Gene Regulation mediated by Bivalent Chromatin Domains

E.N. van Schaik
Utrecht University
Master Program: Molecular and Cellular Life Sciences
Master Thesis

Supervisor:
Dr. B.P.W. Rutjens

Scheres Laboratory at Utrecht University
Utrecht, the Netherlands, July 2012

Abstract

During recent years technical advances have made it possible to study chromatin modifications at high resolution in homogenous cell populations. This thesis will focus on recent developments in the field of polycomb-mediated gene silencing and its associated prototypical chromatin mark Histone 3 lysine 27 tri-methylation (H3K27me3). Interestingly, vertebrate stem-cell precursor chromatin contains so-called bivalently marked genes which are enriched for both H3K27me3 and the opposing and hence gene activating mark H3K4me3. Bivalent domains are mostly located at silent genes involved in embryogenesis and development and depend on Polycomb Repressive Complexes 2 (PRC2) and 1 (PRC1) to be established and maintenaned. It is hypothesized that bivalency is crucial in either timing or poising of gene transcription once required during embryogenesis.

This thesis discusses several lines of research supporting these claims. Also an alternative hypothesis is brought up for the biological relevance of bivalency at specific genes. PRC2 and PRC1 colocalize at a subgroup of bivalent genes. Both complexes are involved in, seemingly, independent pathways leading to RNA Pol II (RNAP) stalling, hence preventing transcriptional elongation to commence. It is therefore that I suggest bivalency ensures a doubled effort in gene repression during crucial developmental stages.

Elucidating the function of bivalent domains will help us to understand their role in regulating embryogenesis. Surprisingly, Drosophila, like other non-vertebrates, lacks bivalent chromatin domains, even though homologous target genes, individual chromatin marks and the necessary machineries involved are al present. Hence, it remains to be clarified what role bivalent domains have in vertebrate embryogenesis and eventually how non-vertebrates manage without them.

Table of content

Introduction

All zygotes face the challenge to grow and differentiate into a fully functional multicellular organism. The tools to do so are present from the very start and are passed on from the parental gametes into the newly formed zygote. Embryonic stem cells (ESCs) need to develop into properly formed tissues and organs and the genes involved to do so are regulated in different fashions. Besides extra- and intracellular cues that may directly influence gene expression, it became clear that cellular memory of gene-expression states transmitted mitotically and sometimes even meiotically, plays an important role in development. Transmission of such cellular states was found to mostly rely on modifications of chromatin. The chromatin core is formed by the nucleosome which comprises ~147 bp of DNA wrapped around a histone octamer (2x H4, 2x H3, 2x H2B and two H2As). Chemical additions of methyl groups to either DNA or histone proteins are now recognized as major player in transmittable gene regulation and errors in these systems often lie at the base of many diseases such as cancer and autism. Because cellular memory systems are heritable but do not affect the underlying genetic code they are referred to as epigenetic mechanisms.

This thesis will focus on several chromatin modifications and their presumed mode of action to control gene expression. The goal of this thesis is to highlight the biological relevance of epigenetic regulation through the gene repressive mark tri-methylation of histone 3 lysine residue 27 (H3K27me3) in general and H3K27me3 combined with the gene activating mark H3K4me3 in particular.

When scientist discovered chromatin modifications, early-obtained data suggested they were part of an on/off type gene regulatory mechanism. Recent data, however, has started to reveal a much more sophisticated and complex interplay between both repressive and activating marks suggesting genes expression can be precisely fine-tuned in the very dynamic context of development.

H3K27 methyl groups are deposited by Polycomb Repressive Complex 2 (PRC2)[1]. However fascinating, a broader perceptive is required to fully understand the impact of epigenetic regulation. Therefore H3K27me3 will be placed among other chromatin marks, considering a joint effort of these marks to regulated genes in a sophisticated fashion. Ultimately this will have consequences on how we view chromatin and the different states it may have. Chromatin was thought to consist of active chromatin (euchromatin) and inactive chromatin (heterochromatin). Over time more and more modifications have been identified. Also combinations of different modifications have been found. Like the previously mentioned H3K27me3 and H3K4me3. Due to their antagonistic effects on gene regulation together they create a bivalent chromatin state. This thesis will try to cover the impact of recent discoveries by discussing the biological relevance of gene regulation mediated by bivalent marked chromatin domains during vertebrate development.

