# The analysis of PWWP INTERACTOR OF POLYCOMBS 1 (PWO1), a putative link between Polycomb-mediated repression and nuclear periphery

## Evolutionary studies of Polycomb-mediated repression in the unicellular algae

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy

of Freie Universität Berlin

by

PAWEL SLAWOMIR MIKULSKI

from Warsaw, Poland

Berlin, 2017

The investigations described in the following thesis were performed under supervision of Prof. Dr. Daniel Schubert in the group of Plant Epigenetics at the Institute of Biology, Free University Berlin (11/2015 – 02/2017) and at the Institute of Biology, Heinrich Heine University Düsseldorf (03/2013 – 11/2015).

1st Reviewer: Prof. Dr. Daniel Schubert

2nd Reviewer: Prof. Dr. Thomas Schmülling

date of defence: 31.03.2017

## Acknowledgements

Firstly, I would like to thank Prof. Dr. Daniel Schubert for the constant support, scientific discussions and creating family-like atmosphere in the group. It has been enjoyable and motivating couple of years to pursue this PhD project.

I would like to thank also Prof. Dr. Thomas Schmülling for accepting the role of the second supervisor of my thesis and all the work connected with its review.

Furthermore, I want to stress that participation in the 'EpiTRAITS' Marie Curie ITN was a great time and really memorable experience. It allowed to broaden my scientific interest and bond with terrific people. I would like to thank my 'EpiTRAITS' advisors, Dr. Manuel Piñeiro and Prof. Dr. Jan Lohmann for the supervision in the consortium.

In the acknowledgements, I cannot forget about my colleagues from the biology institutes at HHU Düsseldorf and FU Berlin, 'EpiTRAITS', as well as former and current AG Schubert members. I want to thank you all for the private and scientific discussions, as well as sharing fun outside of work. My special thanks go to the people who accompanied me in this PhD adventure for the longest time – whole group of 'EpiTRAITS' fellows and Inês, Jenny, Julia, Yu, Olga, Ralf, Sara.

Last but not the least, I would like to thank my friends and family for being irreplaceable listeners and supporters. I especially acknowledge my parents. Their help and values show a perfect example that, in long-standing debate, 'nature and nurture' can come from the same source. ;)

## Table of contents

# 1. Introduction

## 1.1. Complex nature of the chromatin

Genetic information is encoded by DNA which is organized and compacted in the nucleus by proteins termed histones. Such protein-DNA complexes form, so called, chromatin and have roles in the plethora of processes including transcription, DNA repair and cell division. A basic structural unit of chromatin is the nucleosome, a ~ 147bp-long DNA stretch wrapped around an octamer of histone proteins (Luger et al., 1997). Each histone octamer composes oftwo copies of histones: H2A, H2B, H3 and H4. The nucleosomes are arranged in repetitive manner forming 'beads-on-a-string' fibre, the lowest-order chromatin conformation (Finch et al., 1977). In addition, a 20-60bp-long inter-nucleosomal DNA (linker DNA) is bound by a linker histone H1, which facilitates compaction into higher-order structures (Ramakrishnan, 1997).

In the simplest distinction, chromatin can be grouped into two cytologically distinguishable types, eu- and heterochromatin. Euchromatin is an accessible, gene-rich form that correlates with higher transcriptional activity. In contrast, heterochromatin describes a condensed, gene-poor, transcriptionally silent type. Both chromatin types are associated with certain chromatin modifications, including DNA methylation, histone variants and histone marks. A closer look at combinatorial pattern of such modifications, underlying genomic targets and their transcriptional activity revealed further complexity of the chromatin and presence of distinct subtypes (chromatin states) (Baker, 2011).

Among chromatin modifications, the histone marks offer one of the most robust ways to delineate chromatin states and regulate genomic targets. Histone marks are covalent post-translational modifications that predominantly target N-terminal tails of core histone proteins. A plethora of histone marks was discovered and includes: lysine/arginine methylation, lysine acetylation, lysine ubiquitination and serine/threonine phosphorylation. Moreover, some modifications provide an additional complexity, exemplified by the fact that a single lysine residue can be a target of, functionally distinct, mono-, di- or trimethylation (Barski et al., 2007).

Histone marks influence chromatin in direct and/or indirect way. The former is related to the change of sterical features of the nucleosome as seen for, i.e. lysine acetylation. This

modification removes positive charge of the histones and decreases their interaction with negatively charged DNA, which results in more relaxed, transcriptionally permissible chromatin type. In contrast, the indirect way does not necessarily influence the charge within the nucleosome, but is rather based on the recognition of the mark by downstream proteins, which remodel nucleosomes or help to recruit other factors. Regulators of histone marks can be therefore arranged into: writers, which catalyze a modification; readers, responsible for its recognition, and erasers that chemically remove it. The interplay between writers, readers and erasers provides a dynamic modulation of the chromatin.

In contrast to the stability of DNA nucleotide sequence and low frequency of spontaneous DNA mutations, dynamic nature of histone marks offers bigger flexibility in regulating genomic targets in response to changing developmental and environmental cues. As histone marks can be faithfully propagated through the cell divisions (Probst et al., 2009), they can be considered as mean of epigenetic information, which describes heritable changes in genomic activity outside of the DNA sequence. Consequently, histone marks determine the establishment of transcriptional memory crucial for maintaining cell identity in such processes like developmental transitions and stress response (Fang et al., 2014; Kleinmanns and Schubert, 2014; Spivakov and Fisher, 2007).

### 1.2. H3K27me3 histone mark and Polycomb group (PcG) proteins

Trimethylation of lysine-27 on histone H3 (H3K27me3) is one of the most important histone modifications correlated with transcriptional repression (Boyer et al., 2006; Ernst et al., 2011). It is highly conserved among eukaryotes and is present at substantial number of genes (~12% in human (Bracken et al., 2006; Kalushkova et al., 2010), ~20% in *Drosophila* (Ringrose, 2007), 13-25% in *Arabidopsis* (Bouyer et al., 2011; Oh et al., 2008; Zhang et al., 2007)), most of which are developmentally- and stress-regulated. H3K27me3 can be inherited through mitotic and meiotic divisions (Hansen et al., 2008), and therefore provides a transcriptional memory in the next generation of cells.

The writer of H3K27me3 is POLYCOMB REPRESSIVE COMPLEX 2 (PRC2), one of the protein complexes encoded by Polycomb group (PcG) genes, discovered initially in *Drosophila* to control homeotic transformations during embryonic segmentation (Moazed and O'Farrell, 1992). In *Drosophila*, a canonical PRC2 contains: Enhancer of zeste (E(z)), a

catalytic component needed for in vitro and in vivo methylation of H3K27; Extra Sex Combs (ESC), a WD40 motif-containing protein that scaffolds interactions within the complex; Suppressor of zeste (Su(z)12), a Zinc Finger subunit essential for binding to nucleosomes and p55, a nucleosome remodelling factor (reviewed in: (Schwartz and Pirrotta, 2008)) (Fig. A1). In plants, an extensive gene duplication led to frequent formation of small gene families homologous to core PRC2 components from *Drosophila*. Thus, PRC2 in the model plant, *Arabidopsis thaliana*, consists of: one of three E(z) homologs - CURLY LEAF (CLF), SWINGER (SWN) or MEDEA (MEA); one of three Su(z)12-homologs – EMBRYONIC FLOWER 2 (EMF2), VERNALIZATION 2 (VRN2) or FERTILIZATION INDEPENDENT SEED 2 (FIS2); single ESC homolog – FERTILIZATION INDEPENDENT ENDOSPERM (FIE) and a p55 homolog – MULTICOPY SUPPRESSOR OF IRA 1 (MSI1) (Derkacheva and Hennig, 2014) (Fig.A1).

PRC2 can perform its functions dependently on the other PcG complex, termed POLYCOMB REPRESSIVE COMPLEX 1 (PRC1). PRC1 catalyzes monoubiquitination of lysine-118/119 on histone H2A (H2AK118ub/H2AK119ub) to facilitate transcriptional silencing (reviewed in: (Schwartz and Pirrotta, 2013)). PRC1 in *Drosophila* consists of: dRING1/Sex Combs Extra (Sce) and Posterior Sex Combs (Psc), both responsible for catalytic activity; Polyhomeotic (Ph), essential for maintaining protein-protein interactions; Polycomb (Pc), involved in a recruitment to the chromatin; and Sex comb on midleg (Scm), important for spreading of PcG silencing (reviewed in: (Schwartz and Pirrotta, 2013)) (Fig.A1). In *Arabidopsis*, the following PRC1 components were identified: AtRING1a/b (equivalent to dRING/Sce), AtBMI1a/b/c and EMF1 (both equivalent to Psc), and LHP1 (equivalent to Pc) (reviewed in: (Molitor and Shen, 2013)) (Fig.A1).

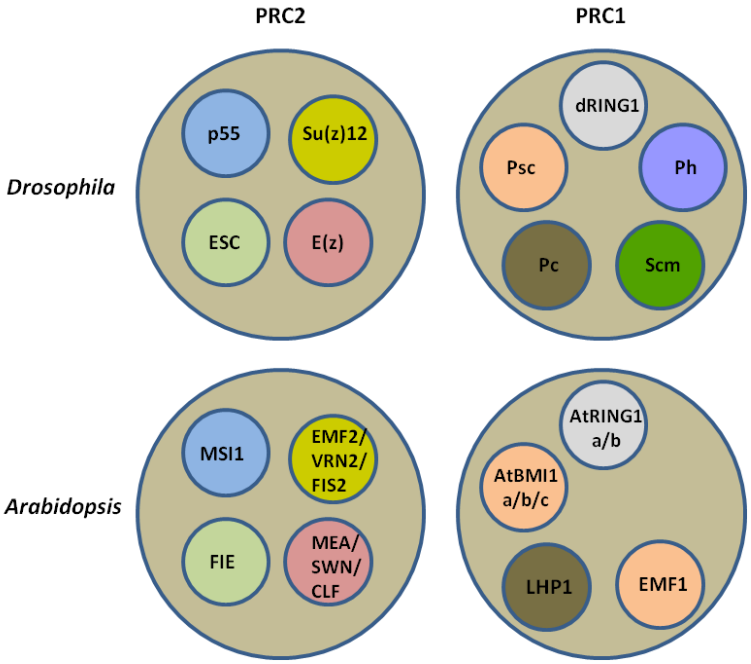



**Fig.A1. The canonical PcG protein complexes in *Drosophila* and *Arabidopsis*.** The colours correspond to the protein homologs between complexes from *Arabidopsis* and *Drosophila*.

Since the initial discovery, PcG proteins and their ability to modify histones gathered considerable attention that led to progressive characterization of their function and mode of action. Investigation of PcG-mediated chromatin processes revealed fascinating complexity in the interplay between PcG proteins and multi-layer mechanism of repression, including the subnuclear chromatin compartmentalization and the formation of higher order chromatin structures. The current state of knowledge in those fields was reviewed in: (Del Prete et al., 2015), attached in the following chapter.

**1.3. PcG-mediated mechanisms of gene repression**

*Genes* **2015**, *6*, 520-542; doi:10.3390/genes6030520

OPEN ACCESS

*genes*

ISSN 2073-4425

www.mdpi.com/journal/genes

*Review*

# One, Two, Three: Polycomb Proteins Hit All Dimensions of Gene Regulation

**Stefania del Prete [1,‡], Pawel Mikulski [2,‡], Daniel Schubert [2,†] and Valérie Gaudin [1,*]**

[1] INRA, UMR1318-AgroParisTech, Institut Jean-Pierre Bourgin (IJPB), INRA-Centre de Versailles-Grignon, Route de St-Cyr, Versailles Cedex F-78026, France; E-Mail: stefania.del-prete@versailles.inra.fr

[2] Institute for Genetics, HHU Duesseldorf, Universitätsstraße 1, Duesseldorf D-40225, Germany; E-Mails: Pawel.Mikulski@uni-duesseldorf.de (P.M.); Dan.Schubert@fu-berlin.de (D.S.)

† Current address: Institute for Biology, Freie Universität Berlin, Königin-Luise-Strasse 12-16, Berlin D-14195, Germany.

‡ These authors contributed equally to this work.

* Author to whom correspondence should be addressed; E-Mail: valerie.gaudin@versailles.inra.fr.

Academic Editor: Jessica Tyler

*Received: 11 May 2015 / Accepted: 30 June 2015 / Published: 10 July 2015*

---

**Abstract:** Polycomb group (PcG) proteins contribute to the formation and maintenance of a specific repressive chromatin state that prevents the expression of genes in a particular space and time. Polycomb repressive complexes (PRCs) consist of several PcG proteins with specific regulatory or catalytic properties. PRCs are recruited to thousands of target genes, and various recruitment factors, including DNA-binding proteins and non-coding RNAs, are involved in the targeting. PcG proteins contribute to a multitude of biological processes by altering chromatin features at different scales. PcG proteins mediate both biochemical modifications of histone tails and biophysical modifications (e.g., chromatin fiber compaction and three-dimensional (3D) chromatin conformation). Here, we review the role of PcG proteins in nuclear architecture, describing their impact on the structure of the chromatin fiber, on chromatin interactions, and on the spatial organization of the genome in nuclei. Although little is known about the role of plant PcG proteins in nuclear organization, much is known in the animal field, and we highlight similarities and differences in the roles of PcG proteins in 3D gene regulation in plants and animals.

**Keywords:** Polycomb; chromatin; *Arabidopsis thaliana*; three-dimensional nuclear architecture; Polycomb bodies; topologically associating domain (TAD); chromatin loops; lamins

## 1. Introduction

In 1947, P. Lewis [1] discovered the Polycomb (*Pc*) gene. While conducting genetic studies in *Drosophila melanogaster*, he found that the *Pc* mutation led to the formation of ectopic sex combs on the second and third legs of adult male flies. Later, while studying the segmentation patterns of Drosophila, it was shown that Pc is a key developmental regulator required to maintain homeotic gene (Hox) repression of the Bithorax complex (BX-C) [2]. Polycomb group (PcG) proteins were unified in a common family on the basis of their similar impact on development and their transcriptional repressive function, but they have various molecular activities [3–6]. Beyond their originally identified roles in controlling the sequence and timing of developmental switches and in maintaining cell and organ identity both in animals and plants [7,8], it is now clear that PcG proteins accomplish a variety of functions. For instance, recent evidence has implicated plant PcG proteins in abiotic and biotic stress responses [9,10]. During development, stress responses, and other processes, PcG proteins participate in a memory system and establish, maintain, and transmit silent epigenetic chromatin states. However, how PcG repression is established and transmitted *in vivo*, what mechanisms underlie repression, and how PcG activity is reset are only partially understood, even though intensive research efforts in diverse eukaryote models, from flies to plants, have attempted to answer these questions for more than sixty years.

During evolution, the genes encoding PcG proteins diversified from single copy genes to small gene families in animals, plants, and fungi [8]. PcG proteins form large and diverse complexes harboring catalytic subunits, which mediate post-translational modifications of the histone tails, and regulatory subunits, which regulate the transcriptional state of genes. Two main Polycomb repressive complexes (PRCs) each consisting of four core components were initially identified in Drosophila: Polycomb Repressive Complex (PRC) 1 and PRC2. The core components are generally conserved in plants and animals, but some kingdom-specific components evolved to meet the particular needs of the organism's life cycle (for detailed reviews: [8,11–13]). For instance, the plant LIKE HETEROCHROMATIN PROTEIN1 (LHP1) is a functional homolog of Pc, despite its structural homology with the HP1 protein family [14,15]. PcG not only biochemically modifies chromatin [16–18], but also induces biophysical changes in chromatin structure [19–22]. Furthermore, some PcG proteins can be part of other additional and unrelated complexes.

PcG function is antagonized by another group of proteins that modifies chromatin and regulates genes, the trithorax group (TrxG) proteins. The TrxG proteins activate gene expression and, like PcG proteins, are conserved in eukaryotes [23]. PcG and TrxG proteins often have the same target genes, and the activity of these genes is finely tuned by the opposing action of these two protein complexes.

Here, we present a brief overview of the molecular functions, recruitment, and regulation of PcG, with an emphasis on new research that highlights the roles of PcG proteins at different nuclear scales.

## 2. Hierarchy *vs.* Plasticity for More Flexibility in Repression?

### *2.1. Canonical Two-Step Mechanism of Repression*

PRC1 and PRC2 often occupy the same target locations and can act in a sequential manner, as proposed in the "canonical model" of PcG repression. The core PRC2 complex trimethylates histone H3 at lysine 27 (H3K27me3) at its target genes in a process that usually takes place on a genomic region called the nucleation site, and the deposited mark spreads over adjacent nucleosomes [24–26]. The deposition of H3K27me3 triggers binding of the PRC1 complex, which further ubiquitinates a lysine residue of histone H2A. These modifications result in other complementary factors being recruited, leading to inhibition of transcription and change in chromatin organization. The inhibition mechanism is elusive. On the other hand, TrxG complexes promote transcription, mostly by introducing the H3K4me3 and H3K36me3 activation marks. Remodeling of chromatin renders it more accessible to transcription factors and triggers transcriptional activation; however, this activation does not necessarily rely on the promotion of Polymerase II complex recruitment to target genes, but rather on transcriptional elongation from poised Polymerase II complexes [27,28]. The functions and hierarchical mode of action of PcG and TrxG proteins were extensively reviewed elsewhere [29–32].

Gene expression dynamics are not only regulated by the antagonistic roles of PcG and TrxG complexes, but also by the active removal of histone modifications deposited by these proteins. For instance, plant histone demethylases (HDMs), including JumonjiC (JmjC) domain containing proteins, counteract the action of PcG proteins in physiological processes, such as the flowering transition, shoot development, cell fate determination, or the circadian clock [33].

### *2.2. Non-Canonical Mechanisms of Repression*

For years, the canonical model was considered to describe the only mechanism that regulates the target genes of PcG complexes. Recent studies have uncoupled the functions of PRC1 and PRC2 and revisited the sequential action of the complexes, revealing more complex working mechanisms [17,18,34–38]. Furthermore, PRC1-like complexes, which differ in composition (absence of some of the core subunits present in PRC1 proteins, or the presence of additional subunits), associated activities, or repressive functions, have been reported in mammals, Drosophila, and plants.

In mammals, various additional components can be combined with the core PRC1 subunits to form a large and diverse PRC1 complex family [38–41]. These additional PRC1 subunits may coordinate other histone modifications and histone crosstalk to repress gene expression. For instance, H3K36 demethylases have been found to associate with PcG proteins in the animal BCOR complex, which regulates a subset of BCL6 target genes [35] or in the Drosophila dRING-associated factor (dRAF) complex [18]. In dRAF, the H3K36 demethylase KDM2 is also required for efficient H2A mono-ubiquitination [18], whereas in BCOR, KDM2B targets the complex to unmethylated CpG islands [42]. The dRAF complex is not the only PRC1-like complex present in Drosophila; others include the Pho-repressive complex (PhoRC) and Polycomb-repressive deubiquitinase complex (PR-DUB) [43–45]. Moreover, in mammals, examples of H3K27me3-independent PRC1 targeting or H2A mono-ubiquitination are emerging [17,36,46].

A recent study of the regulation of seed maturation genes in *A. thaliana* revealed that H2Aub and H3K27me3 are deposited on some PcG target genes independently [34]. Indeed, the *clf/swn* PRC2

mutant, which has reduced levels of H3K27me3 on several target genes, seems not to be affected in overall levels of H2Aub [34]. Furthermore, plant PRC1 is sometimes recruited before PRC2 [34,47–49]. Recent studies have shown that complex relationships exist between PRC1 and PRC2, and that some PRC complexes (*i.e.*, PR-DUB and dRAF) have the capacity to mediate histone modifications other than H3K27 trimethylation and H2A ubiquitination, thus tremendously expanding our understanding of PcG-mediated repression mechanisms.

## 3. Various Ways to Hunt for Targets

PcG targeting has been reported to rely on the presence of *cis*-regulatory elements, *trans*-acting components (DNA-binding proteins, transcription factors, scaffolding proteins, non-coding RNAs (ncRNAs)), or even structural properties of chromatin fiber [50–52].

In Drosophila, PRCs are recruited at Polycomb Response Elements (PREs), which are *cis*-regulatory DNA sequences of up to a few hundred base pairs long. PREs contain combinations of several diverse binding motifs for proteins, such as the DNA-binding PcG protein Pho or GAG factor (GAF), which work cooperatively. PREs can be located several tens of kilobases away from the promoter of the target genes, and their properties depend on the sequence context [45,50,53]. The exact combinations of DNA-binding sites and key regulatory elements that determine the recruitment of the PRC complexes remain to be identified.

The DNA-binding PcG protein Pho participates in the targeting of the Drosophila PhoRC complex, and the PhoRC complex may serve as a tethering platform for other PRC complexes at some genomic locations. However, the targeting mechanism for Drosophila PRC1 and PRC2 at PREs and the identity of putative PRE-DNA-binding candidates remain unknown [53]. In mammals, only a few PREs have been identified; however, 97% of PRC2 targets were shown to correspond to annotated CpG islands or similar CG-rich regions. However, the PREs lack a consensus motif [50,54].

In *A. thaliana*, the presence of PREs has not been confirmed. However, GAGA-motifs were recently shown to overlap with the binding sites of FIE, a PRC2 subunit, and these motifs seem to be necessary for H3K27me3 deposition [55]. GAGA-motifs have also been identified in the target genes of LEAFY (LFY), which are repressed by PRC2 [56]. Recently, it was shown that the PRC1-like subunit LHP1 interacts with the GAGA factor BPC6 and that this interaction is essential and sufficient to recruit LHP1 to DNA sequences that contain GAGA-motifs *in vitro* [57]. The repression of *LEAFY COTYLEDON2* (*LEC2*), a key regulator of *A. thaliana* seed development, also requires a negative *cis*-regulatory element, the Repressive LEC2 Element (RLE), which is located 150 bp upstream of the first codon and is associated with CT-rich elements. The RLE triggers H3K27me3 deposition and inhibits transcriptional activity [58]. Conserved Regulatory Elements are also present in the upstream region of the promoters of two *KNOX* genes (*BREVIPEDICELLUS* (*BP*) and *KNOTTED-LIKE FROM ARABIDOPSIS THALIANA2* (*KNAT2*)), which are required for proper organ formation. These sequences are targets of the AS1-AS2 complex, which mediates the repression of *KNOX* genes [59]. Subsequently, the AS1-AS2 complex was shown to interact with multiple core components of PRC2 to establish the repressed chromatin state at *KNOX* genes [60]. This was one of the first demonstrations of a plant PRC2 being recruited by specific DNA-binding proteins. Similarly, VAL1 is needed to recruit BMI1, a PRC1 subunit in *A. thaliana*, and set the repressive state of embryo-specific genes [34]. More evidence that PcG is recruited by transcription factors

came from a study of the MADS box protein, AGAMOUS (AG). AG binds two CArG boxes located 1 kb downstream of *WUSCHEL* (*WUS*), and is required for the regulation of H3K27me3 levels and repression of *WUS* expression [61]. The MADS box transcription factor, SHORT VEGETATIVE PHASE (SVP), and the GRAS transcription factor, SCARECROW (SCR), also physically interact with LHP1 and bind to specific LHP1 target loci, and thus could participate in the recruitment of PRC complexes and in the maintenance of silent chromatin states at their respective target loci [62,63].

Long non-coding RNAs (lncRNAs) have been shown to participate in numerous mechanisms that regulate gene expression, through interactions with various proteins, including histone-modifying molecules [64,65]. It is still a matter of debate whether lncRNAs participate in the recruitment of PcG complexes either in *cis* or in *trans* [66,67]. The first lncRNAs, identified based on their ability to tether PRC2 to target genes, were mammalian Xist and the lncRNA *HOTAIR* [68–70]. Since then, numerous studies have examined the interaction between lncRNAs and PRC in mammals, and some have revealed that PRC interacts with several hundred large intergenic ncRNAs [71].

In *A. thaliana*, the list of identified lncRNAs is growing rapidly, with the emergence of *in silico* and RNA-seq technology, but their functions remain poorly understood [72–74]. However, two lncRNAs, *COLD ASSISTED INTRONIC NONCODING RNA* (*COLDAIR*) and *COLD INDUCED LONG ANTISENSE INTERGENIC RNA* (*COOLAIR*), are well-characterized regulators of the key floral repressor *FLC* during vernalization [75,76]. *COLDAIR* is a sense-lncRNA transcribed from the vernalization response element (VRE) of the first intron of *FLC*. *COLDAIR* was shown to interact *in vitro* with the CLF PRC2 subunit. Transiently induced by the cold, with peak expression after 20 days of cold exposure, *COLDAIR* was proposed to play a role in recruiting PRC2 to stably silence *FLC* [76]. *COOLAIR* is a set of lncRNA antisense transcripts involved in the early, cold-dependent, and transient transcriptional silencing of *FLC*. Somehow, *COOLAIR* acts as an indirect recruiter of PRC2, but the precise mechanism behind this function is unclear [75,77]. Recently, the lncRNA *APOLO*, which is expressed in response to auxin, was shown to interact with LHP1 [78]. LHP1 also interacts *in vivo* with the RNA-binding protein LIF2, which thus may participate in the formation of a ribonucleoprotein complex involving LHP1. The specificity of the interaction with the RNA partners may thus rely on the three RNA recognition motifs of LIF2 [79]. Recently, specific and promiscuous interactions have been demonstrated in animals, and the strength of the interactions between PcG proteins and lncRNAs was found to depend on the length of the RNA [80]. Whether plant PRC-lncRNA interactions are specific, promiscuous, or non-specific *in vivo* remains to be further investigated. Furthermore, it remains to be determined whether the lncRNA interaction stabilizes PRC complexes with chromatin, participates in the targeting and recognition of the target by a sequence-specific mechanism, or is involved in other scaffolding mechanisms that coordinate different enzymatic activities merits further investigation.

It was recently proposed that PRC2 also recognizes and distinguishes between "open" and "dense" chromatin, with a preference for the latter. Indeed, PRC2 seems to be targeted to specific chromatin features or chromosomal structures [81]. Thus, despite much progress, many open questions remain about how PcG complexes are recruited to silence specific genes.

