhler and L. Hennig. **The Chromodomain of LIKE HETEROCHROMATIN PROTEIN 1 Is Essential for H3K27me3 Binding and Function during Arabidopsis Development**. *PLoS ONE* 4.4 (2009), e5335 (see pp. 179, 181, 186, 199).

[313] X. Liu, Y. J. Kim, R. Müller, R. E. Yumul, C. Liu, Y. Pan, X. Cao, J. Goodrich and X. Chen. **AGAMOUS terminates floral stem cell maintenance in Arabidopsis by directly repressing WUSCHEL through recruitment of Polycomb Group proteins.** *Plant Cell* 23.10 (2011), 3654–3670 (see p. 179).

[314] G. W. Haughn, L. Davin, M. Giblin and E. W. Underhill. **Biochemical Genetics of Plant Secondary Metabolites in Arabidopsis thaliana: The Glucosinolates.** *Plant Physiol.* 97.1 (1991), 217–226 (see p. 179).

[315] J. H. Kim, T. P. Durrett, R. L. Last and G. Jander. **Characterization of the Arabidopsis TU8 glucosinolate mutation, an allele of TERMINAL FLOWER2.** *Plant Mol. Biol.* 54.5 (2004), 671–682 (see p. 179).

[316] V. Gaudin, M. Libault, S. Pouteau, T. Juul, G. Zhao, D. Lefebvre and O. Grandjean. **Mutations in LIKE HETEROCHROMATIN PROTEIN 1 affect flowering time and plant architecture in Arabidopsis.** *Development* 128.23 (2001), 4847–4858 (see pp. 179, 191, 194).

[317] R. Paro and D. S. Hogness. **The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila.** *Proc. Natl. Acad. Sci. U. S. A.* 88.1 (1991), 263–267 (see p. 180).

[318] A. J. Bannister, P. Zegerman, J. F. Partridge, E. A. Miska, J. O. Thomas, R. C. Allshire and T. Kouzarides. **Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain.** *Nature* 410.6824 (2001), 120–124 (see pp. 180, 186).

[319] S. A. Jacobs and S. Khorasanizadeh. **Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail.** *Science* 295.5562 (2002), 2080–2083 (see pp. 180, 186).

[320] P. R. Nielsen, D. Nietlispach, H. R. Mott, J. Callaghan, A. Bannister, T. Kouzarides, A. G. Murzin, N. V. Murzina and E. D. Laue. **Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9.** *Nature* 416.6876 (2002), 103–107 (see p. 180).

[321] W. Fischle, Y. Wang, S. A. Jacobs, Y. Kim, C. D. Allis and S. Khorasanizadeh. **Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains.** *Genes Dev.* 17.15 (2003), 1870–1881 (see pp. 180, 186, 202).

[322] R. Aasland and A. F. Stewart. **The chromo shadow domain, a second chromo domain in heterochromatin-binding protein 1, HP1.** *Nucl. Acids Res.* 23.16 (1995), 3168–3173 (see p. 180).

[323] T. Yamada, R. Fukuda, M. Himeno and K. Sugimoto. **Functional domain structure of human heterochromatin protein HP1(Hsalpha): involvement of internal DNA-binding and C-terminal self-association domains in the formation of discrete dots in interphase nuclei.** *J. Biochem.* 125.4 (1999), 832–837 (see p. 180).

[324] S. V. Brasher, B. O. Smith, R. H. Fogh, D. Nietlispach, A. Thiru, P. R. Nielsen, R. W. Broadhurst, L. J. Ball, N. V. Murzina and E. D. Laue. **The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer.** *EMBO J.* 19.7 (2000), 1587–1597 (see p. 180).

[325] N. P. Cowieson, J. F. Partridge, R. C. Allshire and P. J. McLaughlin. **Dimerisation of a chromo shadow domain and distinctions from the chromodomain as revealed by structural analysis.** *Curr. Biol.* 10.9 (2000), 517–525 (see pp. 180, 182, 183).