The Composition of PRC2

Early work has primarily been done in *Drosophila melangaster*, but it certainly has not been the only model organism that produced interesting insights. Epigenetic regulation is quite conserved among eukaryotes, however certainly not all combinations of marks have the same regulatory purpose[2]. During this thesis the focus will be on the general principles, rather than species specific differences. However those differences can clarify specific mechanisms and are therefore important to mention on occasion.

As stated, pioneering work has been done in *Drosophila*. The polycomb mutant demonstrated a phenotype similar to mutants of developmentally important Hox genes, with male mutants showing extra sex combs. Further research showed that the disrupted gene coded for a regulator of Hox genes[3,4].

The PcG (Polycomb Group) components are part of several complexes. The before mentioned PRC2 is the most important to discus for this thesis. PRC2 contains 4 core subunits and these are conserved over most eukaryotic life. The exceptions are (several) yeasts of which it is speculated that they have lost polycomb genes during evolution[2,5]. *In vitro* experiments showed that these 4 proteins are required for methylation of at least H3K27me2/3. E(z) is the enzymatic sub-unit in Drosophila able to di- and tri-methylate H3K27me1[1]. Varying per organism, a number of co-factors can accompany the core complex. Table 1 and figure 1 show the 3 best studied co-factors. The role of these co-factors is generally speaking thought to be to recruit the polycomb (sub) units or to specify binding on target DNA. However individual mechanisms of these co-factors are still subjected to debate. Beside these three, a number of other proteins (DNMTs, HDAC1 and SIRT1) are reported to interact with PRC2[2]. Most of these are also involved in other protein complexes as well and their function still needs to be confirmed. The general concept is that PRC2 acts as a delicate system that has numerous manners of fine tuning its function.

It is known that in *Drosophila* PRC2 associates with so called Polycomb response elements in the DNA sequence. These elements contain a number of transcription factor (TF) binding sites that are supposed to act as a combinatorial code for PRC2 recruitment and binding[2,4,6]. In mammals no such combination of TF binding sites have been found. There are studies that show a high density of cytosine and guanine nucleotides at PRC2 docking sites, however, the importance of this has not been clarified yet[7]. How the selection of target genes in specific cells takes place is still up for debate. It is not unlikely to consider the known co-factors (and new to discover factors) to play a role in target recognition.

PRC1

Other polycomb complexes interact with PRC2 and have an important role as well. However these are often not as conserved as PRC2. For example PRC1 homologues are present in Drosophila and mice, but there are only suspected homologues in plants and *C. elegans*. Also the composition of PRC1 varies between species, leaving its function subject of debate. Nevertheless it has been proven that PRC1 in combination with PRC2 is required for maintaining the silent state of a subset of PRC2 target genes. This occurs by PRC1 binding to H3K27me3 which will result in further physical compaction of chromatin. Recently a more localized and specific role for PRC1 was proposed in controlling bivalent gene activity. As such PRC1 has become important to understand possible mechanism in marking bivalent genes and the accompanying staling of RNA Pol II (RNAP). This is thought to occur through ubiquitylation of H2A(z) (see *RNA Pol II stalling* for more details). PRC1, however, does not seem to have an effect on the establishment of the bivalent marks[8,9].

Initial findings suggested that PRC1 acted downstream of PRC2. New findings contradicted this and it is now thought that both complexes can act independently as well as collectively depending on the context. In cases that they do co-localize at the chromatin it is at certain bivalent genes[7].

	Drosophila	*Human*
	PRC2	
Core units	E(z)	EZH1/2
	ESC/ESCL	EED
	SUZ12	SUZ12
	NURF55	RbAp46/48
Sub units	JARID2	JARID2
	Jing	AEBP2
	PCL	PCL

Table . Drosophila and animal PRC2 homologs. Since the references in this thesis consist of papers researching both Drosophila and human PRC2 complexes, it is inevitable that both names are used.