## 4. On the Path from 1D to 3D

### *4.1. Never Walk Alone—Polycomb and Chromatin Domains*

Initial discoveries of PcG targets at the gene-scale were soon followed by a genome-wide analysis of H3K27me3 occupancy and binding of several PcG-proteins [15,54,82,83]. As expected, a substantial overlap between PRC2-bound fragments and H3K27me3 profiles was found. A correlation between PRC1-binding and H3K27me3 occupancy was also reported, with some exceptions, suggesting that PRC1 has other functions.

Moreover, it was shown that, in plants, H3K27me3 is deposited predominantly on euchromatin, spans whole genes, and targets up to 20%–30% of all genes in *A. thaliana* [84–87] and 10% in human [82,88], stressing the importance of the Polycomb pathway in genome regulation. The H3K27me3 occupancy pattern was also compared to the profiles of various histone marks, genomic features of underlying sequences, protein binding distributions, and transcriptional activity of target genes to identify common patterns and predict gene expression status, resulting in the identification of so-called chromatin states or domains that include regions of similar characteristics. In the first such study in Drosophila, a Polycomb-repressed chromatin state was identified, in which H3K27me3-targeted genomic fragments were associated with genic regions, lower transcription, occupancy of other repressive marks, and binding regions of PcG proteins (e.g., H3K27me2, Pc, E(Z), PCL [89]), and inversely associated with the active marks H3K4me1/2/3, H3K36me3, and H3K27Ac [90]). Several studies distinguished bivalent domains, in which repressive marks are accompanied by active ones, which are inactive or poised for transcription [91]. One of the main differences between the studied species was that animal H3K27me3-occupied regions consisted of large blocks formed by adjacent regions of a similar epigenetic landscape, whereas the *A. thaliana* (epi-)genome seemed to be organized into small domains of different states interspersed with each other (apart from constitutive heterochromatin, which was spread over the pericentromeric region of the chromosome) [92]. Therefore, the epigenetic topography of the genomes has emerged and the development of high throughput sequencing has tremendously refined it (Figure 1).

### *4.2. PcG and Chromatin Fiber Packaging*

Chromatin compaction and formation of higher-order chromatin structures are proposed to participate in the repression by PcG, by reducing or interfering with DNA accessibility to the transcription machinery. Indeed, *in vitro* studies have shown that the Drosophila PRC1 complex and, more specifically, the Posterior sex combs (PSC) subunit, have the ability to compact nucleosomal arrays and inhibit the chromatin remodeling mediated by the SWI/SNF complex [93]. The intrinsically disordered *C*-terminal region of PSC changes the beads-on-a-string chromatin conformation into high-order chromatin structures [21]. Linker DNA and histone tails might also participate in chromatin compaction; however, Drosophila PRC1 directly interacts with nucleosomes and its activity seems to be independent of histone tail modifications [21]. The important role of PRC1 in the compaction of nucleosomal arrays is conserved across metazoans and plants [19,21,93,94], but is carried out by different PRC1 subunits, such as M33, a Polycomb homolog in mouse, or the PRC1 component EMBRYONIC FLOWER1 (EMF1) in *A. thaliana* [19,94]. Detailed studies of PSC and M33 revealed that one of the main characteristics of a "chromatin compactor" protein is the presence of highly positively charged domains [94]. Using the FISH technique, one study revealed

that the PRC1 Ring1B subunit is also involved in *in vivo* compaction and regulation of Hox loci in murine embryonic stem cells (ESCs). This compaction activity is independent of Ring1B histone H2Aub catalytic activity [95]. Therefore, increasing evidence suggests that PRC1 participates in higher-order chromatin structure organization and that PRC1 has packaging effects on chromatin, which may be sufficient to mediate repression, at least at some specific target genes.

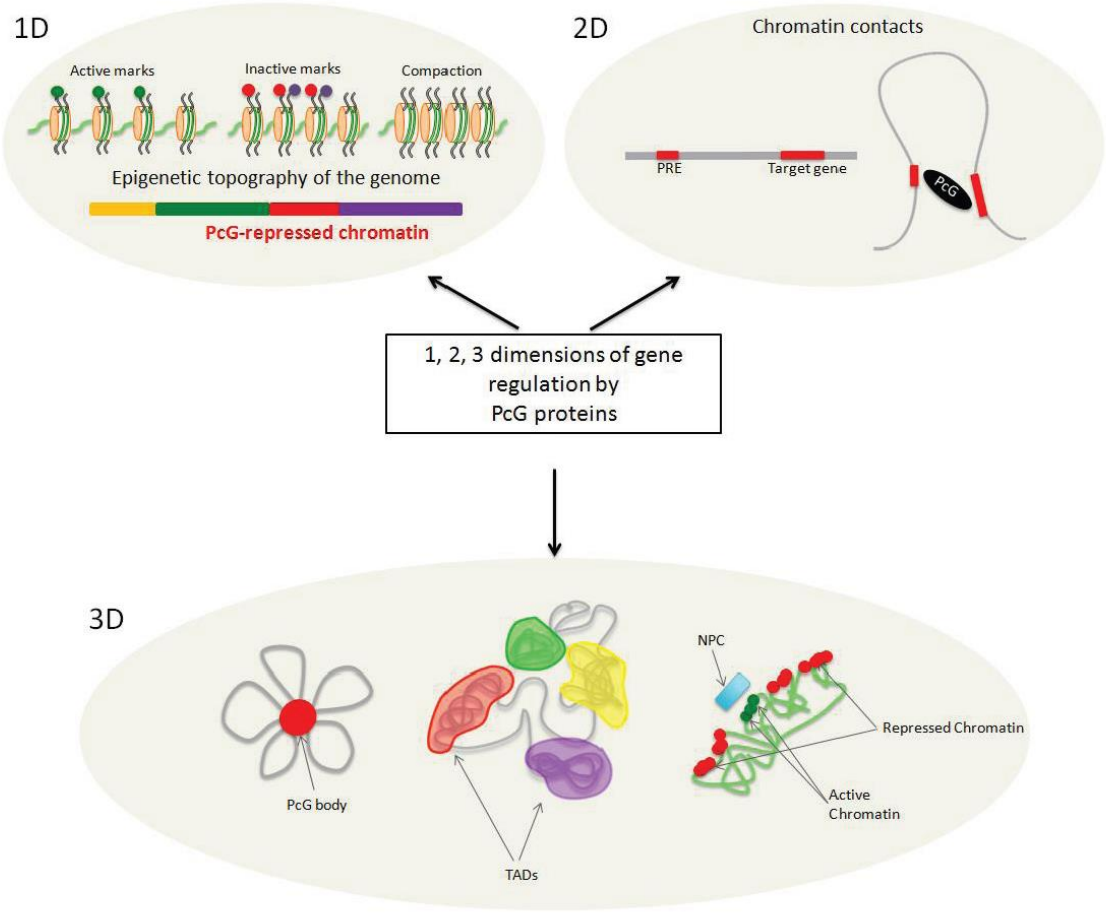



**Figure 1.** Polycomb Group (PcG) proteins affect chromatin regulation in three dimensions. PcG proteins participate in the establishment of the epigenetic topography of the genome by depositing biochemical modifications on histones and inducing chromatin compaction (**1D**). The presence of PcG-associated marks defines a specific chromatin state, called Polycomb-repressed chromatin. PcG proteins can mediate chromatin looping (**2D**). PcG complexes are recruited to *cis*-regulatory elements by various mechanisms (e.g., those involving Polycomb Response Elements (PREs), transcription factors, lncRNAs) to regulate their target genes and change the conformation of the chromatin fiber. PcG proteins act in the nuclear space (**3D**). They can aggregate with their targets to form Polycomb bodies. Chromatin is organized into distinct topologically associated domains (TADs), some of which are Polycomb-repressed chromatin domains. Spatial localization of chromatin at the nuclear periphery is correlated with repressive histone marks and, in the vicinity of the nuclear pore complexes (NPC), with active histone marks.

In support of a role of PcG in chromatin compaction, genome-wide studies demonstrated that H3K27me3-marked chromatin regions have low DNA accessibility both in Drosophila and *A. thaliana* [96,97]. In addition, chromatin in *lhp1* mutants is more sensitive to micrococcal nuclease digestion than is wild-type chromatin, suggesting a possible role for LHP1 in chromatin compaction. Although LHP1 is functionally similar to Pc, it is classified as a HP1 protein, based on its structure [14,15]. The chromo shadow domain (CSD) of the HP1 protein family is involved in homo-dimerization, whereas the chromo domain (CD) recognizes methylated lysine residues on histone tails. Interestingly, the binding of two CDs of Swi6, the *Schizosaccharomyces pombe* HP1 homolog, to nearby modified histone H3 nucleosomes creates a protein interface for tetramerization of two Swi6 homo-dimers connected by their CSDs [98]. The coordinated action of CD and CSD can thus lead to heterochromatin spreading throughout a stepwise oligomerization process from an auto-inhibited homodimer to chromatin-associated oligomers, which is somewhat analogous to the self-association process of tubulin dimers [99]. Whether this model, which is still controversial [100], is conserved for the HP1 family and plant LHP1 remains to be established.

Interestingly, it was recently shown that PRC2 has a higher activity on dense oligonucleosomes than on dispersed oligonucleosomes, suggesting that PRC2 can sense the chromatin environment and that its allosteric activation depends on the density of the substrate nucleosomes [81]. This study also revealed that local chromatin compaction precedes the establishment of H3K27me3 and provides a better substrate for PRC2 activity [81]. Therefore, PRC1 may initiate local chromatin compaction, which could then lead to the establishment and propagation of the H3K27me3 mark. This would support recent data suggesting that PRC1 acts before PRC2 [34,47,48,101]. However, the structural role of PRC complexes, at the local chromatin scale or at a higher order of chromatin organization, requires further investigation.

### *4.3. It Takes Two to Tango—Chromatin Loops Mediated by PcG Proteins*

Besides their roles in linear chromatin topography, increasing evidence suggests that PcG also affects the three-dimensional (3D) organization of chromatin at the nuclear level. Although the spatial genome regulation mediated by PcG proteins remains poorly understood in plants, we highlight recent findings, while taking existing evidence in other species into account.

PcG proteins can organize chromatin into 3D long-range loops, both in plants and in animals. These loops may also participate in PcG-mediated repression mechanisms (reviewed in [102]). Two techniques are used to investigate 3D chromatin organization, *i.e.*, fluorescence *in situ* hybridization (FISH), which allows direct visualization of the spatial organization of genomic sequences by cytology and microscopy approaches in individual nuclei, and methods derived from chromosome conformation capture (3C), which allow high-throughput and high-resolution analyses of genomic interactions at local to genome-wide scales [103–105].

Using both methods, Lanzuolo and colleagues [106] provided one of the first direct lines of evidence of loop formation in Drosophila. All major regulatory PcG-bound DNA elements at the Bithorax (BX-C) locus were found to physically interact with each other via chromatin long-range interactions. After artificial transcriptional reactivation of the BX-C locus, different conformations of the locus were observed, with the active PREs and promoters losing contact with each other [106]. This early study directly linked transcriptional regulation with the regulation of the chromatin loop formation involving PREs. A direct role for PcG in loop formation was provided by the study of the regulation of the human *GATA*-4 locus,

a target of the PRC2 EZH2 subunit [107]. When EZH2 levels are depleted, the long-range genomic contacts at the *GATA*-4 locus are disrupted, leading to transcriptional reactivation [107]. PRC2 has subsequently been shown to change chromatin conformation at the murine *HoxD* locus [108]. Analysis of a PRC2 mutant by Carbon-Copy 3C (5C) (which allows for interaction studies between many selected loci of a specific region) and by FISH showed a strong reduction of the 5C interactions over the *HoxD* region [108]. The loops involve physical contacts between PRE and PcG target genes, but can also form between genes and regulatory elements such as enhancers, silencers, and chromatin insulators [109–111].

In plants, the presence of loops directly mediated by PcG proteins has not yet been demonstrated. However, in *Zea mays* (maize) and *A. thaliana*, examples of loops between flanking regions of genes have been identified. In maize, husk-specific physical interaction between an enhancer located 100 kb upstream of the transcription start site (TSS) and the TSS is required for the high expression of the *b1* gene [112]. In *A. thaliana*, a gene loop has been identified that involves the physical interaction of the 5' and 3' flanking regions of the *FLC* locus [111,113]. Loop formation is regulated by vernalization, as it is disrupted within the first 2 weeks of cold exposure. The disruption occurs in parallel with the switch from an expressed to a Polycomb-silenced state of the *FLC* locus. Concurrently, clustering of the repressed *FLC* alleles to a single nuclear location occurs, and this is disrupted in the *vrn2* and *vrn5* mutants [114]. This finding suggests that the VRN2 and VRN5 PRC2 components influence the nuclear organization of repressed *FLC* alleles. In addition, the PcG protein LHP1 was shown to facilitate auxin-regulated loop formation between a long non-coding RNA and a gene involved in the root development pathway [78]. The presence of a loop was correlated with lower expression of these loci, increased DNA methylation, and higher levels of H3K27me3, suggesting that several repressive pathways, including the one mediated by PRC2, are involved in the process.

Understanding the formation of these chromatin hubs is an exciting and active area of research, due to their potential regulatory function and impact on the structure and evolution of genomes. However, it is still unclear whether PcG proteins are required for the formation or stabilization of the loops, and whether the 3D chromatin conformational changes are causes or consequences of altered gene expression. Furthermore, the methods used to examine chromatin loops are technically challenging. Discrepancies have been observed in the frequency of genomic interactions [115–117]. In addition, a limited number of studies compared the data resulting from FISH and 3C-derived experiments. These methods give concordant data in some cases, but also yield divergent results [108]. Therefore, in future studies of chromatin looping formation and dynamics, both techniques should be used and the results should be compared [108].

## 5. In the Third Dimension: Polycomb Mediates Higher Order Chromatin Organization

### *5.1. Polycomb and Topologically Associating Domains*

Derivatives of 3C techniques are used to establish genome-wide maps of chromatin contacts or physical interactions within the genome. In Drosophila, mouse, and human, these studies highlighted that the genome is organized into topologically associating domains (TADs) that participate in its functional architecture in the nuclear space [115,118–124]. TADs correspond to linear chromatin regions that fold as specific 3D structures that mainly favor internal interactions within a particular TAD. The mean size of TADs varies among species, ranging from 60 kb in Drosophila to 900 kb in mouse and human.

The distribution of TADs along the chromosome is associated with specific underlying histone modifications, density of nucleosomes, and DNA modification: Each TAD thus corresponds to one of the four major epigenetic topographical chromatin types: Active chromatin, Polycomb-repressed chromatin, null chromatin, or heterochromatin [122,125]. In *D. melanogaster* embryonic nuclei, Polycomb-repressed domains are highly correlated with specific TADs. Furthermore, it was proposed that the intrinsic folding regime of inactive TADs is specific and different from that of active TADS, with the probability of contacts forming between inactive TADs being lower and with a stronger association with the chromosome territory [120]. The situation is more complex in mammals, due to the discrepancy between the size of TADs (ranging from 100 kb to 10 Mb) and H3K27me3 domains (1 kb to 100 kb). Only a few overlaps have been reported between the two entities [119,126].

Sharp boundaries containing insulator sites and housekeeping genes have been shown to mark the limits between TADs [120,124]. Moreover, the modularity of TAD organization is conserved in different cell types in animals, but intra-TAD interactions can vary greatly according to cell type [118,119].

Until now, only a few conformation capture studies have been performed in plants [127–129]. In *A. thaliana*, the presence of TADs has not yet been clearly established, the results still dependent on the resolution of the Hi-C methods used [127,128]. At low resolution and at long-range distances (>100 kb), it seems that the *A. thaliana* genome contains relatively small interacting regions, which are distributed over multiple sites in the genome. The main contact regions depend mostly on the presence of DNA methylation and of H3K9me2 or H3K27me1 marks, which allows for the identification of interactive heterochromatin islands (IHI). These main interactions occur between telomeres, whereas pericentromeric heterochromatin regions interact weakly with the rest of the genome, but strongly with each other [127]. AtMORC6, a conserved Microrchidia adenosine triphosphatase required for heterochromatin condensation, seems to play a key role in establishing these heterochromatic genome contacts [130]. A limited number of H3K27me3 regions showed interactions, and these were mostly scattered throughout the genome and were dependent on PRC2 activity [127]. However, using the Hi-C method at higher resolution, about half of the contact regions were shown to be enriched in H3K27me3, H3.1, and H3.3 [128]. Although TADs were not found to be a prevailing structural feature of the *A. thaliana* genome, hundreds of insulator-like regions or regions analogous to TAD boundaries have been discovered [128]. These regions are enriched for accessible chromatin sites, various activating epigenetic marks, and highly expressed genes, whereas TAD-like fragments are characterized by opposite patterns, indicating a repressed chromatin state.

Overall, it remains challenging to identify TADs and specific Polycomb-associated TADs in plants. Conformation capture experiments in plants with larger genomes and at higher resolution in *A. thaliana* may cast light on TAD organization.

### *5.2. Polycomb Clustering*

The 3D organization of PcG target regions has been examined by imaging the subnuclear localization of Polycomb proteins. Interestingly, some Polycomb complex components tend to aggregate in foci instead of being randomly dispersed over the nuclear space [131–134]. These spots of increased signal were called PcG bodies. They differ in size and number in different cell types; as a general rule, fewer and larger foci are present in undifferentiated cells, whereas more numerous and smaller foci occur in differentiated cells [105,132].

Subsequent experiments showed that PcG foci serve as structures in which different Polycomb targets are clustered and co-repressed. A prime example comes from the Drosophila Antp and Abd-B loci. These PcG targets, despite being located 10 Mb apart from each other, co-localize to foci in embryo heads, where they are inactive, and dissociate in differentiated tissue, where at least one of them becomes activated. Further evidence emerged from chromatin conformation capture studies. As mentioned above, it was shown that PREs of different PcG targets can interact with each other even over long distances [115,123], confirming the imaging data. Recently, more than hundred proteins have been identified as regulators of the 3D distribution of PcG proteins in Drosophila [135]. In particular, proteins involved in the sumoylation pathway were shown to be critical for the Pc chromatin binding affinity, residence time, and 3D nuclear distribution. Indeed, Pc foci form enlarged aggregates in the absence of SUMO, whereas they are more dispersed when the activity of the SUMO peptidase Velo is reduced [135]. Interestingly, it was also demonstrated that the Polyhomeotic (Ph) protein assembly at PREs is defective in the absence of its *O*-GlcNAcylation and that its post-translational modification is required for the ordered and functional assembly of Ph via its SAM domain [136].

In plants, the existence of Polycomb bodies remains elusive. However, several lines of evidence suggest that Polycomb components or their targets undergo clustering. Rosa *et al.* [114] showed clustering of the PcG target, *FLC*, that was induced by vernalization and quantitatively correlated with the length of cold exposure. Higher clustering frequency was tightly associated with increased accumulation of H3K27me3 on the *FLC* nucleation site and clustering was impaired in the PcG mutants *vrn2* and *vrn5*, but not *lhp1*. In contrast, imaging of LHP1-GFP in *A. thaliana* showed that LHP1 nuclear distribution patterns vary from a rather uniform pattern in meristematic cells to patterns with multiple distinguishable foci in differentiated cells [137]. The relationship between LHP1 distribution and the differentiation status of the cell is reminiscent of the link between PcG bodies and cell differentiation in animals.

We believe that further development of imaging and 3D interaction techniques will elucidate the clustering of Polycomb targets in plants and the underlying mechanisms. However, genome arrangement in different species might also be the result of different evolutionary paths, which might lead to distinct mechanisms of repressive 3D interactions.

### *5.3. Polycomb Regulation and Spatial Distribution*

Another peculiar aspect of Polycomb regulation involves the distribution of PcG targets in the nuclear space, whose spatial rules have not yet clearly been established. However, the nuclear periphery and nuclear lamina (NL) have been shown to play important roles in gene regulation.

The NL is a protein mesh residing inside the nuclear membrane. In animals, the NL is composed of proteins called lamins, which localize to thin ring structures that cover the inner surface of the nuclear membrane. Despite a lack of substantial sequence homology to animal counterparts, a couple of proteins in plants are believed to fulfill the role of lamins, because of their lamin-like localization and the altered nuclear morphology and size in the corresponding mutants. Prominent examples of plant lamin-like proteins include NMCP1 (*D. carota*) and CROWDED NUCLEI1-4 (CRWN1-4) (*A. thaliana*). The NL is not a uniform structure; it spans the nuclear membrane and is interconnected with Nuclear Pore Complexes (NPCs), which are protein structures that form channels for the exchange of molecules between the nucleoplasm and cytoplasm (for reviews, [138–140]).

Interestingly, a physical association between the DNA and the nuclear periphery was shown to influence gene regulation. State-of-the-art examples come from studies on human [141] and Drosophila [142], in which the authors mapped the interactions between the genome and lamins using the DamID technique [143,144] and subsequent genome-wide profiling. Parts of the genome bound by lamins show lower expression and an abundance of repressive chromatin marks, including H3K27me3 [141,145–147]. Consistently, depletion of Drosophila [148] and *C. elegans* [149] lamin orthologues caused up-regulation of genes at the nuclear periphery and artificial tethering of transgenes to the periphery resulted in decreased expression [150–153].

Localization to the nuclear periphery was also associated with gene activation in several studies. For instance, several groups showed that the inducible yeast genes *INO1* and *GAL1* are recruited to the vicinity of NPCs upon gene induction and remain there for subsequent reactivation [154–157]. Moreover, global chromatin organization seems also to be tissue specific—in rod cells of nocturnal animals, there is an inverted chromatin arrangement, so that repression is correlated with the nuclear interior and activation with the periphery [158]. Rod cells lack lamins and are responsible for harvesting light; the inverted chromatin arrangement is believed to make rods more efficient. Only a few studies identified a connection between the Polycomb repression mechanism and the nuclear periphery [159,160].

In plants, these studies are still in their infancy. One example of gene repositioning related to gene expression has been reported in *A. thaliana* leaf mesophyll cells [161]. Light-inducible loci moved from the nuclear interior to the periphery upon transcriptional activation [161]. Whether these spatial movements are correlated with chromatin modifications remains to be elucidated. Interestingly, the E3 ubiquitin ligase HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENES 1 (HOS1) interacts with components of the *A. thaliana* NPC, thereby facilitating mRNA export. HOS1 activates *FLC* expression by remodeling chromatin at the *FLC* locus during short-term cold stress. The activation takes place by antagonizing the silencing role of FVE, a PcG-related protein and the displacement of its partner, the histone deacetylase HDA6. This, in turn, leads to increased acetylation of histone H3 on *FLC*, which results in higher expression [162]. Whether the two functions in nuclear export and chromatin remodeling are linked is an open question. Finally, *A. thaliana* structural components, such as the CRWN1 and CRWN4 proteins, which control the size of the nucleus, localized at the nuclear periphery and seemed to influence chromosomal organization [129]. Indeed, Hi-C experiments revealed that the *crwn4* and *crwn1* mutants exhibited increased *trans*-chromosomal interaction frequencies, suggesting higher levels of chromosomal compaction [129]. Thus, no link with plant PcG proteins has yet been demonstrated.

## 6. Conclusions

We can view the Polycomb-mediated impact on chromatin dynamics and genome regulation as an integration of three layers or dimensions of repression (Figure 1). The first one is based on the information stored on the linear genome and epigenome. Important aspects of this first dimension include recruitment at specific genomic regions, modification of histone tails, compaction of nucleosomes, and interplay with other factors to form Polycomb-repressed chromatin domains that shape the epigenetic topography of the genome. The second dimension involves the formation of loops on a locus scale, when Polycomb *cis*-regulatory elements interact with each other. Lastly, the third dimension relies on the organization of Polycomb-based TADs and the spatial localization of Polycomb-repressed domains in specific compartments of the nuclear space.

With the development of new methodology and genome-wide approaches combined with modeling to study the 3D arrangement of proteins and genomes, it has become possible to investigate the third dimension of Polycomb repression, and this has deepened our understanding of Polycomb action. PcG proteins continuously surprise us on all layers of repression, by deviating from dogmas and showing non-canonical structures, hierarchies, or mechanisms. As the methodology is now accessible for plants as well, we look forward to the challenge of deciphering the mechanism underlying the 3D PcG-mediated repression in plants, which will certainly bring novel surprises.

**Acknowledgments**

Stefania del Prete and Pawel Mikulski are supported by PhD fellowships provided by the European Commission Seventh Framework-People-2012-ITN Project EpiTRAITS (Epigenetic regulation of economically important plant traits, no-316965). The funding provided by the ITN EpiTRAITS, INRA (to Valérie Gaudin) and the Boehringer Ingelheim Foundation (to Daniel Schubert) is gratefully acknowledged.

**Author Contributions**

All authors contributed to the writing of the review.

**Conflicts of Interest**

The authors declare no conflict of interest.

**References**

1. Lewis, E.B. New mutants: Reports of P. lewis. *Drosoph. Inf. Serv.* **1947**, *21*, 69.
2. Lewis, E.B. A gene complex controlling segmentation in *Drosophila*. *Nature* **1978**, *276*, 565–570.
3. Schwartz, Y.B.; Pirrotta, V. Polycomb silencing mechanisms and the management of genomic programmes. *Nat. Rev. Genet.* **2007**, *8*, 9–22.
4. Köhler, C.; Villar, C.B. Programming of gene expression by Polycomb group proteins. *Trends Cell Biol.* **2008**, *18*, 236–243.
5. Schwartz, Y.B.; Pirrotta, V. A new world of Polycombs: Unexpected partnerships and emerging functions. *Nat. Rev. Genet.* **2013**, *14*, 853–864.
6. Pu, L.; Sung, Z.R. PcG and trxG in plants—Friends or foes. *Trends Genet.* **2015**, *31*, 252–262.
7. Xiao, J.; Wagner, D. Polycomb repression in the regulation of growth and development in *Arabidopsis*. *Curr. Opin. Plant Biol.* **2015**, *23*, 15–24.
8. Hennig, L.; Derkacheva, M. Diversity of Polycomb group complexes in plants: Same rules, different players? *Trends Genet.* **2009**, *25*, 414–423.
9. De Lucia, F.; Gaudin, V. Epigenetic control by plant Polycomb proteins: New perspectives and emerging roles in stress response. In *From Plant Genomics to Plant Biotechnology*; Palmiro Poltronieri, N.B., Corrado, F., Eds.; Woodhead Publishing, Elsevier: Cambridge, UK, 2013; pp. 31–48.
10. Kleinmanns, J.A.; Schubert, D. Polycomb and Trithorax group protein-mediated control of stress responses in plants. *Biol. Chem.* **2014**, *395*, 1291–1300.