[326] D. Canzio, M. Liao, N. Naber, E. Pate, A. Larson, S. Wu, D. B. Marina, J. F. Garcia, H. D. Madhani, R. Cooke, P. Schuck, Y. Cheng and G. J. Narlikar. **A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly**. *Nature* 496.7445 (2013), 377–381 (see pp. 180, 186, 199, 202, 203, 224).

[327] J. C. Eissenberg. **Structural biology of the chromodomain: form and function.** *Gene* 496.2 (2012), 69–78 (see p. 180).

[328] K. Sugimoto, T. Yamada, Y. Muro and M. Himeno. **Human homolog of Drosophila heterochromatin-associated protein 1 (HP1) is a DNA-binding protein which possesses a DNA-binding motif with weak similarity to that of human centromere protein C (CENP-C).** *J. Biochem.* 120.1 (1996), 153–159 (see p. 181).

[329] T. Zhao, T. Heyduk, C. D. Allis and J. C. Eissenberg. **Heterochromatin protein 1 binds to nucleosomes and DNA in vitro.** *J. Biol. Chem.* 275.36 (2000), 28332–28338 (see p. 181).

[330] D. Canzio, E. Y. Chang, S. Shankar, K. M. Kuchenbecker, M. D. Simon, H. D. Madhani, G. J. Narlikar and B. Al-Sady. **Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly.** *Mol. Cell* 41.1 (2011), 67–81 (see pp. 182, 202).

[331] S. Kosugi, M. Hasebe, N. Matsumura, H. Takashima, E. Miyamoto-Sato, M. Tomita and H. Yanagawa. **Six Classes of Nuclear Localization Signals Specific to Different Binding Grooves of Importin**. *J. Biol. Chem.* 284.1 (2008), 478–485 (see p. 192).

[332] S. Takada and K. Goto. **TERMINAL FLOWER2, an Arabidopsis homolog of HETEROCHROMATIN PROTEIN1, counteracts the activation of FLOWERING LOCUS T by CONSTANS in the vascular tissues of leaves to regulate flowering time**. *Plant Cell* 15.12 (2003), 2856–2865 (see p. 194).

[333] J. Adrian, S. Farrona, J. J. Reimer, M. C. Albani, G. Coupland and F. Turck. **cis-Regulatory elements and chromatin state coordinately control temporal and spatial expression of FLOWERING LOCUS T in Arabidopsis.** *Plant Cell* 22.5 (2010), 1425–1440 (see pp. 195, 202).

[334] M. D. Stewart, J. Li and J. Wong. **Relationship between histone H3 lysine 9 methylation, transcription repression, and heterochromatin protein 1 recruitment.** *Mol. Cell. Biol.* 25.7 (2005), 2525–2538 (see p. 199).

[335] K. A. Hines, D. E. Cryderman, K. M. Flannery, H. Yang, M. W. Vitalini, T. Hazelrigg, C. A. Mizzen and L. L. Wallrath. **Domains of heterochromatin protein 1 required for Drosophila melanogaster heterochromatin spreading.** *Genetics* 182.4 (2009), 967–977 (see p. 199).

[336] S. A. Jacobs, W. Fischle and S. Khorasanizadeh. **Assays for the determination of structure and dynamics of the interaction of the chromodomain with histone peptides.** *Meth. Enzymol.* 376 (2004), 131–148 (see p. 202).

[337] R. Margueron and D. Reinberg. **Chromatin structure and the inheritance of epigenetic information**. *Nat. Rev. Genet.* 11.4 (2010), 285–296 (see p. 203).

[338] K. P. MUller, F. Erdel, M. Caudron-Herger, C. Marth, B. D. Fodor, M. Richter, M. Scaranaro, J. Beaudouin, M. Wachsmuth and K. Rippe. **Multiscale analysis of dynamics and interactions of heterochromatin protein 1 by fluorescence fluctuation microscopy.** *Biophys. J.* 97.11 (2009), 2876–2885 (see pp. 203, 204).