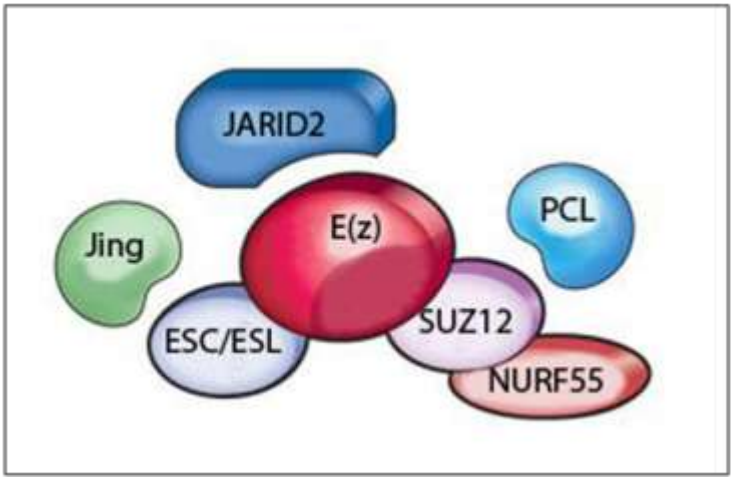

Figure 1. Impression of the four core units of PRC2 in Drosopila. E(z) is enzymatic active component of PRC2. ESC/ESL, SUZ12 and NURF55 are the other core units. JARID2, Jing and PCL are all confirmed to interact with PRC2

Mode of Action of PRC2

The general consensus is that di- and tri-methylation of H3K27 has a negative regulatory effect on gene expression. Interestingly it is still not known how this effect is achieved. The antagonistic methylation of H3K4, by Trithorax Group (TrxG) complexes, has been shown to be similarly elusive. Recognition of the methyl group has been shown, but not how this is translated into gene regulation[10]. Over the years it has been proven to be difficult to elucidate a precise mode of action. This paragraph will discuss a few of the main hypotheses.

One of the early hypotheses was that methyl groups would cause a conformational shift of either the histones and/or the chromatin, reducing the accessibility of DNA for transcription factors. The methyl group would cause physical restrains to the availability of the target genes. Histone acetylation and deacetylation, which act independently of polycomb-mediated gene regulation, have been proven to cause such physical alterations of the nucleosome. The removal of acetyl groups, for instance, redistributes the electrical charge at the nucleosomal surface, tightening the winding of DNA around the histone octamer and causing DNA to be less accessible for DNA binding factors[11]. No evidence for such a straight forward conformational change has been found for neither H3K27 nor H3K4 methylation. Literature does not report any conformational change of chromatin or chromatin proteins upon methylation. However, preventing DNA access by steric hindrance due to H3K27me is still a valid hypothesis.

A second problem with the conformational change hypothesis is that it disregards all (similar) modifications on nearby residues. More and more experimental data suggest that on a single stretch of chromatin several marks can be placed. Even marks that often are thought to be antagonistic (such as H3K4me3 and H3K27me3) can cooccupy the same nucleosomal array. Opposing regulatory labels can even coexist on the same nucleosome, a characteristic of bivalent domains[12]. Interestingly, experiments show that enzymatic activity of histone demethylases targeting H3K27me2/3 (PHF8 and KIAA1718) may associate with bivalent domains[10,13]. These demethylases contain histone binding sites that are able to bind both H3K4me2/3- and H327me2/3 upon which their activity is greatly increased[13].

This all combined suggests that not a conformational change of the nucleosome is the effect of methylation, but rather the formation of a combinatorial recognition key for the appropriate (transcription) factor and accompanying enzymatic activities may be the prevailing function of bivalency. The paragraph *Possible mechanisms of Bivalent Domains* will discuss the proposed mechanisms of epigenetic mark recognition and gene regulation through these marks in more detail.