11. Molitor, A.; Shen, W.H. The polycomb complex PRC1: Composition and function in plants. *J. Genet. Genomics* **2013**, *40*, 231–238.
12. Bemer, M.; Grossniklaus, U. Dynamic regulation of Polycomb group activity during plant development. *Curr. Opin. Plant Biol.* **2012**, *15*, 523–529.
13. Pien, S.; Grossniklaus, U. Polycomb group and trithorax group proteins in *Arabidopsis*. *Biochim. Biophys. Acta* **2007**, *1769*, 375–382.
14. Gaudin, V.; Libault, M.; Pouteau, S.; Juul, T.; Zhao, G.; Lefebvre, D.; Grandjean, O. Mutations in LIKE HETEROCHROMATIN PROTEIN 1 affect flowering time and plant architecture in *Arabidopsis*. *Development* **2001**, *128*, 4847–4858.
15. Zhang, X.; Germann, S.; Blus, B.J.; Khorasanizadeh, S.; Gaudin, V.; Jacobsen, S.E. The *Arabidopsis* LHP1 protein colocalizes with histone H3 Lys27 trimethylation. *Nat. Struct. Mol. Biol.* **2007**, *14*, 869–871.
16. Bratzel, F.; Lopez-Torrejon, G.; Koch, M.; del Pozo, J.C.; Calonje, M. Keeping cell identity in *Arabidopsis* requires PRC1 RING-finger homologs that catalyze H2A monoubiquitination. *Curr. Biol.* **2010**, *20*, 1853–1859.
17. Tavares, L.; Dimitrova, E.; Oxley, D.; Webster, J.; Poot, R.; Demmers, J.; Bezstarosti, K.; Taylor, S.; Ura, H.; Koide, H.; *et al.* RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **2012**, *148*, 664–678.
18. Lagarou, A.; Mohd-Sarip, A.; Moshkin, Y.M.; Chalkley, G.E.; Bezstarosti, K.; Demmers, J.A.; Verrijzer, C.P. dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes Dev.* **2008**, *22*, 2799–2810.
19. Beh, L.Y.; Colwell, L.J.; Francis, N.J. A core subunit of Polycomb repressive complex 1 is broadly conserved in function but not primary sequence. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E1063–E1071.
20. Bantignies, F.; Cavalli, G. Polycomb group proteins: Repression in 3D. *Trends Genet.* **2011**, *27*, 454–464.
21. Francis, N.J.; Kingston, R.E.; Woodcock, C.L. Chromatin compaction by a polycomb group protein complex. *Science* **2004**, *306*, 1574–1577.
22. Schuettengruber, B.; Cavalli, G. Polycomb domain formation depends on short and long distance regulatory cues. *PLoS ONE* **2013**, *8*, e56531.
23. Ringrose, L.; Paro, R. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Ann. Rev. Genet.* **2004**, *38*, 413–443.
24. Finnegan, E.J.; Dennis, E.S. Vernalization-Induced Trimethylation of Histone H3 Lysine 27 at FLC Is Not Maintained in Mitotically Quiescent Cells. *Curr. Biol.* **2007**, *17*, 1978–1983.
25. De Lucia, F.; Crevillen, P.; Jones, A.M.; Greb, T.; Dean, C. A PHD-Polycomb Repressive Complex 2 triggers the epigenetic silencing of FLC during vernalization. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 16831–16836.
26. Pinter, S.F.; Sadreyev, R.I.; Yildirim, E.; Jeon, Y.; Ohsumi, T.K.; Borowsky, M.; Lee, J.T. Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. *Genome Res.* **2012**, *22*, 1864–1876.
27. Dorighi, K.M.; Tamkun, J.W. The trithorax group proteins Kismet and ASH1 promote H3K36 dimethylation to counteract Polycomb group repression in *Drosophila*. *Development* **2013**, *140*, 4182–4192.

28. Srinivasan, S.; Dorighi, K.M.; Tamkun, J.W. *Drosophila* Kismet regulates histone H3 lysine 27 methylation and early elongation by RNA polymerase II. *PLoS Genet.* **2008**, *4*, e1000217.
29. Mozgova, I.; Kohler, C.; Hennig, L. Keeping the gate closed: Functions of the Polycomb repressive complex PRC2 in development. *Plant J.* **2015**, *83*, 121–132.
30. Khanna, N.; Hu, Y.; Belmont, A.S. HSP70 transgene directed motion to nuclear speckles facilitates heat shock activation. *Curr. Biol.* **2014**, *24*, 1138–1144.
31. Schuettengruber, B.; Martinez, A.M.; Iovino, N.; Cavalli, G. Trithorax group proteins: Switching genes on and keeping them active. *Nat. Rev. Mol. Cell Biol.* **2011**, *12*, 799–814.
32. Khan, A.A.; Lee, A.J.; Roh, T.Y. Polycomb group protein-mediated histone modifications during cell differentiation. *Epigenomics* **2015**, *7*, 75–84.
33. Chen, X.; Hu, Y.; Zhou, D.X. Epigenetic gene regulation by plant Jumonji group of histone demethylase. *Biochim. Biophys. Acta* **2011**, *1809*, 421–426.
34. Yang, C.; Bratzel, F.; Hohmann, N.; Koch, M.; Turck, F.; Calonje, M. VAL- and AtBMI1-mediated H2Aub initiate the switch from embryonic to postgerminative growth in *Arabidopsis*. *Curr. Biol.* **2013**, *23*, 1324–1329.
35. Gearhart, M.D.; Corcoran, C.M.; Wamstad, J.A.; Bardwell, V.J. Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. *Mol. Cell. Biol.* **2006**, *26*, 6880–6889.
36. Gao, Z.; Zhang, J.; Bonasio, R.; Strino, F.; Sawai, A.; Parisi, F.; Kluger, Y.; Reinberg, D. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol. Cell* **2012**, *45*, 344–356.
37. Calonje, M.; Sanchez, R.; Chen, L.; Sung, Z.R. EMBRYONIC FLOWER1 participates in polycomb group-mediated AG gene silencing in *Arabidopsis*. *Plant Cell* **2008**, *20*, 277–291.
38. Ogawa, H.; Ishiguro, K.; Gaubatz, S.; Livingston, D.M.; Nakatani, Y. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science* **2002**, *296*, 1132–1136.
39. Gil, J.; O'Loghlen, A. PRC1 complex diversity: Where is it taking us? *Trends Cell Biol.* **2014**, *24*, 632–641.
40. Levine, S.S.; Weiss, A.; Erdjument-Bromage, H.; Shao, Z.; Tempst, P.; Kingston, R.E. The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. *Mol. Cell. Biol.* **2002**, *22*, 6070–6078.
41. Wang, H.; Wang, L.; Erdjument-Bromage, H.; Vidal, M.; Tempst, P.; Jones, R.S.; Zhang, Y. Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **2004**, *431*, 873–878.
42. Farcas, A.M.; Blackledge, N.P.; Sudbery, I.; Long, H.K.; McGouran, J.F.; Rose, N.R.; Lee, S.; Sims, D.; Cerase, A.; Sheahan, T.W.; *et al.* KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *eLife* **2012**, doi:10.7554/eLife.00205.
43. Scheuermann, J.C.; Gutierrez, L.; Muller, J. Histone H2A monoubiquitination and Polycomb repression: The missing pieces of the puzzle. *Fly* **2012**, *6*, 162–168.
44. Klymenko, T.; Papp, B.; Fischle, W.; Kocher, T.; Schelder, M.; Fritsch, C.; Wild, B.; Wilm, M.; Muller, J. A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev.* **2006**, *20*, 1110–1122.
45. Schwartz, Y.B.; Pirrotta, V. Polycomb complexes and epigenetic states. *Curr. Opin. Cell Biol.* **2008**, *20*, 266–273.

46. Schoeftner, S.; Sengupta, A.K.; Kubicek, S.; Mechtler, K.; Spahn, L.; Koseki, H.; Jenuwein, T.; Wutz, A. Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J.* **2006**, *25*, 3110–3122.
47. Kalb, R.; Latwiel, S.; Baymaz, H.I.; Jansen, P.W.; Muller, C.W.; Vermeulen, M.; Muller, J. Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat. Struct. Mol. Biol.* **2014**, *21*, 569–571.
48. Cooper, S.; Dienstbier, M.; Hassan, R.; Schermelleh, L.; Sharif, J.; Blackledge, N.P.; de Marco, V.; Elderkin, S.; Koseki, H.; Klose, R.; *et al.* Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell Rep.* **2014**, *7*, 1456–1470.
49. Blackledge, N.P.; Farcas, A.M.; Kondo, T.; King, H.W.; McGouran, J.F.; Hanssen, L.L.; Ito, S.; Cooper, S.; Kondo, K.; Koseki, Y.; *et al.* Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* **2014**, *157*, 1445–1459.
50. Simon, J.A.; Kingston, R.E. Mechanisms of polycomb gene silencing: Knowns and unknowns. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 697–708.
51. Kim, D.H.; Sung, S. Polycomb-mediated gene silencing in *Arabidopsis thaliana*. *Mol. Cells* **2014**, *37*, 841–850.
52. He, C.; Huang, H.; Xu, L. Mechanisms guiding Polycomb activities during gene silencing in *Arabidopsis thaliana*. *Front. Plant Sci.* **2013**, doi:10.3389/fpls.2013.00454.
53. Kassis, J.A.; Brown, J.L. Polycomb group response elements in *Drosophila* and vertebrates. *Adv. Genet.* **2013**, *81*, 83–118.
54. Ku, M.; Koche, R.P.; Rheinbay, E.; Mendenhall, E.M.; Endoh, M.; Mikkelsen, T.S.; Presser, A.; Nusbaum, C.; Xie, X.; Chi, A.S.; *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* **2008**, *4*, e1000242.
55. Deng, W.; Buzas, D.M.; Ying, H.; Robertson, M.; Taylor, J.; Peacock, W.J.; Dennis, E.S.; Helliwell, C. *Arabidopsis* Polycomb Repressive Complex 2 binding sites contain putative GAGA factor binding motifs within coding regions of genes. *BMC Genomics* **2013**, *14*, 593.
56. Winter, C.M.; Austin, R.S.; Blanvillain-Baufume, S.; Reback, M.A.; Monniaux, M.; Wu, M.F.; Sang, Y.; Yamaguchi, A.; Yamaguchi, N.; Parker, J.E.; *et al.* LEAFY target genes reveal floral regulatory logic, cis motifs, and a link to biotic stimulus response. *Dev. Cell* **2011**, *20*, 430–443.
57. Hecker, A.; Brand, L.H.; Peter, S.; Simoncello, N.; Kilian, J.; Harter, K.; Gaudin, V.; Wanke, D. The *Arabidopsis* GAGA-binding factor BPC6 recruits PRC1 component LHP1 to GAGA DNA-motifs. *Plant Physiol.* **2015**, *168*, 1013–1024.
58. Berger, N.; Dubreucq, B.; Roudier, F.; Dubos, C.; Lepiniec, L. Transcriptional regulation of *Arabidopsis* LEAFY COTYLEDON2 involves RLE, a cis-element that regulates trimethylation of histone H3 at lysine-27. *Plant Cell* **2011**, *23*, 4065–4078.
59. Guo, M.; Thomas, J.; Collins, G.; Timmermans, M.C. Direct repression of KNOX loci by the ASYMMETRIC LEAVES1 complex of *Arabidopsis*. *Plant Cell* **2008**, *20*, 48–58.
60. Hyun, Y.; Yun, H.; Park, K.; Ohr, H.; Lee, O.; Kim, D.H.; Sung, S.; Choi, Y. The catalytic subunit of *Arabidopsis* DNA polymerase alpha ensures stable maintenance of histone modification. *Development* **2013**, *140*, 156–166.

61. Liu, X.; Kim, Y.J.; Muller, R.; Yumul, R.E.; Liu, C.; Pan, Y.; Cao, X.; Goodrich, J.; Chen, X. AGAMOUS Terminates Floral Stem Cell Maintenance in *Arabidopsis* by Directly Repressing WUSCHEL through Recruitment of Polycomb Group Proteins. *Plant Cell* **2011**, *23*, 3654–3670.
62. Liu, C.; Xi, W.; Shen, L.; Tan, C.; Yu, H. Regulation of floral patterning by flowering time genes. *Dev. Cell* **2009**, *16*, 711–722.
63. Cui, H.; Benfey, P.N. Interplay between SCARECROW, GA and LIKE HETEROCHROMATIN PROTEIN 1 in ground tissue patterning in the *Arabidopsis* root. *Plant J.* **2009**, *58*, 1016–1027.
64. Peschansky, V.J.; Wahlestedt, C. Non-coding RNAs as direct and indirect modulators of epigenetic regulation. *Epigenetics* **2014**, *9*, 3–12.
65. Marchese, F.P.; Huarte, M. Long non-coding RNAs and chromatin modifiers: Their place in the epigenetic code. *Epigenetics* **2014**, *9*, 21–26.
66. Brockdorff, N. Noncoding RNA and Polycomb recruitment. *RNA* **2013**, *191*, 429–442.
67. Nazer, E.; Lei, E.P. Modulation of chromatin modifying complexes by noncoding RNAs in trans. *Curr. Opin. Genet. Dev.* **2014**, *25*, 68–73.
68. Rinn, J.L.; Kertesz, M.; Wang, J.K.; Squazzo, S.L.; Xu, X.; Brugmann, S.A.; Goodnough, L.H.; Helms, J.A.; Farnham, P.J.; Segal, E.; *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **2007**, *129*, 1311–1323.
69. Chu, C.; Qu, K.; Zhong, F.L.; Artandi, S.E.; Chang, H.Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* **2011**, *44*, 667–678.
70. Tsai, M.C.; Manor, O.; Wan, Y.; Mosammaparast, N.; Wang, J.K.; Lan, F.; Shi, Y.; Segal, E.; Chang, H.Y. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **2010**, *329*, 689–693.
71. Khalil, A.M.; Guttman, M.; Huarte, M.; Garber, M.; Raj, A.; Rivea Morales, D.; Thomas, K.; Presser, A.; Bernstein, B.E.; van Oudenaarden, A.; *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11667–11672.
72. Liu, J.; Jung, C.; Xu, J.; Wang, H.; Deng, S.; Bernad, L.; Arenas-Huertero, C.; Chua, N.H. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* **2012**, *24*, 4333–4345.
73. Wang, Y.; Fan, X.; Lin, F.; He, G.; Terzaghi, W.; Zhu, D.; Deng, X.W. *Arabidopsis* noncoding RNA mediates control of photomorphogenesis by red light. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 10359–10364.
74. Di, C.; Yuan, J.; Wu, Y.; Li, J.; Lin, H.; Hu, L.; Zhang, T.; Qi, Y.; Gerstein, M.B.; Guo, Y.; *et al.* Characterization of stress-responsive lncRNAs in *Arabidopsis thaliana* by integrating expression, epigenetic and structural features. *Plant J.* **2014**, *80*, 848–861.
75. Swiezewski, S.; Liu, F.; Magusin, A.; Dean, C. Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature* **2009**, *462*, 799–802.
76. Heo, J.B.; Sung, S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* **2011**, *331*, 76–79.
77. Liu, F.; Marquardt, S.; Lister, C.; Swiezewski, S.; Dean, C. Targeted 3' processing of antisense transcripts triggers *Arabidopsis* FLC chromatin silencing. *Science* **2010**, *327*, 94–97.

78. Ariel, F.; Jegu, T.; Latrasse, D.; Romero-Barrios, N.; Christ, A.; Benhamed, M.; Crespi, M. Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. *Mol. Cell* **2014**, *55*, 383–396.
79. Latrasse, D.; Germann, S.; Houba-Herin, N.; Dubois, E.; Bui-Prodhomme, D.; Hourcade, D.; Juul-Jensen, T.; le Roux, C.; Majira, A.; Simoncello, N.; *et al.* Control of flowering and cell fate by LIF2, an RNA binding partner of the polycomb complex component LHP1. *PLoS ONE* **2011**, *6*, e16592.
80. Davidovich, C.; Wang, X.; Cifuentes-Rojas, C.; Goodrich, K.J.; Gooding, A.R.; Lee, J.T.; Cech, T.R. Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. *Mol. Cell* **2015**, *57*, 552–558.
81. Yuan, W.; Wu, T.; Fu, H.; Dai, C.; Wu, H.; Liu, N.; Li, X.; Xu, M.; Zhang, Z.; Niu, T.; *et al.* Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. *Science* **2012**, *337*, 971–975.
82. Bracken, A.P.; Dietrich, N.; Pasini, D.; Hansen, K.H.; Helin, K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.* **2006**, *20*, 1123–1136.
83. Tolhuis, B.; de Wit, E.; Muijrers, I.; Teunissen, H.; Talhout, W.; van Steensel, B.; van Lohuizen, M. Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in *Drosophila melanogaster*. *Nat. Genet.* **2006**, *38*, 694–699.
84. Bouyer, D.; Roudier, F.; Heese, M.; Andersen, E.D.; Gey, D.; Nowack, M.K.; Goodrich, J.; Renou, J.P.; Grini, P.E.; Colot, V.; *et al.* Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. *PLoS Genet.* **2011**, *7*, e1002014.
85. Oh, S.; Park, S.; van Nocker, S. Genic and global functions for Paf1C in chromatin modification and gene expression in *Arabidopsis*. *PLoS Genet.* **2008**, *4*, e1000077.
86. Zhang, K.; Sridhar, V.V.; Zhu, J.; Kapoor, A.; Zhu, J.K. Distinctive core histone post-translational modification patterns in *Arabidopsis thaliana*. *PLoS ONE* **2007**, *2*, e1210.
87. Lafos, M.; Kroll, P.; Hohenstatt, M.L.; Thorpe, F.L.; Clarenz, O.; Schubert, D. Dynamic regulation of H3K27 trimethylation during *Arabidopsis* differentiation. *PLoS Genet.* **2011**, *7*, e1002040.
88. Mohn, F.; Weber, M.; Rebhan, M.; Roloff, T.C.; Richter, J.; Stadler, M.B.; Bibel, M.; Schubeler, D. Lineage-specific polycomb targets and *de novo* DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **2008**, *30*, 755–766.
89. Filion, G.J.; van Bemmel, J.G.; Braunschweig, U.; Talhout, W.; Kind, J.; Ward, L.D.; Brugman, W.; de Castro, I.J.; Kerkhoven, R.M.; Bussemaker, H.J.; *et al.* Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **2010**, *143*, 212–224.
90. Ernst, J.; Kheradpour, P.; Mikkelsen, T.S.; Shoresh, N.; Ward, L.D.; Epstein, C.B.; Zhang, X.; Wang, L.; Issner, R.; Coyne, M.; *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **2011**, *473*, 43–49.
91. Sequeira-Mendes, J.; Araguez, I.; Peiro, R.; Mendez-Giraldez, R.; Zhang, X.; Jacobsen, S.E.; Bastolla, U.; Gutierrez, C. The functional topography of the *Arabidopsis* genome is organized in a reduced number of linear motifs of chromatin states. *Plant Cell* **2014**, *26*, 2351–2366.
92. Roudier, F.; Ahmed, I.; Berard, C.; Sarazin, A.; Mary-Huard, T.; Cortijo, S.; Bouyer, D.; Caillieux, E.; Duvernois-Berthet, E.; Al-Shikhley, L.; *et al.* Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *EMBO J.* **2011**, *30*, 1928–1938.

93. Francis, N.J.; Saurin, A.J.; Shao, Z.; Kingston, R.E. Reconstitution of a functional core polycomb repressive complex. *Mol. Cell* **2001**, *8*, 545–556.
94. Grau, D.J.; Chapman, B.A.; Garlick, J.D.; Borowsky, M.; Francis, N.J.; Kingston, R.E. Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes Dev.* **2011**, *25*, 2210–2221.
95. Eskeland, R.; Leeb, M.; Grimes, G.R.; Kress, C.; Boyle, S.; Sproul, D.; Gilbert, N.; Fan, Y.; Skoultchi, A.I.; Wutz, A.; *et al.* Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* **2010**, *38*, 452–464.
96. Shu, H.; Wildhaber, T.; Siretskiy, A.; Gruissem, W.; Hennig, L. Distinct modes of DNA accessibility in plant chromatin. *Nat. Commun.* **2012**, doi:10.1038/ncomms2259.
97. Bell, O.; Schwaiger, M.; Oakeley, E.J.; Lienert, F.; Beisel, C.; Stadler, M.B.; Schubeler, D. Accessibility of the *Drosophila* genome discriminates PcG repression, H4K16 acetylation and replication timing. *Nat. Struct. Mol. Biol.* **2010**, *17*, 894–900.
98. Canzio, D.; Chang, E.Y.; Shankar, S.; Kuchenbecker, K.M.; Simon, M.D.; Madhani, H.D.; Narlikar, G.J.; Al-Sady, B. Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. *Mol. Cell* **2011**, *41*, 67–81.
99. Canzio, D.; Liao, M.; Naber, N.; Pate, E.; Larson, A.; Wu, S.; Marina, D.B.; Garcia, J.F.; Madhani, H.D.; Cooke, R.; *et al.* A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly. *Nature* **2013**, *496*, 377–381.
100. Teif, V.B.; Kepper, N.; Yserentant, K.; Wedemann, G.; Rippe, K. Affinity, stoichiometry and cooperativity of heterochromatin protein 1 (HP1) binding to nucleosomal arrays. *J. Phys. Condens. Matter* **2015**, doi:10.1088/0953-8984/27/6/064110.
101. Blein, J.P.; Coutos-Thevenot, P.; Marion, D.; Ponchet, M. From elicitins to lipid-transfer proteins: A new insight in cell signalling involved in plant defence mechanisms. *Trends Plant Sci.* **2002**, *7*, 293–296.
102. Cheutin, T.; Cavalli, G. Polycomb silencing: From linear chromatin domains to 3D chromosome folding. *Curr. Opin. Genet. Dev.* **2014**, *25*, 30–37.
103. De Wit, E.; de Laat, W. A decade of 3C technologies: Insights into nuclear organization. *Genes Dev.* **2012**, *26*, 11–24.
104. Hovel, I.; Louwers, M.; Stam, M. 3C technologies in plants. *Methods* **2012**, *58*, 204–211.
105. Del Prete, S.; Arpon, J.; Sakai, K.; Andrey, P.; Gaudin, V. Nuclear architecture and chromatin dynamics in interphase nuclei of *Arabidopsis thaliana*. *Cytogenet. Genome Res.* **2014**, *143*, 28–50.
106. Lanzuolo, C.; Roure, V.; Dekker, J.; Bantignies, F.; Orlando, V. Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat. Cell Biol.* **2007**, *9*, 1167–1174.
107. Tiwari, V.K.; McGarvey, K.M.; Licchesi, J.D.; Ohm, J.E.; Herman, J.G.; Schubeler, D.; Baylin, S.B. PcG proteins, DNA methylation, and gene repression by chromatin looping. *PLoS Biol.* **2008**, *6*, 2911–2927.
108. Williamson, I.; Berlivet, S.; Eskeland, R.; Boyle, S.; Illingworth, R.S.; Paquette, D.; Dostie, J.; Bickmore, W.A. Spatial genome organization: Contrasting views from chromosome conformation capture and fluorescence *in situ* hybridization. *Genes Dev.* **2014**, *28*, 2778–2791.

109. Comet, I.; Schuettengruber, B.; Sexton, T.; Cavalli, G. A chromatin insulator driving three-dimensional Polycomb response element (PRE) contacts and Polycomb association with the chromatin fiber. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 2294–2299.
110. Li, H.B.; Ohno, K.; Gui, H.; Pirrotta, V. Insulators target active genes to transcription factories and polycomb-repressed genes to Polycomb bodies. *PLoS Genet.* **2013**, *9*, e1003436.
111. Zhu, D.; Rosa, S.; Dean, C. Nuclear organization changes and the epigenetic silencing of FLC during vernalization. *J. Mol. Biol.* **2014**, *427*, 659–669.
112. Louwers, M.; Bader, R.; Haring, M.; van Driel, R.; de Laat, W.; Stam, M. Tissue- and expression level-specific chromatin looping at maize *b1* epialleles. *Plant Cell* **2009**, *21*, 832–842.
113. Crevillen, P.; Sonmez, C.; Wu, Z.; Dean, C. A gene loop containing the floral repressor FLC is disrupted in the early phase of vernalization. *EMBO J.* **2013**, *32*, 140–148.
114. Rosa, S.; de Lucia, F.; Mylne, J.S.; Zhu, D.; Ohmido, N.; Pendle, A.; Kato, N.; Shaw, P.; Dean, C. Physical clustering of FLC alleles during Polycomb-mediated epigenetic silencing in vernalization. *Genes Dev.* **2013**, *27*, 1845–1850.
115. Tolhuis, B.; Blom, M.; Kerkhoven, R.M.; Pagie, L.; Teunissen, H.; Nieuwland, M.; Simonis, M.; de Laat, W.; van Lohuizen, M.; van Steensel, B. Interactions among Polycomb domains are guided by chromosome architecture. *PLoS Genet.* **2011**, *7*, e1001343.
116. Simonis, M.; Klous, P.; Splinter, E.; Moshkin, Y.; Willemsen, R.; de Wit, E.; van Steensel, B.; de Laat, W. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* **2006**, *38*, 1348–1354.
117. Bello, B.; Holbro, N.; Reichert, H. Polycomb group genes are required for neural stem cell survival in postembryonic neurogenesis of *Drosophila*. *Development* **2007**, *134*, 1091–1099.
118. Dixon, J.R.; Selvaraj, S.; Yue, F.; Kim, A.; Li, Y.; Shen, Y.; Hu, M.; Liu, J.S.; Ren, B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **2012**, *485*, 376–380.
119. Nora, E.P.; Lajoie, B.R.; Schulz, E.G.; Giorgetti, L.; Okamoto, I.; Servant, N.; Piolot, T.; van Berkum, N.L.; Meisig, J.; Sedat, J.; *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **2012**, *485*, 381–385.
120. Sexton, T.; Yaffe, E.; Kenigsberg, E.; Bantignies, F.; Leblanc, B.; Hoichman, M.; Parrinello, H.; Tanay, A.; Cavalli, G. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **2012**, *148*, 458–472.
121. Lieberman-Aiden, E.; van Berkum, N.L.; Williams, L.; Imakaev, M.; Ragoczy, T.; Telling, A.; Amit, I.; Lajoie, B.R.; Sabo, P.J.; Dorschner, M.O.; *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **2009**, *326*, 289–293.
122. Ciabrelli, F.; Cavalli, G. Chromatin-driven behavior of topologically associating domains. *J. Mol. Biol.* **2014**, *427*, 608–625.
123. Bantignies, F.; Roure, V.; Comet, I.; Leblanc, B.; Schuettengruber, B.; Bonnet, J.; Tixier, V.; Mas, A.; Cavalli, G. Polycomb-dependent regulatory contacts between distant Hox loci in *Drosophila*. *Cell* **2011**, *144*, 214–226.
124. Hou, C.; Li, L.; Qin, Z.S.; Corces, V.G. Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol. Cell* **2012**, *48*, 471–484.