[339] M. Wachsmuth, M. Caudron-Herger and K. Rippe. **Genome organization: balancing stability and plasticity.** *Biochim. Biophys. Acta* 1783.11 (2008), 2061–2079 (see p. 203).

[340] B. D. Fodor, S. Kubicek, M. Yonezawa, R. J. O'Sullivan, R. Sengupta, L. Perez-Burgos, S. Opravil, K. Mechtler, G. Schotta and T. Jenuwein. **Jmjd2b antagonizes H3K9 trimethylation at pericentric heterochromatin in mammalian cells.** *Genes Dev.* 20.12 (2006), 1557–1562 (see p. 203).

[341] A. Zemach, Y. Li, H. Ben-Meir, M. Oliva, A. Mosquna, V. Kiss, Y. Avivi, N. Ohad and G. Grafi. **Different domains control the localization and mobility of LIKE HETEROCHROMATIN PROTEIN1 in Arabidopsis nuclei.** *Plant Cell* 18.1 (2006), 133–145 (see pp. 204, 224).

[342] T. Cheutin, S. A. Gorski, K. M. May, P. B. Singh and T. Misteli. **In Vivo Dynamics of Swi6 in Yeast: Evidence for a Stochastic Model of Heterochromatin**. *Mol. Cell. Biol.* 24.8 (2004), 3157–3167 (see p. 204).

[343] R. Festenstein, S. N. Pagakis, K. Hiragami, D. Lyon, A. Verreault, B. Sekkali and D. Kioussis. **Modulation of heterochromatin protein 1 dynamics in primary Mammalian cells.** *Science* 299.5607 (2003), 719–721 (see p. 204).

[344] V. B. Teif, N. Kepper, K. Yserentant, G. Wedemann and K. Rippe. **Affinity, stoichiometry and cooperativity of heterochromatin protein 1 (HP1) binding to nucleosomal arrays**. *J. Phys.: Condens. Matter* (2015), 1–10 (see p. 204).

[345] H. Guan, Z. Zheng, P. H. Grey, Y. Li and D. G. Oppenheimer. **Conservation and divergence of plant LHP1 protein sequences and expression patterns in angiosperms and gymnosperms.** *Mol. Genet. Genomics* 285.5 (2011), 357–373 (see pp. 205, 214).

[346] J. M. Alonso et al. **Genome-Wide Insertional Mutagenesis of Arabidopsis thaliana**. *Science* 301.5633 (2003), 653–657 (see pp. 205, 238).

[347] C. Engler, M. Youles, R. Gruetzner, T.-M. Ehnert, S. Werner, J. D. G. Jones, N. J. Patron and S. Marillonnet. **A golden gate modular cloning toolbox for plants.** *ACS Synth. Biol.* 3.11 (2014), 839–843 (see pp. 207, 212).

[348] M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins. **Clustal W and Clustal X version 2.0.** *Bioinformatics* 23.21 (2007), 2947–2948 (see p. 214).

[349] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes and A. Drummond. **Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.** *Bioinformatics* 28.12 (2012), 1647–1649 (see p. 214).

[350] S. Petruk, Y. Sedkov, D. M. Johnston, J. W. Hodgson, K. L. Black, S. K. Kovermann, S. Beck, E. Canaani, H. W. Brock and A. Mazo. **TrxG and PcG proteins but not methylated histones remain associated with DNA through replication.** *Cell* 150.5 (2012), 922–933 (see p. 217).

[351] A. Gimelbrant, J. N. Hutchinson, B. R. Thompson and A. Chess. **Widespread monoallelic expression on human autosomes.** *Science* 318.5853 (2007), 1136–1140 (see p. 217).

[352] A.-V. Gendrel, M. Attia, C.-J. Chen, P. Diabangouaya, N. Servant, E. Barillot and E. Heard. **Developmental dynamics and disease potential of random monoallelic gene expression.** *Dev. Cell* 28.4 (2014), 366–380 (see p. 217).