RNA Pol II staling

Several efforts have led to hypotheses that gene repression is established –at least partially- by stalling RNAP upon which elongation is inhibited. RNAP phosphorylation of Serine 5 is associated with the initiation of transcription, upon which Serine 2 is phosphorylated to proceed into the elongation phase[14]. Stalled RNAPs lack the phosphorylation of Ser-2 but maintain Ser-5 phosphorylation. Commonly RING1, of the PRC1 complex, is considered to ubiquitinate H2A and thereby stall RNAP. H2Aub is thought to repress the ubiquitination of H2B. This H2Bub is necessary to start elongation of RNAP[15,16]. If PRC1 requires interaction with PRC2 to achieve H2Aub is yet to be elucidated. The mechanism behind RNAP stalling by H2Aub needss to be clarified as well. Cautious suggestions have been made towards a yet to be defined conformational change of RNAP holo-enzyme, changing the binding site for RNAP and affecting elongation[15].

Although a mechanism is not available, there is a certain consensus on the requirement of PRC2 and PRC1 at bivalent domains[7]. One of the hypotheses is that RNAP stalling is part of a regulatory concept called gene poising[17]. Poised genes would be primed for immediate activation by assembly of all the necessary complexes. A mark for these poised genes is thought to be found in bivalence of both active and repressive marks at the same nucleosomes. Research has shown that these bivalent genes do contain stalled RNAPs. However stalled RNAPs are certainly not exclusively linked to these genes. In fact, more genes contain stalled RNAPs[18]. Unfortunately it is unclear if these genes are solely PRC2 targets or not. This raises the question if RNAP blocking is the mode of action exclusive for PRC2 or if it is a common manner of gene regulation that PRC2 is able to recruit. Current studies point towards the latter.

Kanhere *et al.* (2010) recently reported the finding of short non-coding RNAs of 50 to 200 nt that associated with both RNAPs and PRC2. When so, there is no elongation of mRNA which resembles RNAP stalling[19]. More research is required to clarify the role of these short RNAs and their relation towards PRC2. However, RNAs form a promising lead towards clarifying specific gene targeting by PRC2. For instance it was found that long non-coding RNAs (ncRNA) used in X chromosome silencing are involved in recruitment of epigenetic machineries including PRC2[20]. Xi, a ncRNA transcribed from the X-inactivation centre of one X chromosome, localizes in cis to parts of the X chromosome transcribing Xi which is followed by depositioning of repressive marks, including PRC2 marks and silencing of the chromosome[20,21]. Kanhere and Jenner wrote a review in which they discuss the various hypotheses of an underlying mechanism. They summarize four manners in which ncRNA can influence expression. These vary between a direct interaction of the ncRNA with the DNA and a mediated interaction by RNAP or TFs. In all four hypotheses there are regulatory complexes associating with ncRNAs[22]. PRC2 is one of the

complexes that might fulfill such a function. However it leaves the question open if any of these are the common manner in which ncRNA silencing takes places at H3K27me3 enriched loci.

Vertebrate Genes marked Dually by H3K27me3 and H3K4me3

In studies analyzing genome wide histone modifications certain bivalent marked domains were discovered. These genes are marked with both H3K4me3 and H3K27me3, which are antagonistic marks. Bivalent genes are identified in vertebrates, but not in other model organisms like Drosophila and Xenopus[23,24]. From the beginning this has shed questions on the importance of bivalent chromatin marking.

Before discussing the value of bivalency, it should be clarified what it exactly constitutes. H3K4me3 and H3K27me3 rarely occupy the same stretch of chromatin. They only do coexist on the promoters of certain genes in vertebrate stem cell precursors. In fact, sequential chromatin immunoprecipitation proved that both marks are part of the same nucleosomes[25]. The marked genes are all involved in lineage specification and are transcriptionally inactive[12]. Thus, H3K27me3 seems to override H3K4me3. Most of the bivalent domains disappear when embryonic stem cells differentiate into specific cell types, becoming either active or (permanently) suppressed.

Given the temporal nature of bivalent domains and associated processes the underlying genes control, it is thought that bivalency serves a role in developmental regulation in stem cells. Since bivalent genes are only present in some species, it is interesting what this role might be.

Possible Mechanisms of Bivalent Domains

During development blastocyst cells are formed and become the precursors of embryonic stem cells. These blastocysts are the cells that have bivalent genes. They are pluripotent and the cells have yet to commit to a specific cell lineage. When this happens cells are often unable to change fates. The timing of this turning point seems crucial and many genes influencing differentiation are marked with H3K4me3 and H3K27me3.