125. Lan, X.; Witt, H.; Katsumura, K.; Ye, Z.; Wang, Q.; Bresnick, E.H.; Farnham, P.J.; Jin, V.X. Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages. *Nucleic Acids Res.* **2012**, *40*, 7690–7704.
126. Noordermeer, D.; Leleu, M.; Splinter, E.; Rougemont, J.; de Laat, W.; Duboule, D. The dynamic architecture of Hox gene clusters. *Science* **2011**, *334*, 222–225.
127. Feng, S.; Cokus, S.J.; Schubert, V.; Zhai, J.; Pellegrini, M.; Jacobsen, S.E. Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in *Arabidopsis*. *Mol. Cell* **2014**, *55*, 694–707.
128. Wang, C.; Liu, C.; Roqueiro, D.; Grimm, D.; Schwab, R.; Becker, C.; Lanz, C.; Weigel, D. Genome-wide analysis of local chromatin packing in *Arabidopsis thaliana*. *Genome Res.* **2015**, *25*, 246–256.
129. Grob, S.; Schmid, M.W.; Grossniklaus, U. Hi-C analysis in *Arabidopsis* identifies the KNOT, a structure with similarities to the flamenco locus of *Drosophila*. *Mol. Cell* **2014**, *55*, 678–693.
130. Moissiard, G.; Cokus, S.J.; Cary, J.; Feng, S.; Billi, A.C.; Stroud, H.; Husmann, D.; Zhan, Y.; Lajoie, B.R.; McCord, R.P.; *et al.* MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* **2012**, *336*, 1448–1451.
131. Grimaud, C.; Bantignies, F.; Pal-Bhadra, M.; Ghana, P.; Bhadra, U.; Cavalli, G. RNAi components are required for nuclear clustering of Polycomb group response elements. *Cell* **2006**, *124*, 957–971.
132. Ren, X.; Vincenz, C.; Kerppola, T.K. Changes in the distributions and dynamics of Polycomb repressive complexes during embryonic stem cell differentiation. *Mol. Cell. Biol.* **2008**, *28*, 2884–2895.
133. Saurin, A.J.; Shiels, C.; Williamson, J.; Satijn, D.P.; Otte, A.P.; Sheer, D.; Freemont, P.S. The human Polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. *J. Cell Biol.* **1998**, *142*, 887–898.
134. Vandenbunder, B.; Fourre, N.; Leray, A.; Mueller, F.; Volkel, P.; Angrand, P.O.; Heliot, L. PRC1 components exhibit different binding kinetics in Polycomb bodies. *Biol. Cell* **2014**, *106*, 111–125.
135. Gonzalez, I.; Mateos-Langerak, J.; Thomas, A.; Cheutin, T.; Cavalli, G. Identification of regulators of the three-dimensional Polycomb organization by a microscopy-based genome-wide RNAi screen. *Mol. Cell* **2014**, *54*, 485–499.
136. Gambetta, M.C.; Muller, J. O-GlcNAcylation prevents aggregation of the Polycomb group repressor polyhomeotic. *Dev. Cell* **2014**, *31*, 629–639.
137. Libault, M.; Tessadori, F.; Germann, S.; Snijder, B.; Fransz, P.; Gaudin, V. The *Arabidopsis* LHP1 protein is a component of euchromatin. *Planta* **2005**, *222*, 910–925.
138. Parry, G. The plant nuclear envelope and regulation of gene expression. *J. Exp. Bot.* **2015**, *66*, 1673–1685.
139. Tamura, K.; Goto, C.; Hara-Nishimura, I. Recent advances in understanding plant nuclear envelope proteins involved in nuclear morphology. *J. Exp. Bot.* **2015**, *66*, 1641–1647.
140. Zhou, X.; Groves, N.R.; Meier, I. Plant nuclear shape is independently determined by the SUN-WIP-WIT2-myosin XI-i complex and CRWN1. *Nucleus* **2015**, *6*, 144–153.
141. Guelen, L.; Pagie, L.; Brasset, E.; Meuleman, W.; Faza, M.B.; Talhout, W.; Eussen, B.H.; de Klein, A.; Wessels, L.; de Laat, W.; *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **2008**, *453*, 948–951.

142. Pickersgill, H.; Kalverda, B.; de Wit, E.; Talhout, W.; Fornerod, M.; van Steensel, B. Characterization of the *Drosophila melanogaster* genome at the nuclear lamina. *Nat. Genet.* **2006**, *38*, 1005–1014.
143. Van Steensel, B.; Henikoff, S. Identification of *in vivo* DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat. Biotechnol.* **2000**, *18*, 424–428.
144. Germann, S.; Gaudin, V. Mapping *in vivo* protein-DNA interactions in plants by DamID, a DNA adenine methylation-based method. *Methods Mol. Biol.* **2011**, *754*, 307–321.
145. Peric-Hupkes, D.; Meuleman, W.; Pagie, L.; Bruggeman, S.W.; Solovei, I.; Brugman, W.; Graf, S.; Flicek, P.; Kerkhoven, R.M.; van Lohuizen, M.; *et al.* Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* **2010**, *38*, 603–613.
146. Ikegami, K.; Egelhofer, T.A.; Strome, S.; Lieb, J.D. Caenorhabditis elegans chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2. *Genome Biol.* **2010**, doi:10.1186/gb-2010-11-12-r120.
147. Filion, G.J.; van Steensel, B. Reassessing the abundance of H3K9me2 chromatin domains in embryonic stem cells. *Nat. Genet.* **2010**, doi:10.1038/ng0110-4.
148. Shevelyov, Y.Y.; Lavrov, S.A.; Mikhaylova, L.M.; Nurminsky, I.D.; Kulathinal, R.J.; Egorova, K.S.; Rozovsky, Y.M.; Nurminsky, D.I. The B-type lamin is required for somatic repression of testis-specific gene clusters. *Proc. Natl. Acad. Sc.i USA* **2009**, *106*, 3282–3287.
149. Towbin, B.D.; Meister, P.; Pike, B.L.; Gasser, S.M. Repetitive transgenes in *C. elegans* accumulate heterochromatic marks and are sequestered at the nuclear envelope in a copy-number- and lamin-dependent manner. *Cold Spring Harb. Symp. Quant. Biol.* **2010**, *75*, 555–565.
150. Andrulis, E.D.; Neiman, A.M.; Zappulla, D.C.; Sternglanz, R. Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature* **1998**, *394*, 592–595.
151. Dialynas, G.; Speese, S.; Budnik, V.; Geyer, P.K.; Wallrath, L.L. The role of *Drosophila* Lamin C in muscle function and gene expression. *Development* **2010**, *137*, 3067–3077.
152. Finlan, L.E.; Sproul, D.; Thomson, I.; Boyle, S.; Kerr, E.; Perry, P.; Ylstra, B.; Chubb, J.R.; Bickmore, W.A. Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet.* **2008**, *4*, e1000039.
153. Reddy, K.L.; Zullo, J.M.; Bertolino, E.; Singh, H. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* **2008**, *452*, 243–247.
154. Casolari, J.M.; Brown, C.R.; Drubin, D.A.; Rando, O.J.; Silver, P.A. Developmentally induced changes in transcriptional program alter spatial organization across chromosomes. *Genes Dev.* **2005**, *19*, 1188–1198.
155. Kundu, S.; Horn, P.J.; Peterson, C.L. SWI/SNF is required for transcriptional memory at the yeast GAL gene cluster. *Genes Dev.* **2007**, *21*, 997–1004.
156. Schmid, M.; Arib, G.; Laemmli, C.; Nishikawa, J.; Durussel, T.; Laemmli, U.K. Nup-PI: The nucleopore-promoter interaction of genes in yeast. *Mol. Cell* **2006**, *21*, 379–391.
157. Brickner, J.H.; Walter, P. Gene recruitment of the activated INO1 locus to the nuclear membrane. *PLoS Biol.* **2004**, *2*, e342.
158. Solovei, I.; Wang, A.S.; Thanisch, K.; Schmidt, C.S.; Krebs, S.; Zwerger, M.; Cohen, T.V.; Devys, D.; Foisner, R.; Peichl, L.; *et al.* LBR and lamin A/C sequentially tether peripheral heterochromatin and inversely regulate differentiation. *Cell* **2013**, *152*, 584–598.

159. Wang, J.; Kumar, R.M.; Biggs, V.J.; Lee, H.; Chen, Y.; Kagey, M.H.; Young, R.A.; Abate-Shen, C. The Msx1 homeoprotein recruits polycomb to the nuclear periphery during development. *Dev. Cell* **2011**, *21*, 575–588.
160. Fedorova, E.; Sadoni, N.; Dahlsveen, I.K.; Koch, J.; Kremmer, E.; Eick, D.; Paro, R.; Zink, D. The nuclear organization of Polycomb/Trithorax group response elements in larval tissues of *Drosophila melanogaster*. *Chromosome Res.* **2008**, *16*, 649–673.
161. Feng, C.M.; Qiu, Y.; van Buskirk, E.K.; Yang, E.J.; Chen, M. Light-regulated gene repositioning in *Arabidopsis*. *Nat. Commun.* **2014**, doi:10.1038/ncomms4027.
162. Jung, J.H.; Park, C.M. HOS1-mediated activation of FLC via chromatin remodeling under cold stress. *Plant Signal. Behav.* **2013**, doi:10.4161/psb.27342.

### 1.4. PRC2-associated proteins in *Arabidopsis*

Identification of novel Polycomb-associated proteins offers a possibility to discover previously unknown PcG-related functions, sets of targets and modes of action. In *Arabidopsis*, numerous effector proteins participating in PcG-pathways were identified as a result of target-directed assays or unbiased genetic/physical interaction screens applied in the recent years (Mozgova and Hennig, 2015). These include the proteins associated with core PRC2 components, with several representatives shown below.

The studies on the regulation of FLOWERING LOCUS C (FLC) expression in vernalization process revealed association of PHD-finger proteins VERNALIZATION-INSENSITIVE 3 (VIN3), VERNALIZATION 5 (VRN5) and VRN5/VIN3-LIKE 1 (VEL1) with canonical PRC2 components (De Lucia et al., 2008; Wood et al., 2006). Such PHD-PRC2 complex promotes H3K27me3 deposition on 3' exon 1-5' intron 1 region of FLC locus under prolonged cold and ensures stable by spreading H3K27me3 over the whole locus upon a switch to warm conditions after vernalization.

Another example of PcG-associated protein concerns BLISTER (BLI). BLI was identified as an interactor of CLF and both show an overlapping expression pattern (Schatlowski et al., 2010). Moreover, BLI represses a subset of PRC2 target genes; however it does not affect their H3K27me3 levels, suggesting that BLI is not required for core PRC2 catalytic activity (Schatlowski et al., 2010). As lack of BLI causes developmental phenotypes only partially related to PcG-mutants, the protein is believed to influence *Arabidopsis* development also independently of PRC2.

Furthermore, cullin-RING ubiquitin ligase complex CUL4-DDB1 was shown to associate with CLF (Pazhouhandeh et al., 2011). CUL4-DDB1 physically interacts also with MSI1 and MSI4, ensures maintenance of MEDEA (MEA) paternal imprinting and controls H3K27me3 abundance on FLC and FLOWERING LOCUS T (FT) (Dumbliauskas et al., 2011; Pazhouhandeh et al., 2011).

Recently, LHP1, a PRC1 component, was demonstrated to interact with MSI1 (Derkacheva et al., 2013), providing an interesting link between two PcG-protein complexes. As LHP1 binds H3K27me3 via its chromodomain and *lhp1* mutants show decrease of this histone mark in dividing cells, LHP1 was proposed to provide self-reinforced inheritance of H3K27me3 over DNA replication (Derkacheva et al., 2013). Interestingly, *lhp1* displays relatively mild

developmental phenotypes on whole-plant level (Gaudin et al., 2001) and its importance for PRC2-pathway remains to be elucidated.

Another example of a link between PcG-protein complexes concerns EMBRONIC FLOWER 1 (EMF1). EMF1 is a PRC1 component that, nonetheless, interacts with MSI1 (Calonje et al., 2008) and influences H3K27me3 abundance on a subset of PRC2 targets (Kim et al., 2012). As EMF1 occupies both, PRC2- and non-PRC2, target genes, its function is partially separate from the PRC2-mediated repression.

Moreover, ASYMMETRIC LEAVES 1 (AS1) and ASYMMETRIC LEAVES 2 (AS2) were shown to participate in PRC2-pathway to repress KNOTTED1-like homeobox (KNOX) loci, BREVIPEDICELLUS (BP) and KNOTTED-LIKE FROM ARABIDOPSIS THALIANA 2 (KNAT2) (Lodha et al., 2008). AS1 and AS2 recognize *cis*-regulatory elements present upstream of transcriptional start sites (TSSes) at BP and KNAT2 and recruit PRC2 to their chromatin (Lodha et al., 2008). *as1* and *as2* mutants display reduced H3K27me3 occupancy on BP and KNAT2, as well as decreased binding of CLF to their chromatin (Lodha et al., 2013). Consistently, AS1 and AS2 were shown to physically interact with CLF (Lodha et al., 2013). As mentioned earlier, repression of BP and KNAT2 through AS1/2-PRC2 complex provided the first evidence of PRC2 recruitment by specific DNA-binding factors in plants.

The studies on female gametophyte development let to identify a link between RETINOBLASTOMA-RELATED 1 (RBR1), the regulator of G1-S phase transition in the cell cycle, and PRC2 complex. *rbr1* mutants display defects in cell type specification during female gametophyte maturation and an autonomous endosperm development in embryo sac, similarly to *fis2* mutants (Kuwabara and Gruissem, 2014). RBR1 interacts with FIE (Mosquna et al., 2004) and MSI1 (Jullien et al., 2008). In addition, RBR1 cooperates with PRC2 to establish H3K27me3 during female gametophyte development, at least partially, by suppression of METHYLTRANSFERASE 1 (MET1) in female gametes (Jullien et al., 2008), a process required for the proper imprinting of the target genes.

Another PRC2-associated protein is TBP-ASSOCIATED FACTOR 13 (TAF13). Lack of TAF13 was reported to cause enlarged endosperm in the seed, as observed in *fis* and *mea* mutants (Lindner et al., 2013), and TAF13 protein physically interacts with core PRC2 methyltransferases: SWN and MEA (Lindner et al., 2013). Noteworthy, an association between PRC2 and TAFs was shown also in *Drosophila* (Breiling et al., 2001; Saurin et al.,

2001), demonstrating a connection between PcG-mediated functions and general transcription factors.

Lastly, a search for negative regulators of PcG proteins allowed to characterize UPWARD CURLY LEAF 1 (UCL1) as suppressor of CLF activity (Jeong et al., 2011). The overexpression of UCL1 leads to leaf curling and early flowering, similarly to *clf* mutants. UCL1 is a F-box protein participating in ubiquitin-26S proteasome pathway and reduces CLF protein level via direct binding, a process that eventually causes up-regulation of several PRC2 targets. The important implications of UCL1 in PRC2-pathway came from the studies on endosperm development. As UCL1 is expressed in the endosperm and ectopic expression of CLF in the endosperm leads to *mea*-like phenotypes, UCL1 was proposed to regulate a balance between activities of CLF and MEA (Jeong et al., 2011). In that way, UCL1 would counteract suppression of MEA functions caused by potential overexpression of CLF.

### 1.5. PWO1 – a novel PRC2-associated protein in *Arabidopsis*

PWWP INTERACTOR OF POLYCOMBS 1 (PWO1), formerly known as SWINGER/CURLY LEAF-INTERACTOR 1 (SCI1) was identified as a physical interactor of CLF and SWN via yeast-two-hybrid screen. PWO1 interacts not only physically, but also genetically with CLF and control expression of several common PRC2 targets (Hohenstatt, 2012). Consequently with common function of PWWP motifs in binding chromatin (Wu et al., 2011), PWWP from PWO1 has the ability to bind histones and a range of their modifications, including H3K27me3 (Hohenstatt, 2012). Moreover, PWO1 forms a small gene family in *Arabidopsis* together with its two close homologs, PWO2 and PWO3. Single and double mutants within PWO family display only mild aberrations comparing to the wildtype, with slightly early flowering phenotype. In contrast, *pwo1 pwo2 pwo3* triple mutant show meristem arrest and embryo lethality (Hohenstatt, 2012). Overall, initial characterization of PWO1 suggested that the protein is important for development, has a putative chromatin binding function and associates with PRC2 in physical manner.

### 1.6. Evolutionary aspect of PcG proteins' function

Another aspect of PcG-mediated repression concerns evolutionary conservation of its effectors and function. Initially characterized in multicellular invertebrate *Drosophila* (Moazed and O'Farrell, 1992), PcG proteins were subject of phylogenetic and biochemical analyses aimed at deciphering their presence in the other evolutionary clades.

Homologs of PcG proteins are widespread in the eukaryotic kingdoms (Butenko and Ohad, 2011; Shaver et al., 2010; Sowpati et al., 2015). Nevertheless, Polycomb group proteins were found to be absent from the unicellular yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, suggesting their presence in the last common unicellular ancestor between plant and animal lineages, but subsequent loss in some subclades (Bowman et al., 2007). Such notion was confirmed by discoveries of homologous core PRC2 components in simple unicellular alga *Chlamydomonas reinhardtii* (Shaver et al., 2010), yeast *Cryptococcus neoformans* (Dumesic et al., 2015) and protozoan *Tetrahymena thermofila* (Liu et al., 2007).

Phylogenetic analyses highlighted also the discrepancies in evolutionary dynamics between PRC2 and PRC1, as well as between individual PcG complexes' components. PRC2 is conserved in plants, animals and fungi, albeit only homologs of E(z), ESC and p55 were found to be highly abundant (Shaver et al., 2010). It is worth noteworthy to mention, however, that as p55 participates also in other complexes than PRC2 in higher eukaryotes (Martínez-Balbás et al., 1998; Tyler et al., 1996), it remains to be proven whether this subunit's function in less known species is PcG-related. In contrast, numerous species lack homologous Su(z)12 subunit (Shaver et al., 2010), reflecting its frequent losses in evolution.

PRC1 displays lower degree of conservation. PRC1 was previously thought to emerge in the common ancestor of bilateral animals and be lost in some animal subclades, as no sequence homologs of its components were found in plants and *C. elegans* (Whitcomb et al., 2007). However, recent searches revealed a presence of homologs of individual PRC1 components, RING1 and BMI1 in Chlorophyta (Berke and Snel, 2015). Moreover, the analogous proteins that perform equivalent function to PRC1 components were discovered in land plants (Yong et al., 2016) and several animal species like *C. elegans* (Whitcomb et al., 2007). PRC1 is considered to be a subject of convergent evolution across and within kingdoms. Similarly to p55, the homologs of RING1 and BMI1 might as well fulfil non-PcG-related functions.

## 2. Aims of study

Despite an extensive characterization since their discovery in *Drosophila*, several aspects of Polycomb group proteins and PcG-mediated chromatin repression still remain unclear. The advances in understanding the Polycomb group *modus operandi* revealed an intriguing network of PcG-associated proteins and crucial role in 3D chromatin organization, both of which are currently poorly understood, particularly in plants. Furthermore, deciphering the evolutionary history of Polycomb group proteins uncovered their broad distribution across phylogenetic kingdoms and confirmed the importance of their role in chromatin regulation. Importantly, the majority of phylogenetic analyses on PcG proteins focused on higher branches of Eukaryota and did not biochemically characterize their typical chromatin modifications. Consequently, the evolutionary history of PcG proteins in lower eukaryotes, especially lower plant species, remains largely unknown.

Given the rationale above, this work concerns the further characterization of PWO1, a novel PRC2-associated *Arabidopsis* protein, and its role in connecting PcG-mediated gene repression and the subnuclear periphery. Secondly, this work focuses on the conservation of PcG proteins in lower plants and elucidates H3K27me3 features in model unicellular red alga, *Cyanidioschyzon merolae*.

### 2.1. Understanding the role of PWO1 in PcG-mediated gene repression

In order to further characterize PWO1role in PRC2-pathway and its putative novel functions, a mixture of genome-wide assays and target-directed approaches was used.

Firstly, a co-immunoprecipitation experiment coupled with mass-spectrometry was performed to identify new putative interactors of PWO1 protein. Interestingly, the results revealed an association with several proteins of nuclear lamina (NL). Subsequently, yeast-two-hybrid and Fluorescence Resonance Energy Transfer (FRET) methods were employed to confirm the physical interaction between PWO1 and one of the main NL proteins, CROWDED NUCLEI 1 (CRWN1). The interaction was studied also on a genetic level by generation and phenotyping of *pwo1 crwn1* double mutant.

As stable NL components, including CRWN1, show the localization at the nuclear periphery, microscopic analyses in *Nicotiana benthamiana* and *Arabidopsis thaliana* were done to

unravel a subnuclear localization of PWO1 in heterologous and native organism systems. Moreover, reciprocal crosses (*PWO1::PWO1-GFP* in *crwn1* and *CRWN1::CRWN1-GFP* in *pwo1 pwo3*) were used to decipher an inter-dependence in subnuclear localization between PWO1 and CRWN1.

The functional connection between PWO1, CRWN1 and core PRC2 was assessed further by the RNA-seq experiment and transcriptomic data analysis in *pwo1* and *crwn1 crwn2* stable knockout lines. Differentially expressed genes were classified by gene ontology and cross-compared between both mutant lines and with published set of H3K27me3 targets (Oh et al., 2008).

Finally, as PWO1 decreases expression of numerous PRC2 targets and H3K27me3 is the histone mark catalysed by PRC2 (Hohenstatt, 2012), this work aimed at finding H3K27me3 differentially-bound gene targets in *pwo1*. A chromatin immunoprecipitation (ChIP) method coupled with RT-qPCR or sequencing (ChIP-seq) was used to achieve this goal.

The quality of ChIP-seq and RNA-seq data was subsequently assessed via correlation between biological replicates and independent validation by RT-qPCR.

### 2.2. PcG proteins in lower plant species and H3K27me3 profile in *C. merolae*

Firstly, reciprocal BLAST searches were performed to identify homologs of PRC1 and PRC2 members in the several representative species from lower plant branches. In order to resolve a phylogenetic distance between identified homologs, the Bayesian evolutionary trees were created from their amino acid sequences.

Given a conservation of PRC2 members and unknown H3K27me3 presence in several species selected for this study, the next step concerned biochemical identification of H3K27me3 in their protein extracts and confirmation of detection specificity by peptide competition assay.

For further characterization of H3K27me3 in lower plants, this study focused on the unicellular red alga, *Cyanidioschyzon merolae*. The choice of the species was based on the fact that *C. merolae* contains core PRC2 homologs, displays biochemical presence of H3K27me3, has simple developmental program contains genome with very low abundance of repetitive elements (Nozaki et al., 2007). The latter features are especially interesting in

relation to higher eukaryotes, in which H3K27me3 regulates multicellular developmental transitions and silences repetitive sequences (in animals like *Drosophila* (Yin et al., 2011) or mouse (Ishak et al., 2016)), but not in higher plants (Deleris et al., 2012; Lafos et al., 2011; Park et al., 2012).

In order to identify H3K27me3 targets and its role in *C. merolae*, ChIP-RT-qPCR and ChIP-seq assays were performed on chromatin extracts with anti-H3K27me3 antibody. Subsequently, the identified H3K27me3 targets were characterized bioinformatically to compare their features with the other H3K27me3-containing eukaryotes. Such bioinformatic analyses concerned functional classification of the targets, their genomic type and H3K27me3 occupancy on the chromosomal or gene body-level. Furthermore, in order to assess the role of *C. merolae* H3K27me3 in gene repression, RNA-seq experiment was performed and both, transcriptomic and H3K27me3 occupancy data, were cross-compared.

# 3. Results

## 3.1. Manuscript I

PWO1 was previously identified as a PRC2-associated protein in *Arabidopsis* (Hohenstatt, 2012). The following work aimed at deciphering PWO1 function in PRC2-pathway and novel PWO1 associations. Firstly, a, co-immunoprecipitation experiment with mass spectrometry (CoIP-MS/MS) was performed and revealed a striking association of PWO1 with proteins enriched in the crude nuclear lamina (NL) extract (Sakamoto and Takagi, 2013), a protein mesh located at the nuclear periphery. A physical interaction between PWO1 and a constitutive NL component, CROWDED NUCLEI 1 (CRWN1), was confirmed using independent proteomic methods. Next, in order to determine the genetic interaction between PWO1 and CRWN1, *pwo1 crwn1* double mutant was generated and phenotyped. Furthermore, the function connection between PWO1 and CRWN1 was studied on the level of transcriptome (by RNA-seq) and subnuclear localization (by confocal microscopy). The analyses revealed PWO1 putative role in linking chromatin repression pathway with the nuclear periphery in plants.

The results were summarized in Manuscript I: 'PWWP INTERACTOR OF POLYCOMBS (PWO1) links PcG-mediated gene repression to the nuclear lamina in *Arabidopsis*'.

**Manuscript I**

# PWWP INTERACTOR OF POLYCOMBS (PWO1) links PcG-mediated gene repression to the nuclear lamina in *Arabidopsis.*

**Pawel Mikulski1, Sara Farrona2, Mareike Hohenstatt3, Cezary Smaczniak4, Kerstin Kaufmann4, Gerco Angenent5, Daniel Schubert1a**

1 Institute for Biology, Free University Berlin, Germany

2 School of Natural Sciences, NUI Galway, Ireland

3 R&D Epigenetics Department, Diagenode SA Liège, Belgium

4 Institute of Biochemistry and Biology, University of Potsdam, Germany

5 Laboratory of Molecular Biology, Wageningen University, The Netherlands

a corresponding author: dan.schubert@fu-berlin.de

Keywords: nuclear lamina, H3K27me3, Polycomb, CRWN, PWO, PRC2

Author contributions:

PM, SF, MH, DS designed the research. DS, KK, GA, MH designed CoIP-MS/MS experiment. MH performed CoIP-MS/MS experiment. CS analyzed MS spectral data. PM and MH performed nuclear morphology measurements. PM and SF performed microscopy analyses on PWO1 subnuclear localization in *Nicothiana benthamiana*. PM did FRET and Y2H experiments, genetic interaction studies, subnuclear localization in *Arabidopsis* analyses and transcriptomic experiments. The manuscript was written by PM and revised by DS.