[353] M. A. Eckersley-Maslin, D. Thybert, J. H. Bergmann, J. C. Marioni, P. Flicek and D. L. Spector. **Random monoallelic gene expression increases upon embryonic stem cell differentiation.** *Dev. Cell* 28.4 (2014), 351–365 (see p. 217).

[354] Q. Deng, D. Ramsköld, B. Reinius and R. Sandberg. **Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells.** *Science* 343.6167 (2014), 193–196 (see p. 217).

[355] M. A. Eckersley-Maslin and D. L. Spector. **Random monoallelic expression: regulating gene expression one allele at a time**. *Trends Genet.* 30.6 (2014), 237–244 (see p. 217).

[356] E. J. Finnegan, K. A. Kovac, E. Jaligot, C. C. Sheldon, W. James Peacock and E. S. Dennis. **The downregulation of FLOWERING LOCUS C (FLC) expression in plants with low levels of DNA methylation and by vernalization occurs by distinct mechanisms.** *Plant J.* 44.3 (2005), 420–432 (see p. 218).

[357] G. E. Zentner and S. Henikoff. **Epigenome editing made easy.** *Nat. Biotechnol.* 33.6 (2015), 606–607 (see p. 218).

[358] S. Konermann, M. D. Brigham, A. E. Trevino, P. D. Hsu, M. Heidenreich, L. Cong, R. J. Platt, D. A. Scott, G. M. Church and F. Zhang. **Optical control of mammalian**

endogenous transcription and epigenetic states. *Nature* 500.7463 (2013), 472–476 (see p. 218).

[359] J. D. Sander and J. K. Joung. **CRISPR-Cas systems for editing, regulating and targeting genomes.** *Nat. Biotechnol.* 32.4 (2014), 347–355 (see p. 218).

[360] I. B. Hilton, A. M. D'Ippolito, C. M. Vockley, P. I. Thakore, G. E. Crawford, T. E. Reddy and C. A. Gersbach. **Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers.** *Nat. Biotechnol.* 33.5 (2015), 510–517 (see p. 218).

[361] N. A. Kearns, H. Pham, B. Tabak, R. M. Genga, N. J. Silverstein, M. Garber and R. Maehr. **Functional annotation of native enhancers with a Cas9-histone demethylase fusion.** *Nat. Meth.* 12.5 (2015), 401–403 (see p. 218).

[362] C. Grimaud, N. Nègre and G. Cavalli. **From genetics to epigenetics: the tale of Polycomb group and trithorax group genes.** *Chromosome Res.* 14.4 (2006), 363–375 (see p. 230).

[363] J. A. Simon and R. E. Kingston. **Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put.** *Mol. Cell* 49.5 (2013), 808–824 (see p. 230).

[364] M. Lubas, P. R. Andersen, A. Schein, A. Dziembowski, G. Kudla and T. H. Jensen. **The Human Nuclear Exosome Targeting Complex Is Loaded onto Newly Synthesized RNA to Direct Early Ribonucleolysis**. *Cell Rep.* 10.2 (2015), 178–192 (see p. 230).

[365] K. Adelman and J. T. Lis. **Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans**. *Nat. Rev. Genet.* 13.10 (2012), 720–731 (see p. 230).

[366] J. R. Karr, J. C. Sanghvi, D. N. Macklin, M. V. Gutschow, J. M. Jacobs, B. Bolival, N. Assad-Garcia, J. I. Glass and M. W. Covert. **A whole-cell computational model predicts phenotype from genotype.** *Cell* 150.2 (2012), 389–401 (see p. 231).

[367] Y. H. Chew, B. Wenden, A. Flis, V. Mengin, J. Taylor, C. L. Davey, C. Tindal, H. Thomas, H. J. Ougham, P. de Reffye, M. Stitt, M. Williams, R. Muetzelfeldt, K. J. Halliday and A. J. Millar. **Multiscale digital Arabidopsis predicts individual organ and whole-organism growth.** *Proc. Natl. Acad. Sci. U. S. A.* 111.39 (2014), E4127–36 (see p. 231).