Cell commitment occurs in steps with intermitting cell identities. With each consecutive cell type becoming more committed and losing pluripotency. The tri-methylation serves a crucial role in this process. Interestingly independent experiments of mutants of PRC2 show cells that are either held up in pluripotency or show early differentiation[26]. This means that ES show phenotypes that range from unviable cells to cells failing to complete the proper developmental path. A lack of silencing of pluripotency factors may cause this effect and at the same time, repression of developmental genes is released to early[27–29].

The main hypotheses of bivalent gene function consider the one tri-methylation to keep the other (temporally) in check. H3K27me3 was thought to keep lineage specific genes silenced until the appropriate time. In this way H3K27me3 would keep cells pluripotent and at the same time its removal results in immediate differentiation. In PRC2 mutants a subset of the bivalent genes lose H3K27me3 and become actively transcribed. However this did not disrupt the pluripotency of the ES cells. The cells did showed difficulty finishing their designated development path[26,27]. Another important remark is that in these experiments H3K27me3 is not just removed from bivalent genes, but also from ‘monovalent’ genes. It is known that important pluripotency factors like Oct4, Sox2 and Nanog are silenced with H3K27me3 when differentiation is started. At least it should be considered that this could have an effect on cell type specification. It has yet not been possible to study the effect of loss of tri-methylation on selected gene targets. Therefore other experiments must be designed. Since these genes are activated over time, perhaps inducible mutants can be of use (for further details see *Future Research*)

Interestingly not all bivalent genes are up regulated in PRC2 mutants[26,30]. This raises the question what other characteristics are involved. A comprehensive profile of the chromatin could emerge when bivalent genes are studied more in depth of their precise chromatin status. For instance it already has been shown that PRC2 and PRC1 interact but not in all cases[7](*more details on PRC1 and PRC2 is given in the next paragraph*). This indicates a more complex interaction and hints towards involvement of more factors. A clearer image of all factors could show a much more delicate key to regulation then solely the H3K4me3 and H3K27me3 marks.

The other consideration was that H3K4me3 would maintain the possibility of H3K27me3 marked genes to be rapidly activated upon differentiation onset. It would serve as a protection of the gene to become completely silenced. This could for instance be done by protecting the DNA from becoming methylated (DNA methylation is regarded the strongest form of epigenetic gene silencing). Evidence for this hypothesis has been scarce but a few studies indirectly point this way. Experiments reducing H3K4me3 have shown a deregulation of pluripotency. This was shown in two different fashions. *dpy-30* mutants lack a functional core unit of the H3K4 tri-methylation complex Trithorax Group (TrxG). *dpy-30 cells* show a decrease in differentiation potential[31]. This contradicts studies of another TrxG core unit, Wdr5, in which levels of pluripotency genes are reduced[32]. More research is required on both Dpy and Wdr5 to clarify their role in H3K4 tri-methylation regulation. The fact that they give apposing phenotypes should be studied in more detail. For instance the genes that they separately effect might well be involved in opposing processes.

It is quite likely that bivalent genes function in early embryogenesis by regulating differentiation and possibly pluripotency. However it has been proven hard to substantiate this claim. Until now, no experiments have been able to reduce tri-methylation on a local scale and target bivalent genes specifically. Although indirectly, experiments show a role in the maintenance of an ES cell population during development and a role in the timing of differentiation. Many have speculated bivalent genes are primed genes or poised genes. All indicating that bivalency ensures a readiness for changes in gene expression. Most of these claims are based on studies of mutants with impaired methylase complexes, causing a genome wide effect. This causes difficulty in interpreting the results on just a subset of the effected genes. Of course the changes in expression of bivalent genes in PcG mutants are a good indication of the role of PcG complexes. However it has remained hard to pinpoint an exact mechanism. The fact that only a subset of species has evolved bivalent marks and are still capable of maintaining ES cells and complete differentiation on time adds to the question of their importance. Perhaps scientists are still misguided by the outdated thought that H3K27 tri-methylation is irreversible. A possibility could be that H3K27me3 is a more transient mark that is removed unless otherwise indicated. If so compared with 'monovalent' genes, bivalent genes could be considered 'extra primed', given the enzymatic enhancement it gives to demethylases. As discussed in paragraph 'mode of action of PRC2', some demethylases show higher activity when they can bind to both H3K4- and K27me3[10,13]. It is unlikely that this is the sole purpose of bivalent genes as it has not been excluded that there are more chromatin modifications involved. Future research should focus on the possible occupancy of the bivalent mark by an inhibitor that protects H3K27 to be demethylated. Kanhere *et al.* (2010) proposed a similar necessity for a factor and considered PRC2 itself to be a candidate. They described a model in which short ncRNAs mediate stalling of RNAPs by PRC2, forming an inhibitory and protective complex. Without these short ncRNA no stable complex is formed and PRC2 dissociates from the chromatin[19,22]. Although this does not explain the role of bivalent marking, it does teach us something about the nature of dualistic methylation.