### 3.1.1. Abstract

Polycomb group (PcG) proteins facilitate chromatin-mediated gene repression through the modification of histone tails in a wide range of eukaryotes, including plants and animals. One of the PcG protein complexes, Polycomb Repressive Complex 2 (PRC2), promotes repressive chromatin formation via tri-methylation of lysine-27 on histone H3 (H3K27me3). The animal PRC2 is implicated in impacting subnuclear distribution of chromatin as its complex components and H3K27me3 are functionally connected with the nuclear lamina (NL) - a peripheral protein mesh that resides underneath the inner nuclear membrane (INM) and consists of lamins and lamina-associated proteins. In contrast to animals, NL in plants has an atypical structure and its association with PRC2-mediated gene repression is largely unknown. Here, we present a connection between lamin-like protein, CROWDED NUCLEI 1 (CRWN1), and a novel PRC2-associated component, PWWP INTERACTOR OF POLYCOMBS 1 (PWO1), in *Arabidopsis thaliana*. We show that PWO1 and CRWN1 proteins associate physically with each other, act in the same pathway to maintain nuclear morphology and control expression of similar set of target genes. Moreover, we demonstrate that PWO1 proteins form speckle-like foci located partially at the subnuclear periphery in *Nicotiana benthamiana* and *Arabidopsis thaliana.* Ultimately, as CRWN1 and PWO1 are plant-specific, our results argue that plants developed an equivalent, rather than homologous, mechanism of linking PRC2-mediated chromatin repression and nuclear lamina.

### 3.1.2. Introduction

Polycomb group (PcG) proteins facilitate chromatin-mediated gene repression by modifying histone tails in a wide range of eukaryotes, including plants and animals. A subset of PcG proteins constitutes Polycomb Repressive Complex 2 (PRC2). A canonical form of PRC2 was initially discovered in *Drosophila melanogaster*, where it consists of: E(z) – a component with catalytic functions, Esc –a WD40 motif-containing scaffolding protein, Su(z)12 – a Zinc-finger subunit facilitating binding to nucleosomes, and p55 – a nucleosome remodelling factor (Schwartz and Pirrotta, 2007). PRC2 is present also in the model higher plant, *Arabidopsis thaliana*, albeit with several components being multiplicated and partially diversified in expression pattern and set of targets. PRC2 in *Arabidopsis thaliana* is made up of: CLF/MEA/SWN - E(z) homologs, VRN2/EMF2/FIS - Su(z)12 homologs, FIE - ESC homologs, and MSI1-5 - p55 homologs (Hennig and Derkacheva, 2009). PRC2 mediates its

repressive function by catalysing methylation of lysine 27 on histone H3 (H3K27), which is in turn recognized by reader proteins that silence underlying DNA sequences (Xiao et al., 2016). Depending on the species, PRC2 methylates H3K27 in various contexts (Ebert et al., 2004; Ferrari et al., 2014; Jacob and Michaels, 2009), with H3K27 trimethylation (H3K27me3) being the most typical modification.

PRC2 functions on different layers of gene repression (reviewed in (Del Prete et al., 2015)), including regulation of target spatial distribution in the nucleus, namely at the nuclear periphery. Nuclear periphery is a subnuclear space in the vicinity to nuclear envelope (NE), a double lipid bilayer forming inner and outer nuclear membrane (INM and ONM, respectively), and nuclear lamina (NL), a nuclear protein mesh physically associated with INM. NL comprises of lamin-associated membrane proteins and lamins themselves, the latter one being categorized as A- and B-type, depending on the structural properties and expression pattern (Dechat et al., 2010). Apart from a role in the nuclear architecture, both lamin types were implicated to influence subnuclear organization of the chromatin, exemplified by the existence of lamin-bound chromatin regions, called lamina-associated domains (LADs) in animals.

Several lines of evidence indicate that PRC2 plays an important role in the regulation of LADs and peripheral localization of specific chromatin types. These include: (1) H3K27me3 enrichment at tissue-specific LADs (variable LADs, vLADs) and the borders of constitutive LADs in mammalian cell cultures (Fakhouri et al., 2010; Harr et al., 2015); (2) H3K27me3 enrichment at chromatin domains bound by LEM-2, lamina-associated protein, in *C. elegans* (Ikegami et al., 2010); (3) H3K27me3 accumulation at transgene multicopy arrays that are associated with NE in *C. elegans* (Meister et al., 2010; Towbin et al., 2010); (4) decrease of peripheral localization for vLAD fragment upon knockdown of EZH2, a mammalian homolog of E(z), in mouse fibroblast (Harr et al., 2015) and (5) recruitment of vLAD fragment to NL mediated by Ying Yang 1 (YY1), an interactor of PRC2 (Harr et al., 2015).

However, general H3K27me3 occupancy and PRC2 targeting are not the exclusive determinants of chromatin association with animal lamina. On one side, a number of PRC2-independent pathways was reported to be also involved in that process (Harr et al., 2016), with some being specific to cell type or the species. On the other, only a subset of all PRC2 targets can be found in LADs (Hänzelmann et al., 2015) and spatial distribution of PRC2 components or H3K27me3 in some tissue types concerns localization to both, nuclear

periphery and nuclear interior (Eberhart et al., 2013; Luo et al., 2009; Wu et al., 2013). Therefore, understanding the role of PRC2 in peripheral positioning requires characterization of target-specific PRC2-associated proteins and deciphering the interplay with other factors involved in tethering to NL.

In plants, general impact of chromatin on gene spatial distribution and contribution of Polycomb group-mediated repression in this process was investigated only in a handful of studies (Pecinka et al., 2002; Rosa et al., 2013; Rosin et al., 2008) and remains largely unknown. In fact, no lamina-associated domains were identified in the green lineage and it was long believed that plants do not have the lamina, as they lack sequence homologs of the lamin proteins (Melcer et al., 2007). Instead, plant NL contains the lamina-like proteins (Ciska and Moreno Díaz de la Espina, 2014), which, similarly to their animal counterparts, possess coiled-coil protein motif and show general localization to the nuclear periphery (Ciska and Moreno Diaz de la Espina, 2013; Gardiner et al., 2011). Lamina-like genes in *Arabidopsis* form plant-specific family CROWDED NUCLEI (CRWN), containing 4 members (*CRWN1-4*). The most prominent phenotypes of CRWN family loss-of-function mutants are: reduced nuclear size, increased nuclear DNA density and abnormal nuclear shape, with the most pronounced effect seen in *crwn 1 crwn2* (Dittmer et al., 2007; Sakamoto and Takagi, 2013; Wang et al., 2013). Furthermore, CRWN proteins associate with a range of other plant NE/NL components (Goto et al., 2014; Graumann, 2014). Interestingly, CRWN genes also affect constitutive heterochromartin organization and 3D chromosome arrangement - different CRWN mutant combinations show altered chromocentres' integrity (Dittmer et al., 2007; Wang et al., 2013) and *crwn1 crwn4* exhibits elevated *trans*-chromosomal interactions, but unchanged contact frequency *in cis* (Grob et al., 2014). However, the exact mechanism of CRWN proteins' function in chromatin organization remains unknown.

In our previous work, we identified *Arabidopsis* protein PWWP INTERACTOR OF POLYCOMB1 (PWO1) as a novel plant-specific PRC2-associated factor. We showed that PWO1 interacts physically with PRC2 methyltransferases, displays epistatic interaction with CLF, controls expression of several PRC2-dependent target genes and is needed for full H3K27me3 occupancy at several PcG target genes (Hohenstatt, 2012).

Here, we present a connection between PWO1 and the plant nuclear lamina. We demonstrate that PWO1 interacts with CRWN1 physically and genetically, and that they both control

expression of a similar set of genes. Furthermore, we show that PWO1 regulates nuclear size and partially localizes to peripheral speckles in the nucleus. Taken together, our results provide a putative link between PcG-mediated gene repression and chromatin organization at the subnuclear periphery.

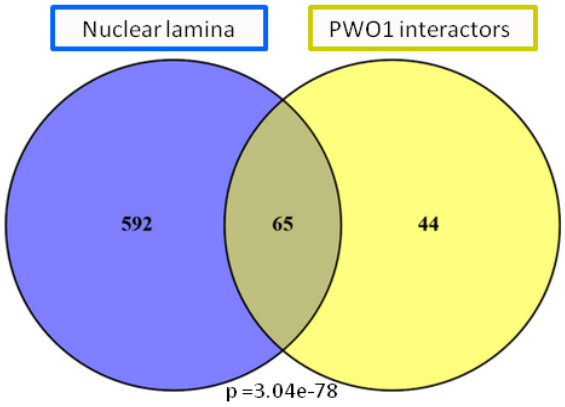
### 3.1.3. Results

#### PWO1 physically associates with NL/NE proteins

After initial characterization of PWO1 (Hohenstatt, 2012), we sought to identify its protein interactors by unbiased quantitative proteomics. We undertook co-immunoprecipitation experiment coupled with mass spectrometry using a bait of PWO1-GFP fusion protein in *PWO1::PWO1-GFP Arabidopsis* transgenic line (Hohenstatt, 2012). The protein abundance was scored via label-free quantification (LFQ) analysis and a comparison to background sample (Col-0, wildtype) was used to identify significantly enriched proteins in the *PWO1::PWO1-GFP* line. As a result, we obtained a list of 109 putative PWO1 interactors. Interestingly, ~ 60% of those overlap with putative components of crude plant nuclear lamina fraction (Sakamoto and Takagi, 2013) (Fig.1A), including constitutive NL/NE members and chromatin-associated proteins (Fig.1B, Table S1). Our attention was drawn to the presence of CROWDED NUCLEI 1 (CRWN1), a nuclear lamina protein with prominent role in nuclear morphology (Wang et al., 2013) that affects chromocenter organization (Dittmer et al., 2007) and interchromosomal contact frequencies (Grob et al., 2014).

In order to confirm PWO1-CRWN1 interaction, we performed Yeast-Two-Hybrid (Y2H) and Acceptor Photobleaching FRET (FRET-APB) experiments. Both methods revealed no or very low interaction between CRWN1 and full-length PWO1, but much stronger association when only a C-terminal PWO1 fragment (634 – 769 aa) was used instead (Fig.2, Fig.3, Fig.S1). Such result might indicate steric constrains of overall PWO1 protein fold caused by introduction of fusion tags (Gal4-domains or FRET fluorophores) and masking of CRWN1-binding site in full-length PWO1 construct versions. In contrast, the usage of truncated PWO1 protein containing CRWN1-interaction region resolved potential steric effects. The interaction with CRWN1 was shown to be specific to C-terminal part of PWO1 and not a simple outcome of random binding of any truncated PWO1 versions, as the other PWO1 fragments (frag.1 and 2) show still a background-level interaction (Fig.2B).

A




B

| TAIR ID | Name | p-value | log2 ratio | Unique Peptides | Sequence Coverage [%] |
|---|---|---|---|---|---|
| **Bait protein** | | | | | |
| AT3G03140 | PWO1 | 2.76E-03 | 21.58 | 17 | 28.7 |
| **Constitutive NL/NE components** | | | | | |
| AT1G67230 | CRWN1 | 3.12E-02 | 3.63 | 3 | 3.4 |
| AT5G65770 | CRWN4 | 4.13E-04 | 6.50 | 4 | 5.2 |
| AT5G04990 | SUN1 | 3.54E-03 | 2.54 | 2 | 7.6 |
| **Chromatin-associated proteins** | | | | | |
| AT5G58230 | MSI1 | 5.13E-03 | 8.08 | 2 | 7.1 |
| AT1G77180 | SKIP | 2.54E-03 | 2.84 | 4 | 8.2 |
| AT5G18620 | CHR17 | 3.76E-03 | 12.04 | 4 | 4.3 |
| AT5G14170 | CHC1 | 1.99E-03 | 16.23 | 8 | 15.7 |
| AT1G20960 | BRR2 | 7.19E-03 | 1.86 | 22 | 10.9 |
| AT5G64420 | Pol V | 4.42E-02 | 2.56 | 3 | 2.4 |

**Fig.1. Identification of PWO1 interactors.** (A) An overlap between *Arabidopsis* nuclear lamina (NL) fraction (Sakamoto and Takagi, 2013) and PWO1 interactors. PWO1 interactors correspond to proteins showing significant ($p<0.05$) enrichment over the wildtype in CoIP-MS/MS experiment with PWO1-GFP (*PWO1::PWO1-GFP* line) used as a bait. Significance level of the overlap was calculated using the Student's t-test. (B) Mass spectrometry results for bait protein and selected candidates from an overlap of PWO1 interactors and NL components. For the full list of 65 overlapped proteins, see Table S1.

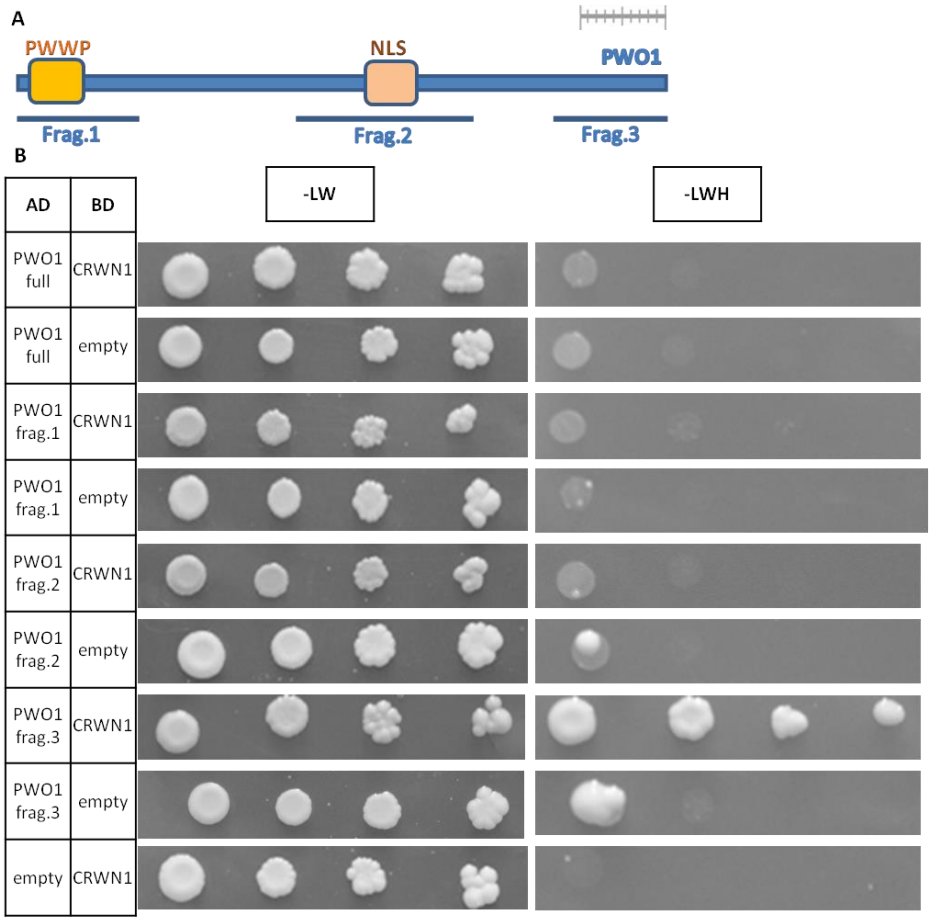



**Fig.2. Yeast-two-hybrid (Y2H) analysis on PWO1-CRWN1 interaction.** (A) Scheme of PWO1 protein sequence with marking of PWO1 fragments used for Y2H experiments. Scale bar on the top of the scheme corresponds to 100 aminoacids. (B) Y2H experiment. Y2H constructs containing inserts N-terminally fused to Gal4 activating domain (AD) or Gal4 binding domain (BD) were transformed in AH109 S.cerevisae strain. Coding sequences of full length CRWN1 (CRWN1), full length PWO1 (PWO1 full) and PWO1 fragments (PWO1 frag.1-3) were used as the inserts. Transformation with insert-containing constructs and empty vectors served as a negative control. Overall yeast growth and protein interaction between CRWN1, PWO1 and PWO1 fragments was assessed on transformation medium (-LW) and selection medium (-LWH).

Interestingly, the CRWN1-interaction region in PWO1 sequence does not contain any annotated protein domain (Fig.2A). However, an alignment to the closest PWO1 plant homologs revealed a high level of conservation for PWO1 C-terminal sequence in *Brassicaceae* species other than *Arabidopsis* (Fig.S2), highlighting a functional importance of this fragment. In contrast, PWO1 C-terminal fragment is largely absent from PWO2 and PWO3, suggesting partially separate functions between PWO family proteins in *Arabidopsis*.

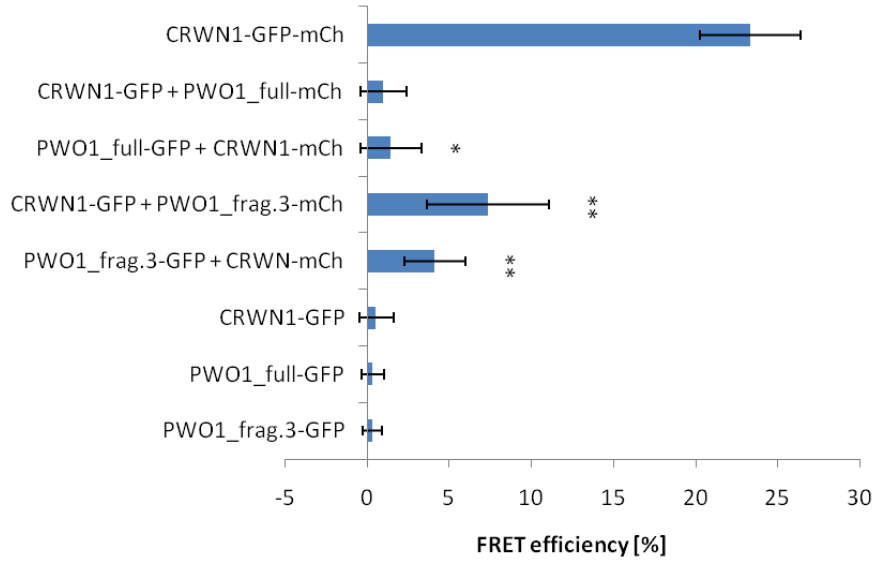



**Fig.3. FRET-APB-based interaction between PWO1 and CRWN1.** FRET efficiency was calculated for nuclei from the positive control (n=10), negative controls (n=15-17) and interaction samples (n=15-27). Error bars represent standard deviation values. Significance between interaction samples and their respective negative controls was inferred from Student's t-test. Single asterisk (*) represents significance threshold: $p<0.05$, double (**): $p<0.01$.

## PWO1 interacts genetically with CRWN1

Given the physical association of PWO1 and CRWN1, we sought to investigate their functional connection by studying *pwo1*, *crwn1* and *pwo1 crwn1* mutants on the phenotypic level. As PWO1 is PRC2-associated protein (Hohenstatt, 2012), we included also *swn* and *swn crwn1* transgenic lines in our analysis to characterize the link between nuclear lamina and canonical PRC2.

A prominent feature of mutants in NE or NL components is the reduction in nuclear size and change of nuclear shape. In order to characterize alterations in nuclear morphology, we isolated whole-seedlings nuclei from abovementioned mutants, stained them with DAPI and measured their nuclear area and circularity (Fig.4). Consistently with the other studies (Wang et al., 2013), we observed a reduction in nuclear area and increase in circularity index in *crwn1*, compared to the wildtype control (Col-0). Interestingly, lower average nuclear area was also observed in *pwo1*, albeit with no significant changes in nuclear shape. The analysis on *pwo1 crwn1* phenotype revealed that both genes act in the same pathway to control the nuclear size as the double mutant phenocopies either of the single mutants, rather than displays an additive phenotype (Fig.4B). However, as both, *pwo1* and *crwn1*, show similar nuclear area reduction, we could not determine the phenotypic dominance of one gene over

the other in the double knock-out line. Interestingly, we also observed a genetic interaction in nuclear shape phenotype. Namely, as *pwo1 crwn1* shows circularity index average on the similar level to *pwo1*, CRWN1-dependent nuclear shape alteration was suppressed by the lack of functional *PWO1* (Fig.4C). Consistently, we found that the nuclei from *pwo1 crwn1* can be characterized by a broad distribution of circularity index levels, similarly to the *pwo1* single mutant. In contrast, the nuclei from *crwn1* are narrowly distributed and enriched at the high circularity index levels (Fig.4D). Such result suggests that PWO1 is required for CRWN1-mediated control of nuclear shape.

Furthermore, we observed no significant changes in nuclear morphology in *swn*. The double mutant *crw1 swn* shows an increased average circularity and reduced average nuclear area, hence resembling *crwn1*. Consistently, a distribution of individual nuclei in *crwn1 swn* is more similar to the distribution observed in *crwn1*, rather than *swn*. The phenocopy of nuclear morphology changes between *swn crwn1* and *crwn1*, yet lack of such phenotype in *swn,* suggest altogether that the canonical PRC2 doesn't influence CRWN1-mediated control of nuclear architecture.

Overall, our results suggest that PWO1 and CRWN1 determine nuclear size and shape within the same genetic pathway. In contrast, canonical PRC2 component, SWN, does not influence morphology of the nucleus, irrespective of CRWN1 function.

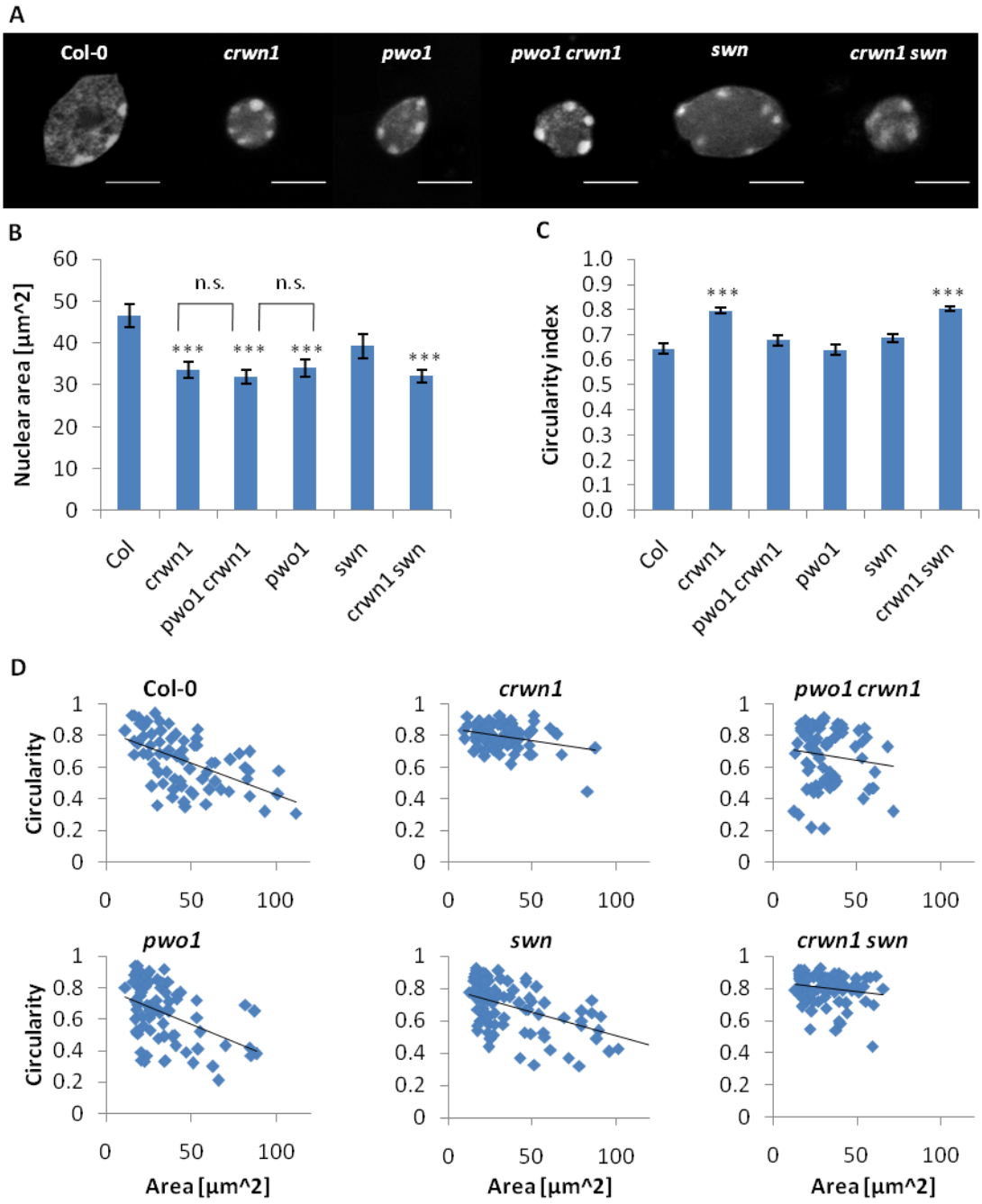



**Fig.4. Genetic interaction of *pwo1*, *crwn1* and *swn*.** (A) DAPI-stained nuclei in different genotypes. Scale bar corresponds to 10 µm. Nuclear area (B) and circularity (C) measurements. Calculations were performed on Z plane showing the largest area of particular nucleus in the Z-stack. 72-92 nuclei per genotype were measured. Circularity index was calculated as 4π(area/perimeter^2), with the maximal value of 1.0 corresponding to the perfect circle. Significance was calculated in Student's t-test for comparison of the mutants to the wildtype control (Col). Significance was calculated also for the comparison of *crwn1* to *pwo1 crwn1*, and of *pwo1* to *pwo1 crwn1* (indicated by brackets above respective chart bars). Non-significant changes with threshold p = 0.05 are indicated (n.s.), whereas significant changes with p < 0.001 are marked with triple asterisk (***). (D) The relationship between circularity index and nuclear area in nuclei from genotypes used in the study. Blue points represent individual nuclei, black line displays the trendline.