PRC1 and PRC2 in Vertebrates and Fly

It is an interesting question why only vertebrates have developed this system. Many have tried to claim that it is crucial for timing of developmental genes. Besides the absence of proof, it is also the question why other organisms, like Drosophila, seem to function perfectly without it. Assuming it is indeed a timing mechanism, it can be reasoned that Drosophila never evolved one but instead uses one that is very different.

Literature shows a number of factors that are associated with bivalent genes. These factors in turn are known to be involved in many regulatory processes. Although a direct link to bivalent genes is not proved, they are valuable clues. Even more interesting, these factors are known in Drosophila as well and also associated with key developmental genes, as they do in vertebrates[17,33,34]. The next paragraph will try to compare Drosophila factors with known vertebrate factors.

Ku *et al.* (2008) showed that of the bivalent genes 39% is occupied by both PRC2 and PRC1. Table 2 shows a summary of an analyses of the bivalent genes occupied by PRC1 by performing a chromatin immunoprecipitation for Ring 1B (a PRC1 subunit)[7]. Of the 39% genes that are Ring1B positive, 40.5% are involved in developmental processes and another 30.4% in regulation of gene expression. Both PRC1 and PRC2 are involved in repression, possibly both through RNAP stalling. Evidence so far indicates that the stalling by PRC1 is facilitated by it ubiquitination of H2A, while PRC2 is thought to work through short ncRNA[9,19]. These are two distinct pathways. It is easy to hypothesize that their combined effort leads to a doubled assurance in repressing crucial genes in embryogenesis until the appropriate time.

Drosophila seems to have a similar combination of factors recruited to their developmental genes[17,33,34]. Like in vertebrates PRC2 and PRC1 are present and RNAP is stalled at the promoter. This shows that at least in Drosophila bivalency is not required to assemble these factors in this configuration. It is possible that bivalency provides an extra stability to PRC1 and PRC2 binding. If bivalency does give extra stability to associated complexes, it would give it through a yet to be identified factor. However, another hypothesis would be that bivalency itself does facilitate this stability. In absence of bivalency, Drosophila would require a factor X to retain PRC1 and PRC2. In this scenario PCL proteins are interesting candidates. Drosophila ES cells respond strongly to the absence of PCL2, displaying similar phenotypes like PRC2 mutants. Pluripotency factors are up regulated and differentiation is disrupted. It already has been shown that PCL2 associates with various PcG components[35].

Ultimately the ancestors of organisms with bivalent genes were fitter then their peers were without. On the one side it assures that bivalent genes have a function, otherwise it would have been lost during evolution. On the other hand currently prevailing hypotheses of poising or stalling might be difficult to prove in a laboratory environment. Perhaps the advantage it gives is only present in wild type conditions.

Table 1. Table is taken from Ku *et al* (2008).

GO term	% of genes	Ratio	P-value
Ring1B+			
developmental process	40.5%	2.4	$p<10^{-20}$
regulation of gene expression	30.4%	2.2	$p<10^{-20}$
cell-cell signaling	7.7%	3.8	$p<10^{-20}$
embryonic morphogenesis	5.4%	5.0	$p<10^{-20}$
Ring1B-			
membrane	50.1%	1.2	$p<10^{-10}$
cell adhesion	7.9%	2.3	$p<10^{-5}$
transporter activity	12.3%	1.8	$p<10^{-5}$

Future Research

A lot of the polycomb proteins are known for quite some time and their functions are at least partially determined. However at the moment critical questions remain and scientists are limited by technical possibilities to answer them. This thesis tried to concretize this by discussing our knowledge of bivalent genes. In this paragraph a summery will be given of the remaining questions and suggest possible lines of inquiry.

The last few years high resolution genomic maps have been obtained for many of the known chromatin marks. Polycomb mutants showed specific target genes and the effects of loss of function on the genomic methylation profile shown with great precision. At the same time this points to the weakness of polycomb research. Due to the genomic wide effect it is hard to differentiate between primary and secondary responses to the absence of H3K27 tri-methylation.

One of the ways to (partially) circumvent a general effect is to use inducible mutants. When the mutant is induced at a set time in cell line specification this could show the effect of polycomb proteins in specific time frames. The mutants will still have a genomic wide effect, but it is very possible that genes already regulated in ES cells will not disrupt gene regulation in progeny cells.

Key examples are bivalent genes. The combination of the two antagonistic marks H3K4me3 and H3K27me3 has resulted in many hypotheses, of which none has been definitively proven. The problem of targeting just the bivalent genes and not other genes with tri-methylation is yet to be resolved. More than any other bivalent genes are claimed to be highly time sensitive[30]. Up till now, no studies have been made public about an attempt to influence them *in time*. Besides inducing Polycomb mutants, it is relatively easy to retain polycomb components in the cytosol using specific phusion proteins. They can be released into the nucleus at specific times by chemical interference and hence create a temporal complementation of the mutant.

At the moment a lot of research is done on ES cell cultures. A common method is to compare ES cell to fully differentiated cells. Often ES cells are induced with a cocktail of differentiation factors to push to a desired cell type in high speed. This leaves a lot of questions of intermediate cell types. A more controlled culture procedure combined with cell sorting could elucidate a more temporal role of tri-methylation.

Finally this thesis suggested that bivalent marks could be more of a stabilization factor for both PRC1 and PRC2, possibly with PCL proteins. The best way to test this hypothesis is to determine the structural interactions of these proteins. Research in that direction needs to be carefully planned as the success of these experiments is often related to the size and number of proteins.

Conclusion

In time, scientist moved away from a simple on/off system by H3K27me3 silencing its targets and H3K4me3 is activating its targets, to a more complex picture. The fact that they can coexist on the same nucleosome indicates a more sophisticated mechanism than previously anticipated[25]. This thesis has reviewed some of the factors involved in establishing and controlling bivalently marked genes[12]. One of the most important clues of bivalent genes is their function. There are several distinct regulatory mechanisms involved in controlling polycomb target genes even within the bivalent genes pool [7]. How this regulation is achieved, is the biggest question yet to be answered. Many have suggested timing or poising mechanism[30]. This thesis gave another hypothesis. Bivalent marks are likely not acting on their own accord. Therefore I suggest that developmental genes are protected through a double mechanism by PRC2 and PRC1. Unfortunately the only evidence is for this comes from their co-localization on bivalent genes[7]. It is easy to speculate on recruitment and stabilization of PRC1 and PRC2, but no evidence for this has been found so far.

The difficulties of polycomb research have been discussed as well. Also possible lines of inquiry have been given. However, reports given in literature seem to point in avoidance of these issues. This leaves a feeling of some scientist focusing on the big, chromatin wide picture, without discussing the (biological) relevance of their findings. Others focus on the role of a single complex or even a subunit with disregard of possible effects on the chromatin as a whole. It is understandable that some of these 'blind spots' originated in technical limitations (as discussed before). This should not be an excuse to pursue this line of investigation. For instance a lot can be gained of combining databases by bio-informaticians. It could be a time consuming and tedious effort, but the results would be invaluable.

References

1. Müller, J. *et al.* Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* **111**, 197-208 (2002).

2. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-9 (2011).

3. Lewis, E.B. A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565-570 (1978).

4. Schuettengruber, B. & Cavalli, G. Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development (Cambridge, England)* **136**, 3531-42 (2009).