## PWO1 subnuclear localization

As PWO1 interacts physically and genetically with CRWN1, a nuclear lamina component that localizes to the nuclear periphery, we asked whether both proteins share similar subnuclear localization. In *Arabidopsis*, CRWN1 expressed from its endogenous promoter forms a thin ring underlying the border of the nucleus (Dittmer et al., 2007). If expressed from a strong, 35S promoter in the heterologous system (*Nicotiana benthamiana*), CRWN1 localizes partially to the nuclear periphery, forms ring-like or filamentous structures and induces moderate deformation of NE (Goto et al., 2014).

In order to study PWO1 localization in a heterologous system, we expressed PWO1-GFP fusion protein in *N. benthamina* leaves. In order to prevent formation of over-accumulated protein aggregates, the expression was driven by a beta-estradiol-inducible i35 promoter. We observed that PWO1 localizes to the nucleoplasm and forms multiple speckles present at the nuclear periphery and the nucleolus (Fig.5, Fig.S3). We detected 17-37 speckles per nucleus (n=13), with a vast majority of them located at the nuclear periphery (Fig.5B).

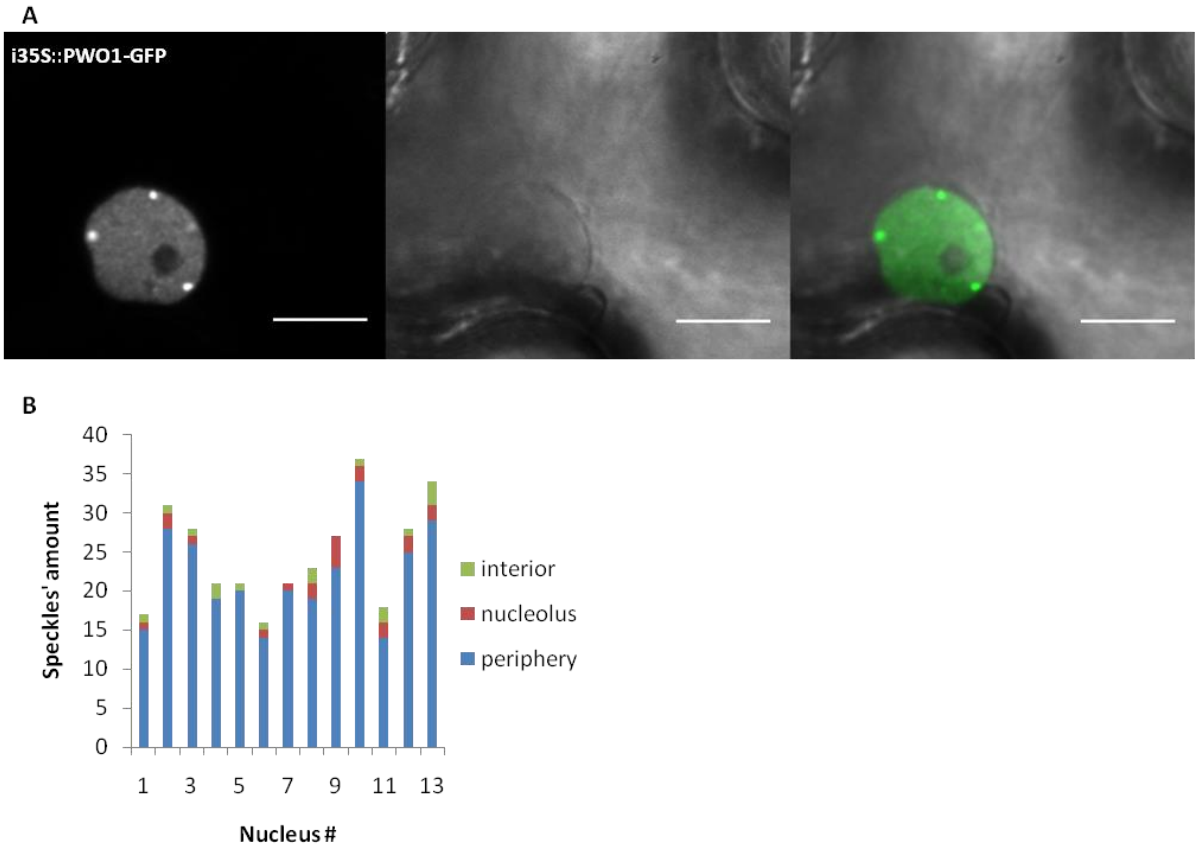



**Fig.5. PWO1 localization to speckles in *Nicotiana benthamiana*.** Leaves of *N. benthamiana* were infiltrated with beta-estradiol-inducible i35S::PWO-GFP construct. The images of epidermis nuclei were acquired 16-20

hours post induction. Raw images were filtered by Gaussian blur (sigma = 0.75). Scale bar corresponds to 10 µm. (B) The number of speckles and their localization per individual nucleus.

Next, we acquired confocal microscopy images from stable *PWO1::PWO1-GFP Arabidopsis* line. We observed localization to the nucleoplasm and subnuclear speckles in cortex and epidermis tissues of root meristematic zone (Fig.6A,C). However, due to the small size of meristematic nuclei, it remains unclear whether the speckles locate preferentially at the nuclear periphery and/or the nucleolus. In contrast, imaging of nuclei from epidermis of root elongation zone and leaves revealed a uniform nucleoplasmic distribution without speckles' formation (Fig.S4, Fig.S6A)]. In summary, we showed that PWO1 does not localize exclusively to the nuclear periphery in *Arabidopsis* or *N. benthamiana*; instead it occupies a whole nucleoplasmic space and forms partially peripheral subnuclear speckles in a tissue-specific manner.

Furthermore, we asked if disruption of *PWO1* or *CRWN1* affects localization of each other. Our confocal microscopy analyses of the lines: *PWO1::PWO1-GFP crwn1* and *CRWN1::CRWN1-GFP pwo1 pwo3* revealed no changes in subnuclear localization comparing to the control lines (Fig.6, Fig.S5, Fig.S6). Therefore, we concluded that the subnuclear localizations of PWO1 and CRWN1 are independent of each other.

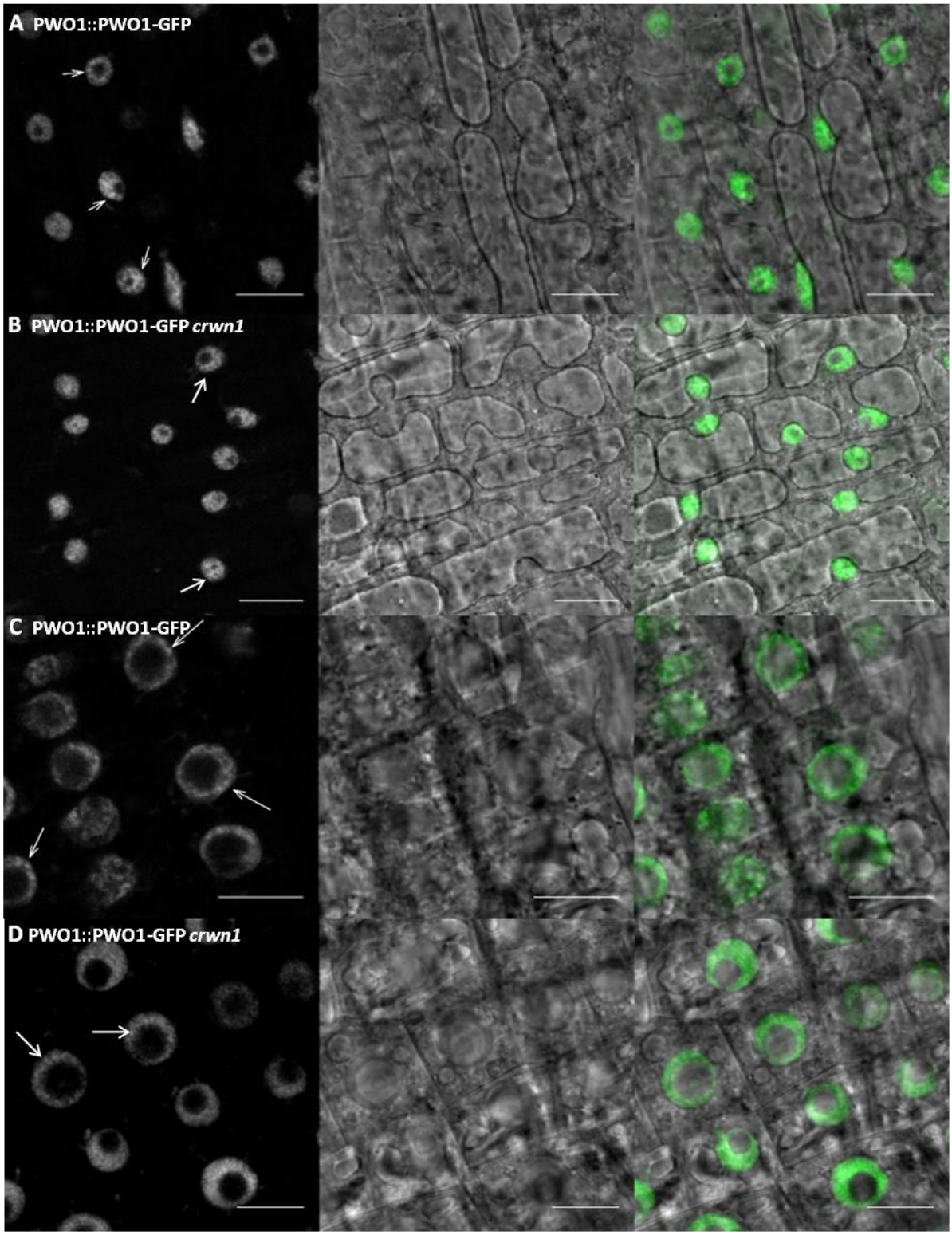



**Fig.6. PWO1 localization in *Arabidopsis*.** Confocal microscopy was performed on roots from the line: *PWO1::PWO1-GFP* in *crwn1* (B,D) and the control: *PWO1::PWO1-GFP* (A,C). The images represent different root tissues: meristematic zone epidermis (A and B), mersitematic zone cortex (C and D). The images were filtered using gaussian blur (sigma = 1). Scale bar corresponds to 10 µm.

**PWO1 and CRWN1 affect expression of a similar set of genes**

In order to further understand a connection of PWO1 and CRWN1, we sought to investigate global transcriptomic changes using the RNA-seq method. As a material, we used 2 week-old *Arabidopsis* seedlings from *pwo1* and strong *crwn1 crwn2* mutants. Differential expression analysis (see materials and methods) allowed to identify 179 significantly upregulated- and 242 significantly downregulated-differentially expressed genes (DEGs) in *pwo1*, compared to the wildtype control (Col-0). In turn, similar analysis on *crwn1 crwn2* revealed 602 significantly upregulated- and 214 significantly downregulated-DEGs. Cross-comparison of the datasets from both genotypes allowed to identify a significant overlap of 71 upregulated- and 73 downregulated-DEGs (Fig.7A,B). Such results show that *PWO1* and *CRWN1* regulate partially similar set of target genes.

Next, we asked whether common DEGs affected by both, *pwo1* and *crwn1/2*,can be organized into functional categories. Our gene ontology analysis (see materials and methods) on the overlap of upregulated DEGs showed significant enrichment of the functions related to: response to stress (biotic and abiotic), iron homeostasis and transport of amino acids/nitrate (Fig.7C). In turn, DEGs from the overlap of downregulated targets are involved specifically in the response to auxin.

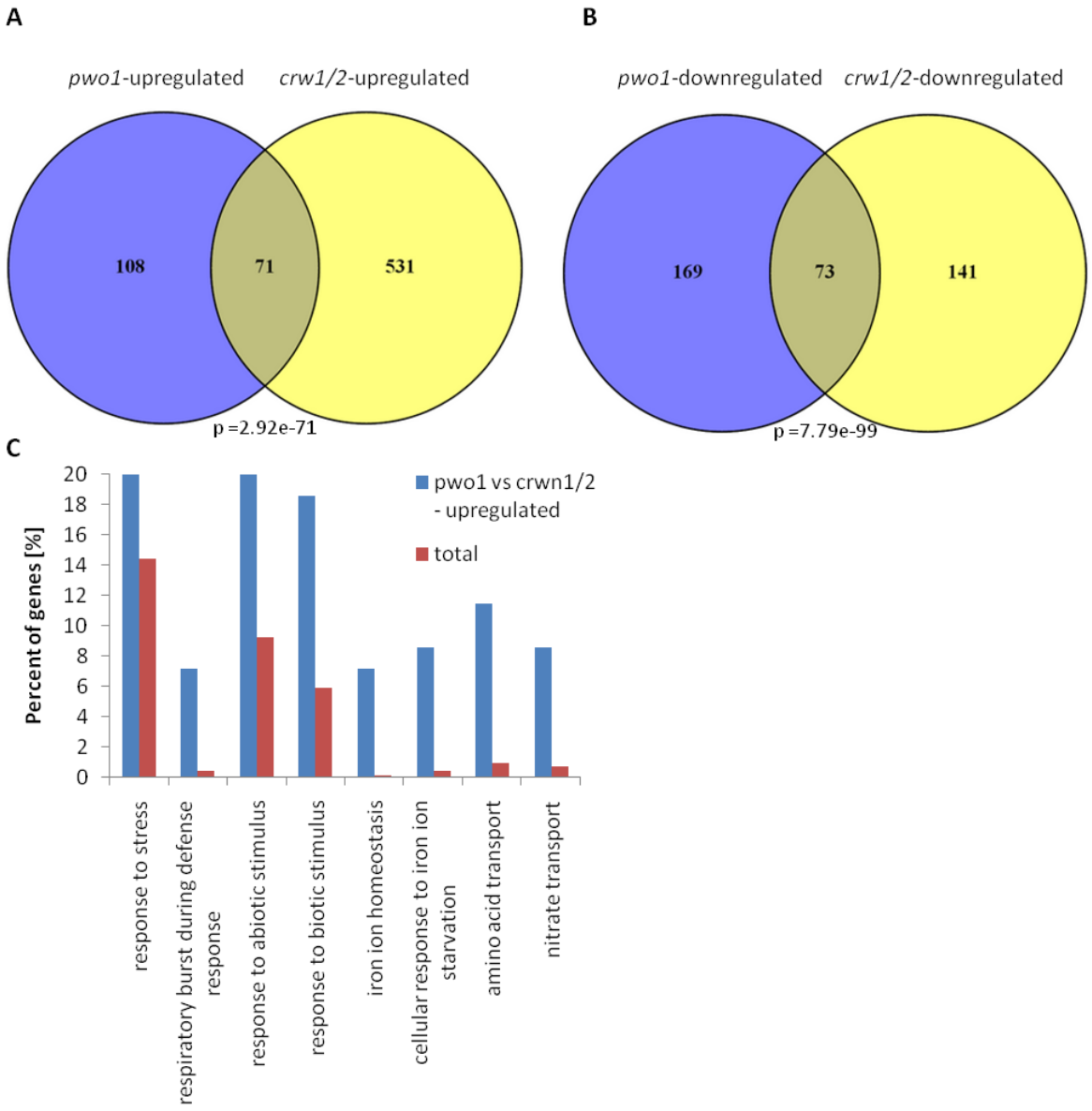
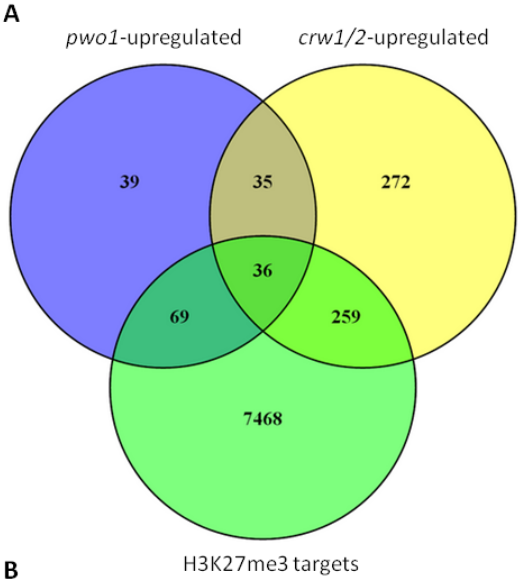



**Fig.7. Differentially expressed gene (DEG) analysis – overlap between *pwo1* and *crwn1/2*.** (A) Venn diagram comparing upregulated-DEGs in *pwo1* (*pwo1*-upregulated) and *crwn1/2* (*crwn1/2*-upregulated). Significance level of the overlap was calculated in R using hypergeometric test. (B) Venn diagram comparing downregulated-DEGs in *pwo1* (*pwo1*-downregulated) and *crwn1/2* (*crwn1/2*-downregulated). Significance level of the overlap was calculated in R using hypergeometric test. (C) GO term enrichment analysis on commonly upregulated-DEGs in *pwo1* and *crwn1/2* compared to all TAIR10 protein-coding genes ("total"). The percentage corresponds to the number of genes enriched in significantly abundant (FDR<0.005) GO terms in the overlap shown in (A) or the whole protein-coding gene number based on TAIR10 annotation.

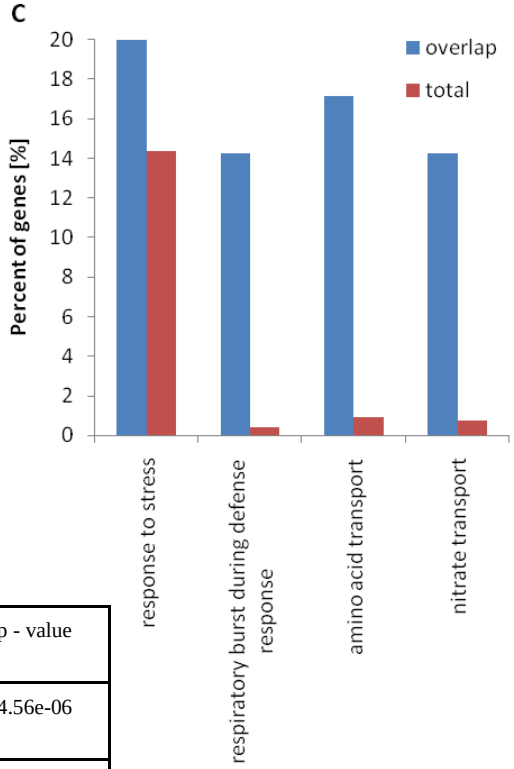
Given the PWO1 involvement in PRC2 pathway (Hohenstatt, 2012), we hypothesized that upregulated-DEGs in *pwo1* are covered by H3K27me3 in the wildtype plants, in which a functional PWO1 is present. Therefore, we compared our list of upregulated-DEGs in *pwo1* and published dataset of H3K27me3 targets (Oh et al., 2008). Indeed, we detected a

significant overlap of 105 genes present in both lists (Fig.8A,B), showing that PWO1 is required for the repression of a subset of PRC2 targets. As PWO1 associates with CRWN1, we asked whether mutants in CRWN gene family affect expression of H3K27me3 targets as well. Interestingly, the comparison between upregulated-DEGs in *crwn1/2* and H3K27me3-covered genes revealed a significant number of genes shared between two datasets (Fig.8A,B). Finally, cross-comparison of upregulated-DEGs in *pwo1* and *crwn1/2*, and H3K27me3 targets gave 36 genes shared by all three datasets (Fig.8A,B). Their function is related to stress response and amino acid/nitrate transport, but not to iron metabolism (Fig.8C), which implicates a role of PWO1 and CRWN1 in iron homeostasis facilitated in PRC2-independent manner. We noted also a substantial number of H3K27me3 targets which expression is affected only by either, PWO1 or CRWN1 CRWN2. Such a result suggests that PWO1 and CRWN1 mediate repression (directly or indirectly) of PRC2 targets also separately of each other.

A




C




B

| Comparison | p - value |
|---|---|
| H3K27me3 targets vs *pwo1*-upregulated | 4.56e-06 |
| H3K27me3 targets vs *crwn1/2*-upregulated | 2.20e-06 |
| H3K27me3/*pwo1*-upreg. vs H3K27me3/*crwn1/2* upreg. | 5.64e-28 |

**Fig.8. Differentially expressed gene (DEG) analysis – cross-comparison between mutants and H3K27me3 targets.** (A) Venn diagram comparing upregulated-DEGs in *pwo1* (*pwo1*-upregulated), in *crwn1/2* (*crwn1/2*-upregulated) and H3K27me3 targets in Col-0 (Oh et al., 2008). (B) Significance level of overlapping gene number shown in (A). P-values were calculated in R by hypergeometric test. (C) GO term enrichment analysis

on commonly upregulated DEGs in *pwo1* and *crwn1/2*, and H3K27me3 targets in Col-0, compared to all TAIR10 protein-coding genes ("total"). The percentage corresponds to the number of genes enriched in significantly abundant (FDR<0.0005) GO terms in the common target genes shared by all three datasets in (A) or the whole protein-coding gene number based on TAIR10 annotation.

### 3.1.4. Discussion

Here, we demonstrate that the *Arabidopsis* PRC2-associated protein PWO1 interacts physically with numerous nuclear lamina components, including CRWN1 – a coiled-coil analog of lamin proteins in *Arabidopsis*. We report that *PWO1* and *CRWN1* are epistatic in controlling nuclear size and affect expression of a similar set of PRC2 target genes. Moreover, we show that PWO1 partially localizes to the speckles located at the nuclear periphery.

Our findings suggest a conservation of the association between a repressive chromatin environment and the nuclear periphery across biological kingdoms, including plants. Strikingly, the exact mechanism and linking factors involved in this phenomenon are frequently species-specific, as has been reported for, i.e. *C. elegans*-specific CEC4 (Gonzalez-Sandoval et al., 2015) or Sir4, present in *Saccharomycetacea e*only (Hickman et al., 2011; Taddei et al., 2004). Identification of PWO1, a novel plant-specific bridging factor, adds another evolutionary-distinct element in otherwise widespread machinery linking gene silencing with the nuclear lamina.

The pathways linking repressive chromatin and NL vary not only between the species, but also within the same organism. Many protein effectors associated with NL in fungi or animals regulate only a specific subset of targets and play roles also outside of the nuclear periphery (Harr et al., 2016). Consequently, our results on PWO1 subnuclear localization, transcriptomic analyses in *pwo1* and *crwn1/2* and whole nucleoplasmic distribution of H3K27me3 in plants (Mathieu et al., 2005) collectively suggest a NL- and PWO1-dependent repression of only a specific set of PRC2 targets. We speculate that the other, PWO1-independent, pathways bridging NL and PRC2 also exist, as suggested by the PWO1-independent CRWN1/2-mediated repression of PRC2 target genes. Needless to say, an identification of such pathways will be the essential step for the future. The feature of altered nuclear size in mutants of stable NL components (like *crwn1*) and its associated members (like *pwo1*) provides a measurable phenotype for unbiased mutagenic screens (Goto et al., 2014).

Target-specificity of NL-chromatin bridging pathways is demonstrated also by the phenotypes of respective mutants. Lack of bridging factors is frequently represented by a normal viability and only mild aberrations in standard conditions, with a substantial effect seen only at selected developmental stages or environments (Gonzalez-Sandoval et al., 2015; Ozawa et al., 2006; Towbin et al., 2012). Similarly, the *pwo1* mutant shows general phenotype indistinguishable from the wildtype, with obvious changes in one specific trait only (flowering time) (Hohenstatt, 2012). As disruption of all PWO family members in *pwo1/2/3* displays severe effect (seedling lethality) (Hohenstatt, 2012), we cannot exclude that the mild phenotype in *pwo1* single mutant is caused by the redundancy between PWO family proteins. However, a nuclear area reduction seen already in *pwo1* single mutant and the absence of conserved CRWN1-interacting motif in PWO2 and PWO3 suggest that PWO1 has a NL-connected role separate from the other PWO family members.

Furthermore, our results on PWO1 subnuclear localization showed formation of peripheral and nucleolar speckles in *N. benthamiana* and whole nucleoplasmic speckles in *A. thaliana*. Formation of distinguishable nucleoplasmic foci is a prominent feature of PRC1 proteins in animals (Pirrotta and Li, 2012) and was reported to be dependent on PRC2 activity (Hernández-Muñoz et al., 2005). Moreover, animal H3K27me3 is found in large aggregates at the nuclear periphery if associated with late-replicating chromatin (Hernández-Muñoz et al., 2005). Similarly to animal models, PRC1 members in *Arabidopsis* were found to localize into multiple speckles as well (Calonje et al., 2008; Libault et al., 2005), albeit without any preference toward nuclear periphery. Moreover, using genomic methods, H3K27me3 was shown to be enriched in the chromatin associated with nucleolus in mammals (Németh et al., 2010) and plants (Pontvianne et al., 2016). It remains to be elucidated whether PWO1 acts together with PRC1 complex, if PRC1 and PWO1 speckles are functionally connected and whether PWO1 has a role in the organization of nucleolar chromatin.

Our transcriptomic analyses, presented together with the nuclear size phenotype, highlight another important point. As reduced nuclear area causes elevated frequency of interchromosomal contacts (Grob et al., 2014), changes in the gene expression in *pwo1* and *crwn1/2* might reflect indirect effects coming from altered spatial organization of the chromatin. Thus, a generation and cross-comparison of transcriptomic data from the other mutants affected in nuclear size should be performed to de-couple gene-specific influence from the indirect effects.

In short, we provide a putative link between nuclear lamina and H3K27me3-mediated gene repression in plants, highlighting an importance of nuclear architecture proteins in chromatin-mediated control of gene expression across biological kingdoms. Deepening our understanding of PWO1 serves as a fascinating opportunity to decipher mechanistic in the association between gene silencing and nuclear periphery.

### 3.1.5. Acknowledgements

We would like to cordially thank Prof. Dr. Ruediger Simon (HHU Duesseldorf, Germany) for providing FRET vectors and Keygene N.V. company (Wageningen, Netherlands) for the training in RNA-seq analysis. We thank Dr. Julia Kleinmanns, Dorota Komar and Suraj Jamge for critical revision of the manuscript. This work was supported by European Commission 7$^{th}$ FP project EpiTRAITS, Boehringer Ingelheim Foundation and the CRC973.