5. Shaver, S., Casas-Mollano, J.A., Cerny, R.L. & Cerutti, H. Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga Chlamydomonas. *Epigenetics : official journal of the DNA Methylation Society* **5**, 301-12 (2010).

6. Simon, J. a & Kingston, R.E. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nature reviews. Molecular cell biology* **10**, 697-708 (2009).

7. Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS genetics* **4**, e1000242 (2008).

8. Hengbin Wang, Liangjun Wang, Hediye Erdjument-Bromage, Miguel Vidal, Paul Tempst, Richard S. Jones, Y.Z. Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 862-8 (2004).

9. de Napoles, M. *et al.* Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Developmental cell* **7**, 663-76 (2004).

10. Khorasanizadeh, S. Recognition of methylated histones: new twists and variations. *Current opinion in structural biology* **21**, 744-9 (2011).

11. Shogren-Knaak, M., Ishii, H. & Sun, J. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **16**, 844-847 (2006).

12. Bernstein, B.E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-26 (2006).

13. Justin, N., De Marco, V., Aasland, R. & Gamblin, S.J. Reading, writing and editing methylated lysines on histone tails: new insights from recent structural studies. *Current opinion in structural biology* **20**, 730-8 (2010).

14. Spilianakis, C. *et al.* CIITA regulates transcription onset viaSer5-phosphorylation of RNA Pol II. *The EMBO journal* **22**, 5125-36 (2003).

15. Stock, J.K. *et al.* Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nature cell biology* **9**, 1428-35 (2007).

16. Fuda, N.J., Ardehali, M.B. & Lis, J.T. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**, 186-92 (2009).

17. Zeitlinger, J. *et al.* RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nature genetics* **39**, 1512-6 (2007).

18. Min, I.M. *et al.* Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes & development* **25**, 742-54 (2011).

19. Kanhere, A. *et al.* Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Molecular cell* **38**, 675-88 (2010).

20. Senner, C.E. & Brockdorff, N. Xist gene regulation at the onset of X inactivation. *Current opinion in genetics & development* **19**, 122-6 (2009).

21. Zhao, J., Sun, B.K., Erwin, J. a, Song, J.-J. & Lee, J.T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science (New York, N.Y.)* **322**, 750-6 (2008).

22. Kanhere, A. & Jenner, R.G. Noncoding RNA localisation mechanisms in chromatin regulation. *Silence* **3**, 2 (2012).

23. Schuettengruber, B. *et al.* Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. *PLoS biology* **7**, e13 (2009).

24. Akkers, R.C. *et al.* A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in Xenopus embryos. *Developmental cell* **17**, 425-34 (2009).

25. Pan, G. *et al.* Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell stem cell* **1**, 299-312 (2007).

26. Chamberlain, S.J., Yee, D. & Magnuson, T. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem cells (Dayton, Ohio)* **26**, 1496-505 (2008).

27. Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M. & Helin, K. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Molecular and cellular biology* **27**, 3769-79 (2007).

28. Leeb, M. *et al.* Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes & development* **24**, 265-76 (2010).

29. Shen, X. *et al.* EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Molecular cell* **32**, 491-502 (2008).

30. Vastenhouw, N.L. & Schier, A.F. Bivalent histone modifications in early embryogenesis. *Current opinion in cell biology* 1-13 (2012).doi:10.1016/j.ceb.2012.03.009

31. Jiang, H. *et al.* Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. *Cell* **144**, 513-25 (2011).

32. Ang, Y.-S. *et al.* Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145**, 183-97 (2011).

33. Hendrix, D. a, Hong, J.-W., Zeitlinger, J., Rokhsar, D.S. & Levine, M.S. Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 7762-7 (2008).

34. Chopra, V.S., Cande, J., Hong, J.-W. & Levine, M. Stalled Hox promoters as chromosomal boundaries. *Genes & development* **23**, 1505-9 (2009).

35. Walker, E., Manias, J.L., Chang, W.Y. & Stanford, W.L. PCL2 modulates gene regulatory networks controlling self-renewal and commitment in embryonic stem cells. *Cell Cycle* **10**, 45-51 (2011).