### 3.1.6. Materials & methods

#### Plant material

*A. thaliana* seedlings were grown on ½ MS plates in long day conditions (22°C) and transferred to soil, if later stages of development were needed. For the knock-out mutations, following transgenic lines were used: *pwo1-1 (*SAIL_342_C09), *crwn-1* (SALK_025347), *crwn1-1 crwn2-1* (SALK_025347, SALK_076653), *swn-7* (SALK_109121). For transient assays, *N. benthamiana* plants were grown in soil in long day conditions (22°C) up to 4$^{th}$ week and subsequently used for infiltration. The sequences of oligonucleotides used for genotyping are as described in: (Hohenstatt, 2012) and (Wang et al., 2013).

#### Cloning

Regarding the vectors used for FRET-APB, pMDC7 derivatives containing GFP, mCherry and GFP+mCherry coding sequences were created as described in: (Bleckmann et al., 2010). Next, genomic sequence of *CRWN1* (AT1G67230), coding region of full *PWO1* sequence (AT3G03140) or coding region of PWO1 C-terminal fragment were ligated into pCR8-GW-TOPO using TA Cloning Kit (Thermo Fisher) to create entry vectors. To construct

translational fusion plasmids, insert sequences from the entry vectors were introduced into pMDC7 derivatives via LR reaction from Gateway cloning protocol (Thermo Fisher). For yeast-two-hybrid experiments, CRWN1, PWO1 and PWO1 fragments' coding sequences were used to create entry vectors via TA Cloning and subsequently introduced into pGADT7 and pGBKT7 Y2H vectors (Clontech) using LR reaction (Thermo Fisher).

**CoIP-MS/MS**

Nuclear proteins were isolated from 3g of whole seedlings from *PWO1::PWO1-GFP* or Col-0 (negative control) lines grown in LD conditions. 4 biological replicates per genotype were used. Nuclei were extracted without prior fixation of the tissue as described in: (Kaufmann et al., 2010) and the further procedure was performed as described in: (Smaczniak et al., 2012). For immunoprecipitation, µMACS GFP Isolation Kit (Miltenyi Biotec) was used. Peptide spectra were obtained on LTQ Orbitrap XL mass spectrometer (Thermo Scientific). Protein identification was performed by the database search with Andromeda search engine (Cox et al., 2011). Raw data processing was done on MaxQuant v1.1.1.36 software (Cox and Mann, 2008). To distinguish PWO1-interacting proteins from the background, Student's t-test with FDR adjustment was performed on label-free quantification (LFQ) values.

**Subnuclear localization and FRET-APB**

For FRET-APB and examination of subnuclear localization in *Nicotiana benthamiana*, estradiol-inducible pMDC7 vector derivatives were transformed into Agrobacterium tumefasciens (GV3101 PMP90 strain with p19 silencing suppressor plasmid) and grown on YEB medium plates for 2 days. Bacterial lawn was scraped into infiltration medium (5% sucrose, 0,1% Silwet-L77, 450µM acetosyringone) to OD = 0,8 and kept on ice for 1 hour. Bacteria were infiltrated into 4-week old *N. benthamiana* leaves using 1ml-suringes without needles. After 48 hours, an induction was performed by painting abaxial side of the leaves with 20-50µM Beta-estradiol solution in 0,1% Tween-20. Confocal miscroscopy was done 18-24 hours post-induction on LSM780 (Carl Zeiss) or SP8 (Leica). For FRET-APB, GFP was excited at 488 nm with argon laser and mCherry at 561 nm with helium laser. Photobleaching was performed using 561nm laser in 5 frames with 100% power on Leica SP8. FRET efficiency was calculated as described in: (Bleckmann et al., 2010).

Subnuclear localization of candidate proteins in *Arabidopsis* was done by confocal microscopy on seedling of stable transgenic lines using same laser parameters as described above. Images were acquired with 3-4 line averaging.

**Yeast-two-hybrid**

Yeast cultures were grown at 28°C on YPD on selective SD media. AH109 strain was used for transformation, following protocol described in Yeast Protocols Handbook (Clontech Laboratories, Inc .Version No. PR973283 21), with both, Gal4-BD and Gal4-AD, constructs added. Transformants were selected on SD medium lacking tryptophan and leucine (SD-LW). Protein interaction was assessed by the transformant growth on selective SD media additionally lacking histidine (SD-LWH).

**Nuclear morphology analysis**

0.5g of 2 week-old seedlings were fixed in 10mL freshly prepared ice-cold 4% formaldehyde in PBS buffer for 20 min. under vacuum. The seedlings were washed 3 times for 5 min in PBS buffer. After removal of PBS, the material was chopped on ice with razor blade in 50µL Nuclear Isolation Buffer (NIB: 500mM sucrose, 100mM KCl, 10mM Tris-HCl pH 9.5, 10mM EDTA, 4mM spermidine, 1mM spermine, 0.1%(v/v) Beta-mercaptoethanol). Additional 450 µL NIB was added and cellular solution was filtered through 50 µm cell strainers (#04-0042-2317, Partec). The filtrate was centrifuged at 500g, for 3 min, 4°C. The pellet was resuspended in 40 µL NIB and 2-3 µL was spread on microscopic slide, left to dry and mixed subsequently with 4 µg/mL DAPI solution (#6335, Roth) in Vectashield mounting medium (#H-1000, Vector Laboratories). DAPI-stained nuclei were visualized in Z-stacks done on SP8 confocal microscope (Leica) using 405 nm diode. Acquired Z-stacks were manually thresholded and used for nuclear shape and size measurements in Fiji implementation of ImageJ2 (Schindelin et al., 2012). Nuclear area and circularity index measurements were performed on Z plane showing the largest area of particular nucleus in the Z-stack. Circularity index was calculated as: $4\pi(\text{area}/\text{perimeter}^2)$, with maximal value of 1.0 corresponding to the perfect circle. The measurements were done on 72-92 nuclei per genotype.

## Transcriptomics

RNA was extracted from 2-week old seedlings using RNeasy Plant Mini Kit (Qiagen, #74904), according to manufacturer's manual. RNA was treated with DNase and its quality was assessed on Nanodrop 1000 (Thermo Fisher Scientific), by Qubit RNA BR Assay (Thermo Fisher Scientific, #Q12210) or using RNA 6000 Nano Kit (Agilent, #5067-1511) on 2100 Bioanalyzer (Agilent). First strand cDNA was synthesized using RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, #K1622) with oligo-d(T) primers, according to manufacturer's manual. cDNA was sent for library preparation and sequencing on HiSeq 2000 (Illumina) at BGI (www.genomics.cn). Alternatively, cDNA was used for RT-qPCR with KAPA SYBR FAST Master Mix (Kapa Biosystems, #KK4600) on iQ5 detection system (Biorad). Expression levels were calculated by application of ΔΔCt method. Sequences of oligonucleotides are shown in Table S2. All the transcriptomic analyses were done in 3 or 5 independent biological replicates for RNA-seq or RT-qPCR samples, respectively.

## RNA-seq analysis

2 x90bp paired-end raw sequencing reads were scored by quality, cleaned and trimmed from 5' and 3'ends in Trimmomatic v0.35 and FastQC v0.10. Clean reads were aligned to TAIR10 reference genome by TopHat v2.0.12 with mate inner distance (-r ) = 20bp, segment length = 30bp and minimal intron length (-i) = 20bp, according to: (Trapnell et al., 2009). Alignment files were processed in two different pipelines: 1) Cufflinks (Tuxedo protocol, (Trapnell et al., 2013)) and 2) EdgeR (Robinson et al., 2010). 1) FPKM values from mapped reads were calculated in Cufflinksv2.2.1 with enabled: reference annotation (-g), multi read correction (-u), fragment bias correction (-b) and minimal intron length set to 20 bp. Cufflinks output was used to create common transcript reference file in Cuffmerge v2.2.1.0 and perform DEG analysis in Cuffdiff v2.2.1 with multi read correction and fragment bias correction. 2) Mapped reads were transformed into counts using HTSeq v0.6.1 and used as input for edgeR v3.3. Features with less than 1 count per mln were discarded. For remaining features differential expression was computed with adjusted p-values <0.05. DEG analysis in either of pipelines concerned comparison between wildtype and mutant (*pwo1* or *crwn1/2)* samples. In order to ensure the stringency of the analysis, only those genes that were present in DEG lists from both pipelines were taken for subsequent steps. The top DEGs in *pwo1* were

validated by RT-qPCR using 5 independent biological replicates (Fig.S7). The oligonucleotides used for the validation are depicted in Table S2.

Correlation between replicates (Fig.S8) was done based on the matrix containing "counts per million" values calculated in HTSeq v0.6.1. Features with less than 1 count per mln were discarded. Heatmap with correlation between replicates was obtained in Heatmap3 v1.1.1 using default methods for distance computing and dendrogram re-ordering.

**Secondary bioinformatic analyses**

Gene ontology of DEGs was inferred from Singular Enrichment Analysis on AgriGO server (Du et al., 2010). Statistical significance was calculated using Fischer test with Yekutieli adjustment method and the threshold of FDR<0.01 was applied. *Arabidopsis thaliana* TAIR10 genes were used as a reference.

For cross-comparison of DEGs in different mutant backgrounds, Venn diagrams were created in Venny v2.1 software (Oliveros, 2007). Statistical significance of overlapping gene number was calculated in R using hypergeometric test.

### 3.1.7. References


Bleckmann, A., Weidtkamp-Peters, S., Seidel, C. a M., and Simon, R. (2010). Stem cell signaling in Arabidopsis requires CRN to localize CLV2 to the plasma membrane. *Plant Physiol.* 152, 166–76. doi:10.1104/pp.109.149930.

Calonje, M., Sanchez, R., Chen, L., and Sung, Z. R. (2008). EMBRYONIC FLOWER1 participates in polycomb group-mediated AG gene silencing in Arabidopsis. *Plant Cell* 20, 277–291. doi:10.1105/tpc.106.049957.

Ciska, M., and Moreno Diaz de la Espina, S. (2013). NMCP/LINC proteins: putative lamin analogs in plants? *Plant Signal. Behav.* 8, e26669. doi:10.4161/psb.26669.

Ciska, M., and Moreno Díaz de la Espina, S. (2014). The intriguing plant nuclear lamina. *Front. Plant Sci.* 5, 166. doi:10.3389/fpls.2014.00166.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372. doi:10.1038/nbt.1511.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011). Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* 10, 1794–1805. doi:10.1021/pr101065j.

Dechat, T., Adam, S. A., Taimen, P., Shimi, T., and Goldman, R. D. (2010). Nuclear lamins. *Cold Spring Harb. Perspect. Biol.* 2, a000547. doi:10.1101/cshperspect.a000547.

Dittmer, T. A., Stacey, N. J., Sugimoto-Shirasu, K., and Richards, E. J. (2007). LITTLE NUCLEI genes affecting nuclear morphology in Arabidopsis thaliana. *Plant Cell* 19, 2793–803. doi:10.1105/tpc.107.053231.

Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: A GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* 38. doi:10.1093/nar/gkq310.

Eberhart, A., Feodorova, Y., Song, C., Wanner, G., Kiseleva, E., Furukawa, T., et al. (2013). Epigenetics of eu- and heterochromatin in inverted and conventional nuclei from mouse retina. *Chromosom. Res.* 21, 535–554. doi:10.1007/s10577-013-9375-7.

Ebert, A., Schotta, G., Lein, S., Kubicek, S., Krauss, V., Jenuwein, T., et al. (2004). Su(var) genes regulate the balance between euchromatin and heterochromatin in Drosophila. *Genes Dev.* 18, 2973–2983. doi:10.1101/gad.323004.

Fakhouri, T. H. I., Stevenson, J., Chisholm, A. D., Mango, S. E., Mann, R., Carroll, S., et al. (2010). Dynamic Chromatin Organization during Foregut Development Mediated by the Organ Selector Gene PHA-4/FoxA. *PLoS Genet.* 6, e1001060. doi:10.1371/journal.pgen.1001060.

Ferrari, K. J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., et al. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol. Cell* 53, 49–62. doi:10.1016/j.molcel.2013.10.030.

Gardiner, J., Overall, R., and Marc, J. (2011). Putative Arabidopsis homologues of metazoan coiled-coil cytoskeletal proteins. *Cell Biol. Int.* 35, 767–74. doi:10.1042/CBI20100719.

Gonzalez-Sandoval, A., Towbin, B. D., Kalck, V., Cabianca, D. S., Gaidatzis, D., Hauer, M. H., et al. (2015). Perinuclear Anchoring of H3K9-Methylated Chromatin Stabilizes Induced Cell Fate in C. elegans Embryos. *Cell* 163, 1333–1347. doi:10.1016/j.cell.2015.10.066.

Goto, C., Tamura, K., Fukao, Y., Shimada, T., and Hara-Nishimura, I. (2014). The Novel Nuclear Envelope Protein KAKU4 Modulates Nuclear Morphology in Arabidopsis. *Plant Cell* 26, 2143–2155. doi:10.1105/tpc.113.122168.

Graumann, K. (2014). Evidence for LINC1-SUN associations at the plant nuclear periphery. *PLoS One* 9, e93406. doi:10.1371/journal.pone.0093406.

Grob, S., Schmid, M. W., and Grossniklaus, U. (2014). Hi-C Analysis in Arabidopsis Identifies the KNOT, a

Structure with Similarities to the flamenco Locus of Drosophila. *Mol. Cell* 55, 678–693. doi:10.1016/j.molcel.2014.07.009.

Hänzelmann, S., Beier, F., Gusmao, E. G., Koch, C. M., Hummel, S., Charapitsa, I., et al. (2015). Replicative senescence is associated with nuclear reorganization and with DNA methylation at specific transcription factor binding sites. *Clin. Epigenetics* 7, 19. doi:10.1186/s13148-015-0057-5.

Harr, J. C., Gonzalez-Sandoval, A., and Gasser, S. M. (2016). Histones and histone modifications in perinuclear chromatin anchoring: from yeast to man. *EMBO Rep.* 17, e201541809. doi:10.15252/embr.201541809.

Harr, J. C., Luperchio, T. R., Wong, X., Cohen, E., Wheelan, S. J., and Reddy, K. L. (2015). Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. 208, 33–52. doi:10.1083/jcb.201405110.

Hennig, L., and Derkacheva, M. (2009). Diversity of Polycomb group complexes in plants: same rules, different players? *Trends Genet.* 25, 414–23. doi:10.1016/j.tig.2009.07.002.

Hernández-Muñoz, I., Taghavi, P., Kuijl, C., Neefjes, J., and van Lohuizen, M. (2005). Association of BMI1 with polycomb bodies is dynamic and requires PRC2/EZH2 and the maintenance DNA methyltransferase DNMT1. *Mol. Cell. Biol.* 25, 11047–58. doi:10.1128/MCB.25.24.11047-11058.2005.

Hickman, M. A., Froyd, C. A., and Rusche, L. N. (2011). Reinventing heterochromatin in budding yeasts: Sir2 and the origin recognition complex take center stage. *Eukaryot. Cell* 10, 1183–92. doi:10.1128/EC.05123-11.

Hohenstatt, M. L. (2012). Functional analysis of SCI1 – A PWWP domain protein involved in Polycomb group mediated gene regulation in Arabidopsis.

Ikegami, K., Egelhofer, T. a, Strome, S., and Lieb, J. D. (2010). Caenorhabditis elegans chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2. *Genome Biol.* 11, R120. doi:10.1186/gb-2010-11-12-r120.

Jacob, Y., and Michaels, S. D. (2009). H3K27me1 is E(z) in animals, but not in plants. *Epigenetics* 4, 366–369. doi:10.4161/epi.4.6.9713.

Kaufmann, K., Muiño, J. M., Østerås, M., Farinelli, L., Krajewski, P., and Angenent, G. C. (2010). Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). *Nat. Protoc.* 5, 457–472. doi:10.1038/nprot.2009.244.

Libault, M., Tessadori, F., Germann, S., Snijder, B., Fransz, P., and Gaudin, V. (2005). The Arabidopsis LHP1 protein is a component of euchromatin. *Planta* 222, 910–925. doi:10.1007/s00425-005-0129-4.

Luo, L., Gassman, K. L., Petell, L. M., Wilson, C. L., Bewersdorf, J., and Shopland, L. S. (2009). The nuclear periphery of embryonic stem cells is a transcriptionally permissive and repressive compartment. *J. Cell*

*Sci.* 122.

Mathieu, O., Probst, A. V, Paszkowski, J., Bartee, L., Malagnac, F., Bender, J., et al. (2005). Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. *EMBO J.* 24, 2783–91. doi:10.1038/sj.emboj.7600743.

Meister, P., Towbin, B. D., Pike, B. L., Ponti, A., and Gasser, S. M. (2010). The spatial dynamics of tissue-specific promoters during C. elegans development. *Genes Dev.* 24, 766–82. doi:10.1101/gad.559610.

Melcer, S., Gruenbaum, Y., and Krohne, G. (2007). Invertebrate lamins. *Exp. Cell Res.* 313, 2157–2166. doi:10.1016/j.yexcr.2007.03.004.

Németh, A., Conesa, A., Santoyo-Lopez, J., Medina, I., Montaner, D., Péterfia, B., et al. (2010). Initial Genomics of the Human Nucleolus. *PLoS Genet.* 6, e1000889. doi:10.1371/journal.pgen.1000889.

Oh, S., Park, S., and van Nocker, S. (2008). Genic and global functions for Paf1C in chromatin modification and gene expression in Arabidopsis. *PLoS Genet.* 4, e1000077. doi:10.1371/journal.pgen.1000077.

Oliveros, J. C. (2007). VENNY. An interactive tool for comparing lists with Venn Diagrams. Available at: http://bioinfogp.cnb.csic.es/tools/venny/index.html.

Ozawa, R., Hayashi, Y. K., Ogawa, M., Kurokawa, R., Matsumoto, H., Noguchi, S., et al. (2006). Emerin-lacking mice show minimal motor and cardiac dysfunctions with nuclear-associated vacuoles. *Am. J. Pathol.* 168, 907–17. doi:10.2353/ajpath.2006.050564.

Pecinka, A., Schubert, V., Meister, A., Kreth, G., Klatte, M., and Lysak, M. A. (2002). Chromosome territory arrangement and homologous pairing in nuclei of Arabidopsis thaliana are predominantly random except for NOR-bearing chromosomes. 22, 258–269. doi:10.1007/s00412-004-0316-2.

Pirrotta, V., and Li, H.-B. (2012). A view of nuclear Polycomb bodies. *Curr. Opin. Genet. Dev.* 22, 101–9. doi:10.1016/j.gde.2011.11.004.

Pontvianne, F., Carpentier, M.-C., Durut, N., Pavlištová, V., Jaške, K., Schořová, Š., et al. (2016). Identification of Nucleolus-Associated Chromatin Domains Reveals a Role for the Nucleolus in 3D Organization of the A. thaliana Genome. *Cell Rep.* 16, 1574–87. doi:10.1016/j.celrep.2016.07.016.

Del Prete, S., Mikulski, P., Schubert, D., and Gaudin, V. (2015). One, two, three: Polycomb proteins hit all dimensions of gene regulation. *Genes (Basel).* 6, 520–542. doi:10.3390/genes6030520.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi:10.1093/bioinformatics/btp616.

Rosa, S., Lucia, F. De, Mylne, J. S., Zhu, D., Ohmido, N., Pendle, A., et al. (2013). Physical clustering of FLC alleles during Polycomb-mediated epigenetic silencing in vernalization. 1845–1850. doi:10.1101/gad.221713.113.Freely.

Rosin, F. M., Watanabe, N., Cacas, J.-L., Kato, N., Arroyo, J. M., Fang, Y., et al. (2008). Genome-wide transposon tagging reveals location-dependent effects on transcription and chromatin organization in Arabidopsis. *Plant J.* 55, 514–25. doi:10.1111/j.1365-313X.2008.03517.x.

Sakamoto, Y., and Takagi, S. (2013). LITTLE NUCLEI 1 and 4 regulate nuclear morphology in Arabidopsis thaliana. *Plant Cell Physiol.* 54, 622–33. doi:10.1093/pcp/pct031.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. doi:10.1038/nmeth.2019.

Schwartz, Y. B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. *Nat. Rev. Genet.* 8, 9–22. doi:10.1038/nrg1981.

Smaczniak, C., Immink, R. G. H., Muiño, J. M., Blanvillain, R., Busscher, M., Busscher-Lange, J., et al. (2012). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. *Proc. Natl. Acad. Sci. U. S. A.* 109, 1560–5. doi:10.1073/pnas.1112871109.

Taddei, A., Hediger, F., Neumann, F. R., Bauer, C., Gasser, S. M., Andrulis, E., et al. (2004). Separation of silencing from perinuclear anchoring functions in yeast Ku80, Sir4 and Esc1 proteins. *EMBO J.* 23, 1301–12. doi:10.1038/sj.emboj.7600144.

Towbin, B. D., González-Aguilera, C., Sack, R., Gaidatzis, D., Kalck, V., Meister, P., et al. (2012). Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell* 150, 934–47. doi:10.1016/j.cell.2012.06.051.

Towbin, B. D., Meister, P., Pike, B. L., and Gasser, S. M. (2010). Repetitive transgenes in C. elegans accumulate heterochromatic marks and are sequestered at the nuclear envelope in a copy-numberand lamin-dependent manner. *Cold Spring Harb. Symp. Quant. Biol.* 75, 555–565. doi:10.1101/sqb.2010.75.041.

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31, 46–53. doi:10.1038/nbt.2450.

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi:10.1093/bioinformatics/btp120.

Wang, H., Dittmer, T. a, and Richards, E. J. (2013). Arabidopsis CROWDED NUCLEI (CRWN) proteins are required for nuclear size control and heterochromatin organization. *BMC Plant Biol.* 13, 200. doi:10.1186/1471-2229-13-200.

Wu, F., Yao, J., Akhtar, A., Gasser, S., Dechat, T., Pfleghaar, K., et al. (2013). Spatial compartmentalization at the nuclear periphery characterized by genome-wide mapping. *BMC Genomics* 14, 591. doi:10.1186/1471-2164-14-591.

Xiao, J., Lee, U.-S., and Wagner, D. (2016). Tug of war: adding and removing histone lysine methylation in Arabidopsis. *Curr. Opin. Plant Biol.* 34, 41–53. doi:10.1016/j.pbi.2016.08.002.

### 3.1.8. Supplementary data

**Table S1. List of proteins from an overlap between *Arabidopsis* NL proteins and PWO1 interactors**

| TAIR ID | TAIR name | p-value | log2 ratio | Unique Peptides | Sequence Coverage [%] |
|---|---|---|---|---|---|
| AT1G09200 | Histone superfamily protein | 0.00994 | 2.296754 | 1 | 6.6 |
| AT1G09770 | ATCDC5, ATMYBCDC5, CDC5, cell division cycle 5 | 0.041435 | 6.60631 | 4 | 6.2 |
| AT1G10580 | Transducin/WD40 repeat-like superfamily protein | 0.004748 | 3.339939 | 2 | 3.7 |
| AT1G18450 | ARP4, ATARP4, actin-related protein 4 | 0.040975 | 12.24983 | 4 | 15 |
| AT1G20960 | emb1507, U5 small nuclear ribonucleoprotein helicase, putative | 0.00719 | 1.855619 | 22 | 10.9 |
| AT1G29940 | NRPA2, nuclear RNA polymerase A2 | 0.014381 | 0.641556 | 13 | 11.4 |
| AT1G48610 | AT hook motif-containing protein | 0.032085 | 4.83082 | 2 | 6.6 |
| AT1G51060 | HTA10, histone H2A 10 | 0.001402 | 5.861771 | 1 | 17.4 |
| AT1G59610 | ADL3, CF1, DL3, DRP2B, dynamin-like 3 | 0.046452 | 2.148804 | 2 | 2.2 |
| AT1G67230 | CRWN1, LINC1, little nuclei1 | 0.03123 | 3.633692 | 3 | 3.4 |
| AT1G68830 | STN7, STT7 homolog STN7 | 0.006244 | 5.808836 | 4 | 8.5 |
| AT1G77180 | SKIP, chromatin protein family | 0.002545 | 2.836021 | 4 | 8.2 |
| AT1G80070 | EMB14, EMB177, EMB33, SUS2, Pre-mRNA-processing-splicing factor | 0.005751 | 1.688779 | 40 | 19.7 |
| AT2G21390 | Coatomer, alpha subunit | 0.021772 | 5.920757 | 2 | 1.9 |
| AT2G23070 | Protein kinase superfamily protein | 0.005545 | 2.995675 | 2 | 6.2 |
| AT2G38040 | CAC3, acetyl Co-enzyme a carboxylase carboxyltransferase alpha subunit | 0.004266 | 3.526746 | 3 | 4.7 |
| AT3G03920 | H/ACA ribonucleoprotein complex, subunit Gar1/Naf1 protein | 0.024703 | 0.340992 | 7 | 40.1 |
| AT3G05060 | NOP56-like pre RNA processing ribonucleoprotein | 0.000527 | 0.543844 | 15 | 38.8 |
| AT3G08580 | AAC1, ADP/ATP carrier 1 | 0.015708 | 2.905601 | 5 | 14.4 |
| AT3G09790 | UBQ8, ubiquitin 8 | 0.030182 | 3.199069 | 3 | 8.2 |
| AT3G18790 | unknown protein | 0.000749 | 3.215798 | 2 | 7.3 |
| AT3G20670 | HTA13, histone H2A 13 | 0.01159 | 6.761234 | 1 | 17.4 |
| AT3G49910 | Translation protein SH3-like family protein | 0.010782 | 0.505465 | 7 | 37.7 |
| AT3G57150 | AtCBF5, AtNAP57, CBF5, NAP57, homologue of NAP57 | 0.000346 | 0.393501 | 25 | 58.6 |
| AT4G15900 | PRL1, pleiotropic regulatory locus 1 | 0.008488 | 4.553156 | 4 | 12.1 |
| AT4G31880 | Tudor/PWWP/MBT superfamily protein | 0.010285 | 54.23155 | 9 | 14.4 |
| AT4G40030 | Histone superfamily protein | 0.000228 | 3.3332 | 1 | 5.5 |
| AT5G02560 | HTA12, histone H2A 12 | 0.023285 | 5.036331 | 4 | 30.1 |
| AT5G04990 | ATSUN1, SUN1, SAD1/UNC-84 domain protein 1 | 0.003538 | 2.542025 | 2 | 7.6 |
| AT5G13490 | AAC2, ADP/ATP carrier 2 | 0.031099 | 2.762182 | 1 | 2.9 |
| AT5G14040 | PHT3;1, phosphate transporter 3;1 | 0.011107 | 3.035378 | 2 | 4.5 |
| AT5G14170 | CHC1, SWIB/MDM2 domain superfamily protein | 0.001985 | 16.22798 | 8 | 15.7 |

| | | | | | |
|---|---|---|---|---|---|
| AT5G18620 | CHR17, chromatin remodeling factor17 | 0.003764 | 12.0375 | 4 | 4.3 |
| AT5G19770 | TUA3, tubulin alpha-3 | 0.014529 | 2.815504 | 5 | 14.7 |
| AT5G22880 | H2B, HTB2, histone B2 | 0.002872 | 2.459329 | 5 | 39.3 |
| AT5G23060 | CaS, calcium sensing receptor | 0.005188 | 1.770383 | 5 | 16.8 |
| AT5G27120 | NOP56-like pre RNA processing ribonucleoprotein | 0.008346 | 0.732308 | 11 | 28 |
| AT5G27670 | HTA7, histone H2A 7 | 0.001726 | 4.943652 | 3 | 28 |
| AT5G28740 | Tetratricopeptide repeat (TPR)-like superfamily protein | 0.040115 | 7.453405 | 3 | 4.7 |
| AT5G47690 | PDS5, cohesion cofactor | 0.015028 | 4.566385 | 29 | 22.6 |
| AT5G52470 | ATFBR1, ATFIB1, FBR1, FIB1, SKIP7, fibrillarin 1 | 0.017328 | 0.629368 | 5 | 17.9 |
| AT5G58230 | ATMSI1, MEE70, MSI1, Transducin/WD40 repeat-like superfamily protein | 0.005131 | 8.083125 | 2 | 7.1 |
| AT5G59870 | HTA6, histone H2A 6 | 0.02551 | 2.641463 | 2 | 20 |
| AT5G63420 | emb2746, RNA-metabolising metallo-beta-lactamase family protein | 0.032657 | 1.955262 | 8 | 10.6 |
| AT5G64420 | DNA polymerase V family | 0.044173 | 2.564358 | 3 | 2.4 |
| AT5G65770 | CRWN4, LINC4, little nuclei4 | 0.000413 | 6.500275 | 4 | 5.2 |

The list of NL proteins was taken from: (Sakamoto and Takagi, 2013). The records corresponding to mitochondrial and ribosomal proteins were removed.

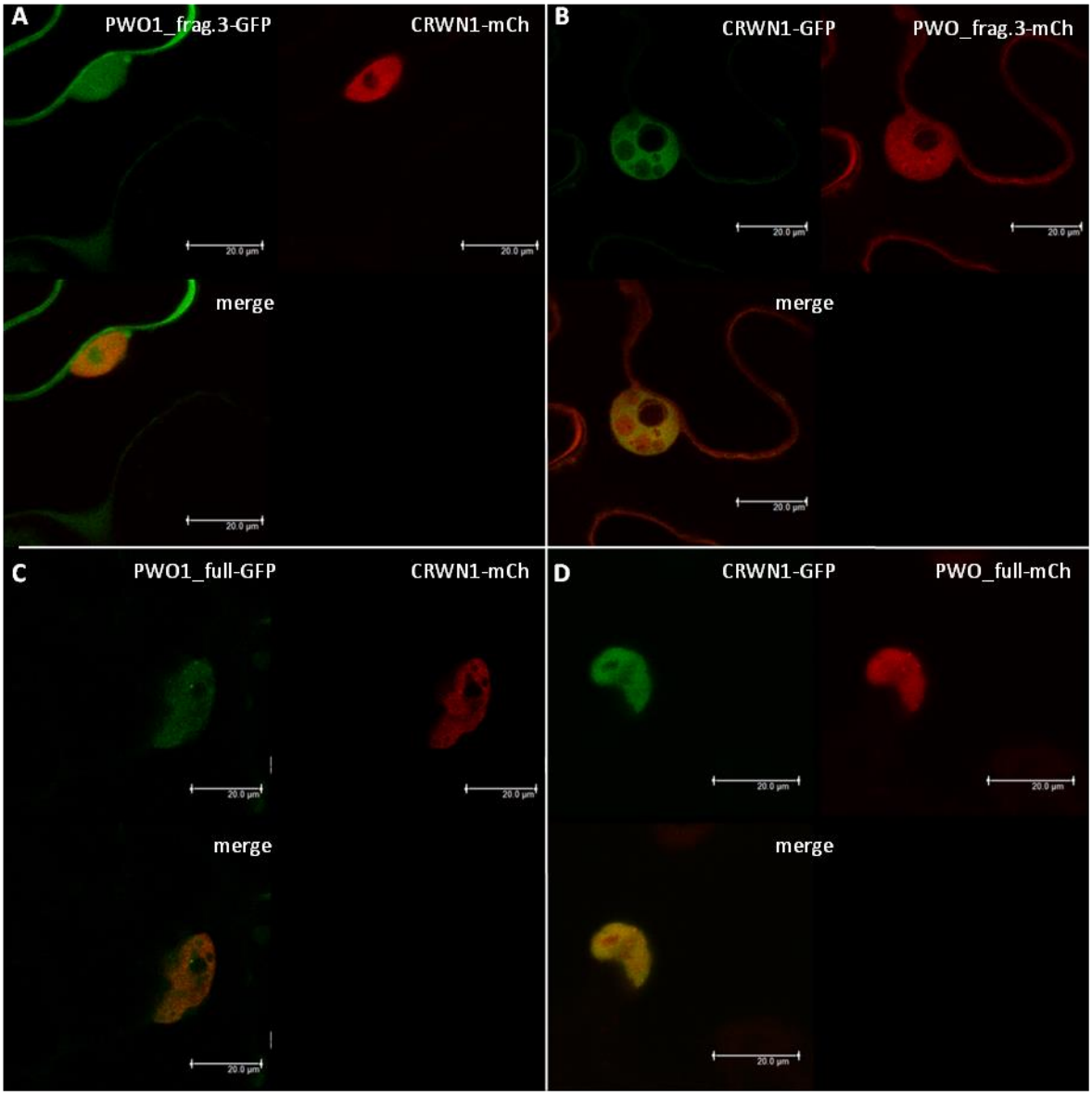



**Fig.S1. Subcellular localization of different constructs' combinations used for FRET-APB.** Beta-estradiol-inducible vectors were infiltrated to *N. benthamiana* leaves. Images were acquired 16-20 hours after induction. Scale bar corresponds to 20 µm. FRET-APB constructs' combinations: (A) PWO1_frag.3-GFP+CRWN1-mCh, (B) CRWN1-GFP+PWO_frag.3-mCh, (C) PWO1_full-GFP+CRWN1-mCh, (D) CRWN1-GFP+PWO1_full-GFP.

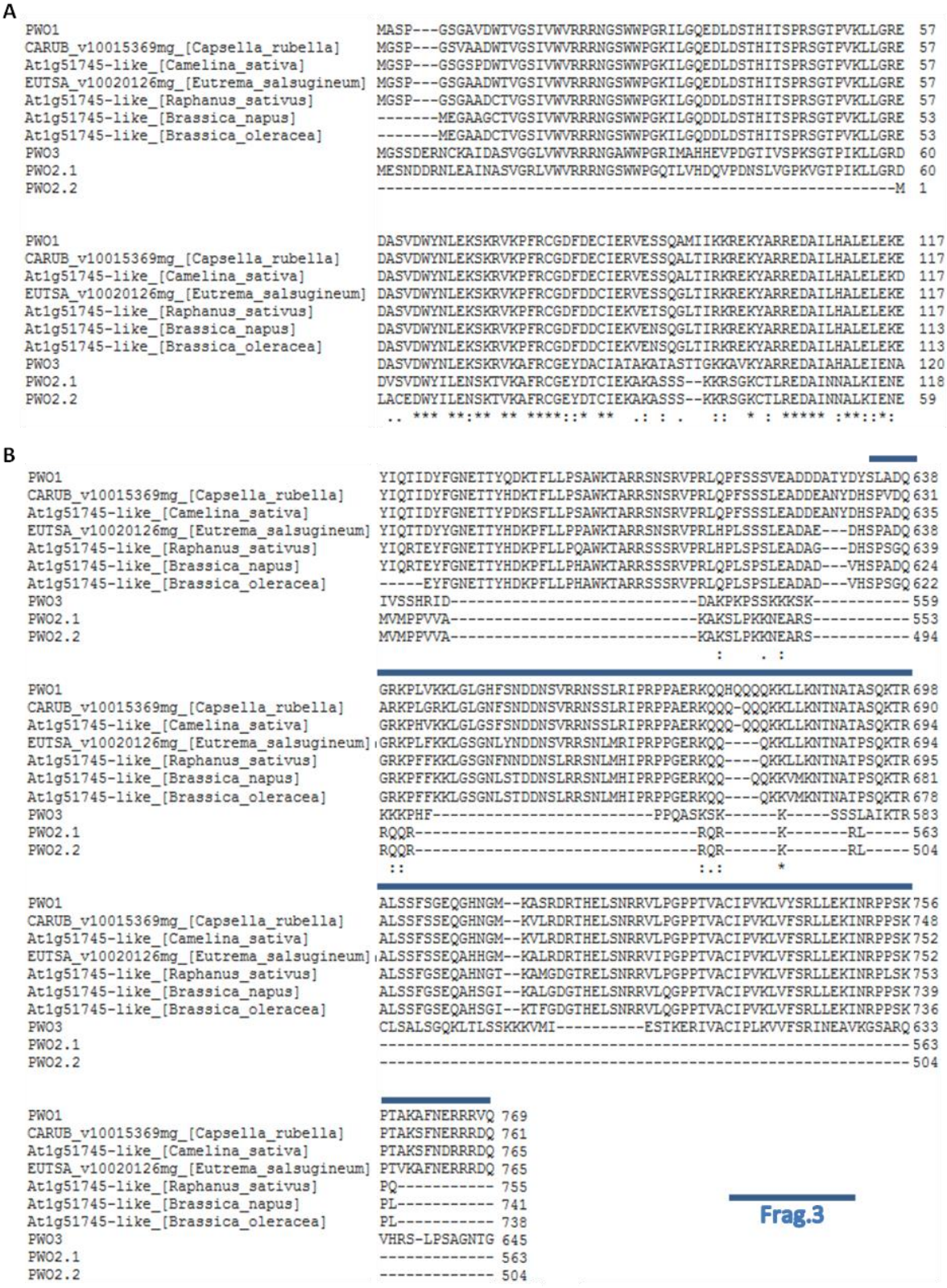



**Fig.S2. PWO1 multiple alignment.** A full PWO1 protein sequence was used as a query for BLAST-P search, the closest homologs were selected and aligned using Clustal Omega sofware. The figure represents a subset of the alignment corresponding to relatively variable C-terminal PWO1 sequence (B) and to highly conserved N-terminal PWO1 sequence (A), for comparison. CRWN1-interacting fragment (fragment 3) is depicted as a blue bar above the alignment. PWO2.1 and PWO2.2 correspond to sequences of two different PWO2 alternatively spliced isoforms.

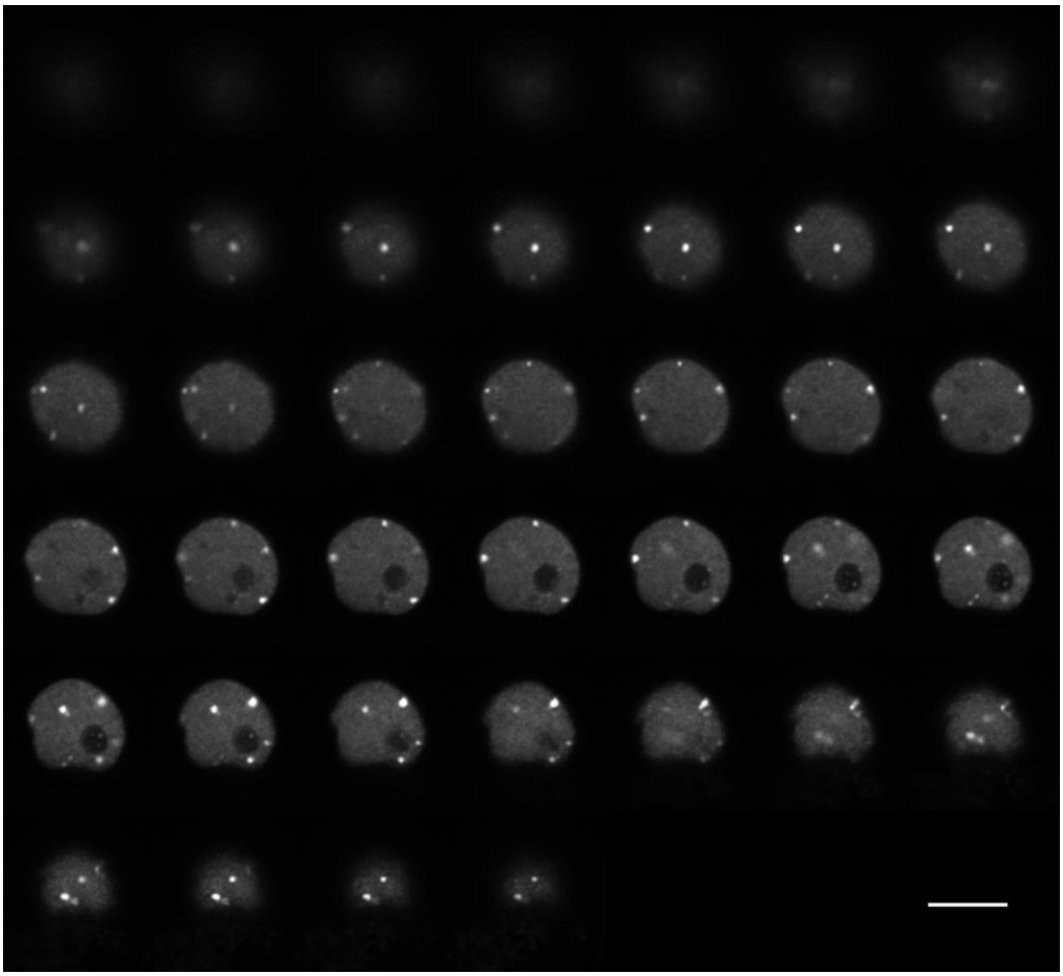


**Fig.S3. Z-stack montage of PWO1-GFP localization to speckles in *Nicotiana benthamiana*.** Leaves of *N. benthamiana* were infiltrated with estradiol-inducible *i35S::PWO1-GFP* construct. Z-stacks of epidermis nuclei were aquired 16-20 hours post induction. Raw Z-stacks were filtered by gaussian blur 3D (X sigma = 0.5, Y sigma = 0.5, Z sigma = 0.5) and visualized in grayscale. Scale bar corresponds to 10 µm.

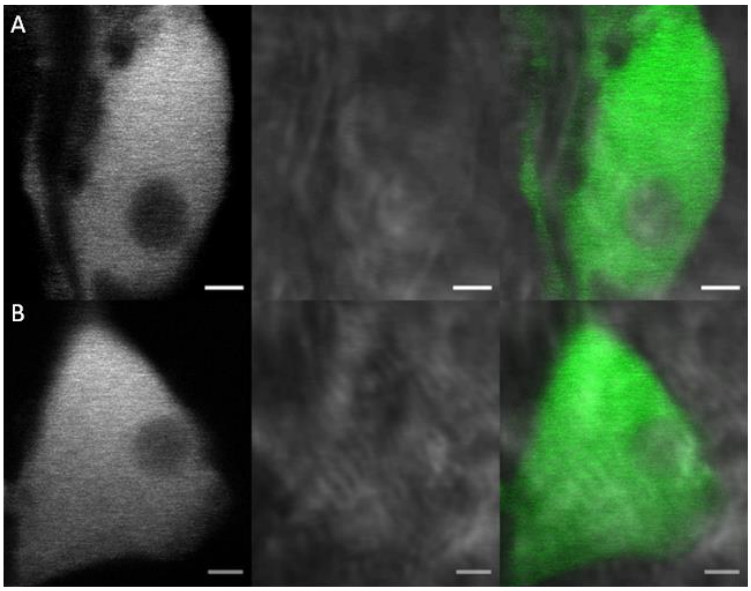



**Fig.S4. PWO1 localization in epidermis of *Arabidopsis thaliana* leaves.** Two representative nuclei (A and B) from *PWO1::PWO1-GFP* seedling leaves are shown. Scale bar corresponds to 20 µm.

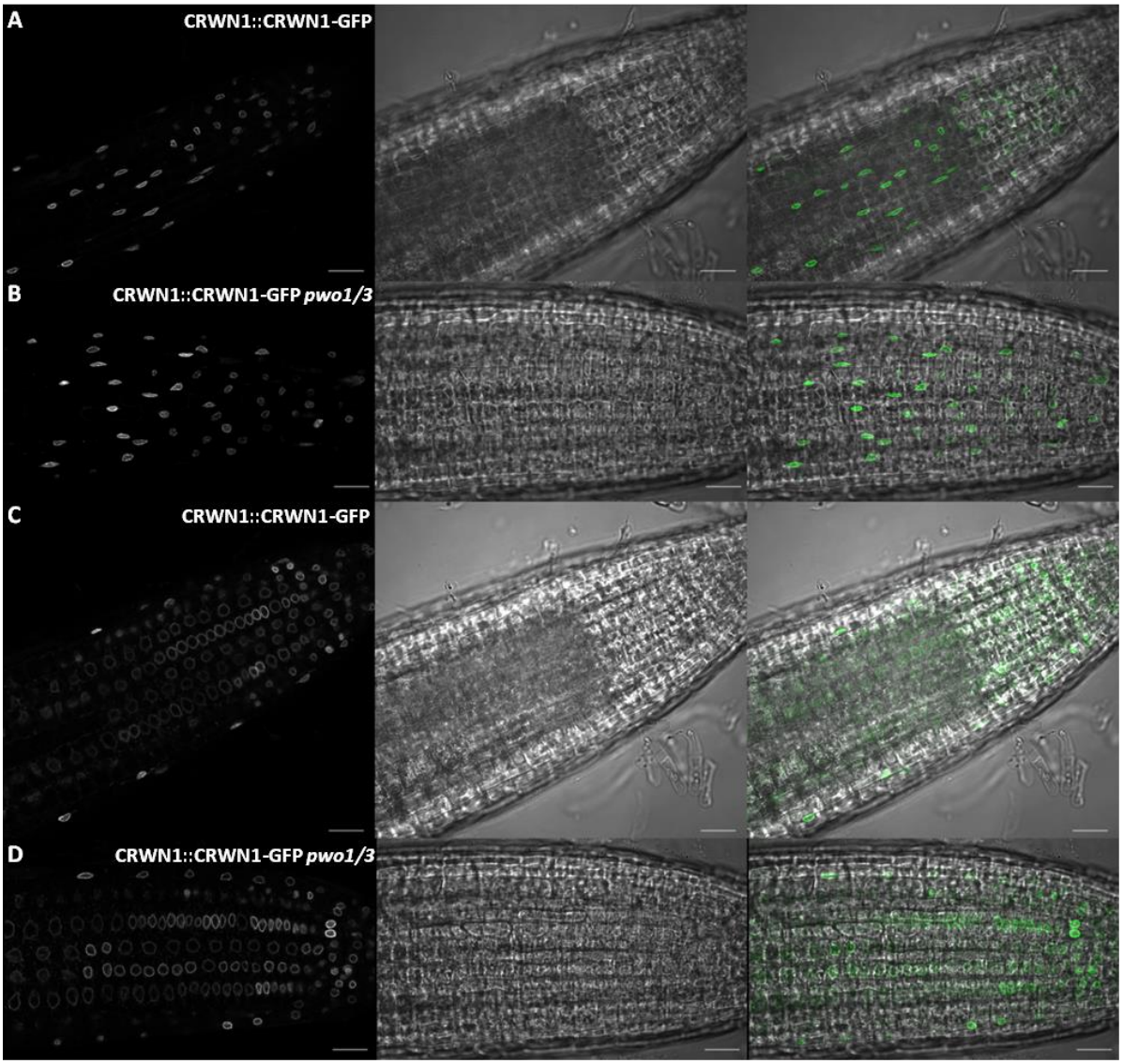



**Fig.S5. CRWN1 localization in *pwo1/3*.** Confocal microscopy was performed on roots from line: *CRWN1::CRWN1-GFP* in *pwo1/3*. Images acquired from *CRWN1::CRWN1-GFP* served as a control. The images represent different root tissues: epidermis of meristematic and elongation zones (A and B) or cortex of meristematic and elongation zones (C and D). Scale bar corresponds to 20 µm.

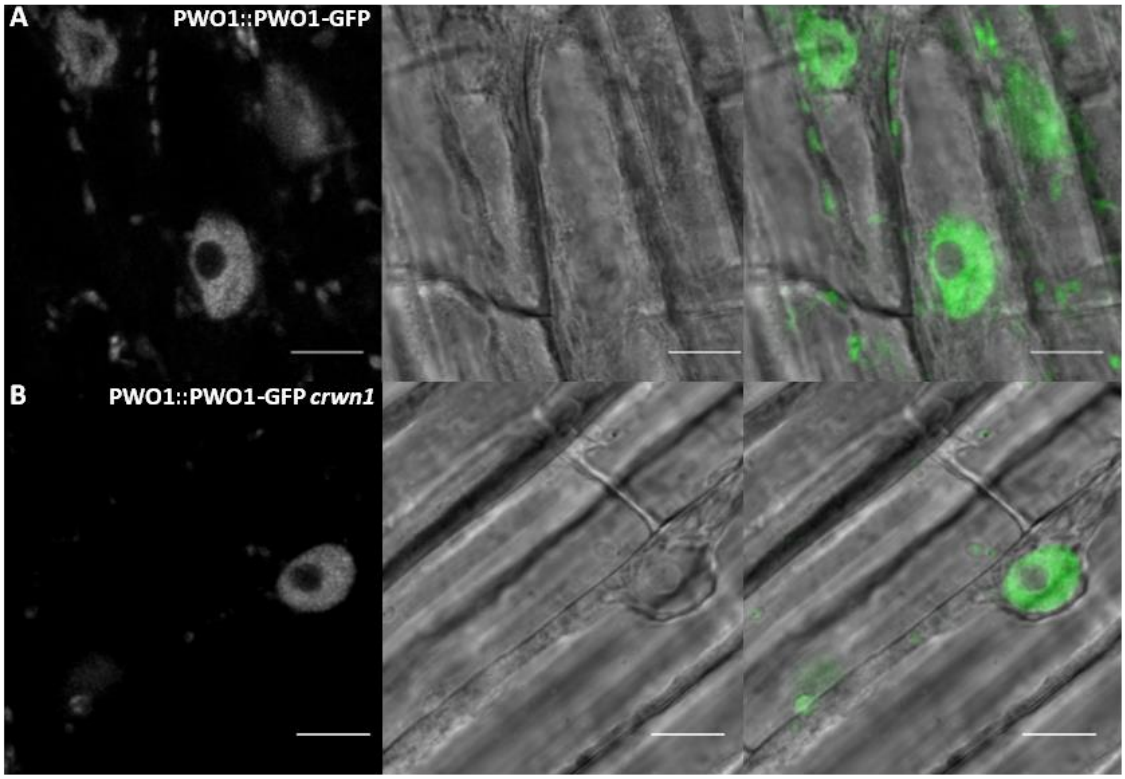



**Fig.S6. PWO1 localization in *crwn1*.** Confocal microscopy was performed on roots from the line: *PWO1::PWO1-GFP* in *crwn1* and the control: *PWO1::PWO1-GFP*. The images represent nuclei of the root elongation zone epidermis. The images were filtered using gaussian blur (sigma = 1). Scale bar corresponds to 10 µm.

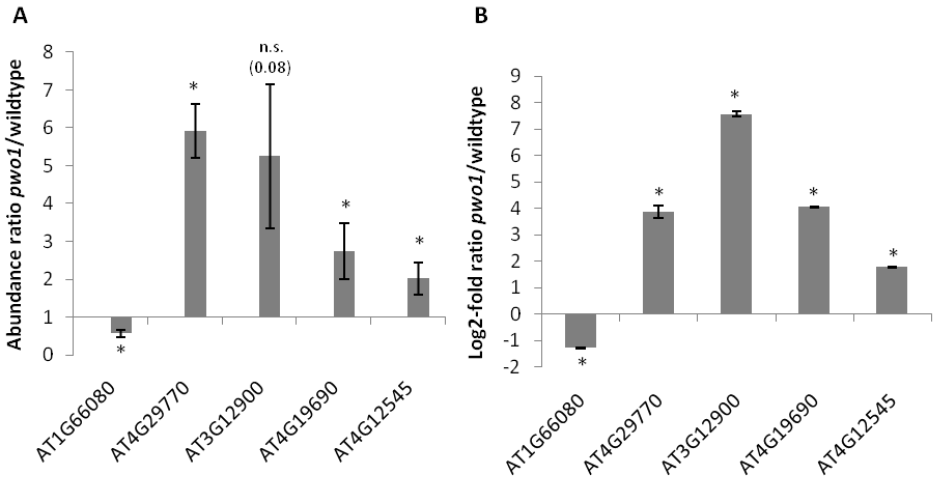



**Fig.S7. RT-qPCR-validation of RNA-seq data for selected DEGs.** (A) Transcript abundance ratio *pwo1*/wildtype inferred from RT-qPCR. Expression level was calculated using ΔΔCT method and normalized to EUKARYOTIC INITIATION FACTOR-4A (eIF4A). Col-0 samples were used as a wildtype control. Error bars represent standard error from 5 independent biological replicates. Asterisk marks significant ($p<0.05$) changes in transcript abundance calculated using Student's t-test. (B) Log2-fold ratio *pwo1*/wildtype inferred from RNA-seq. Col-0 samples were used as a wildtype control. The bars represent average log2-fold values from two RNA-seq analysis methods: Tuxedo (Cufflinks) and edgeR. Error bars correspond to standard deviation of log2-fold values from two RNA-seq analysis methods mentioned above. Asterisk represents significant (FDR<0.05) transcript abundance changes in *pwo1* over the wildtype, calculated by edgeR and Cuffdiff in Tuxedo method.

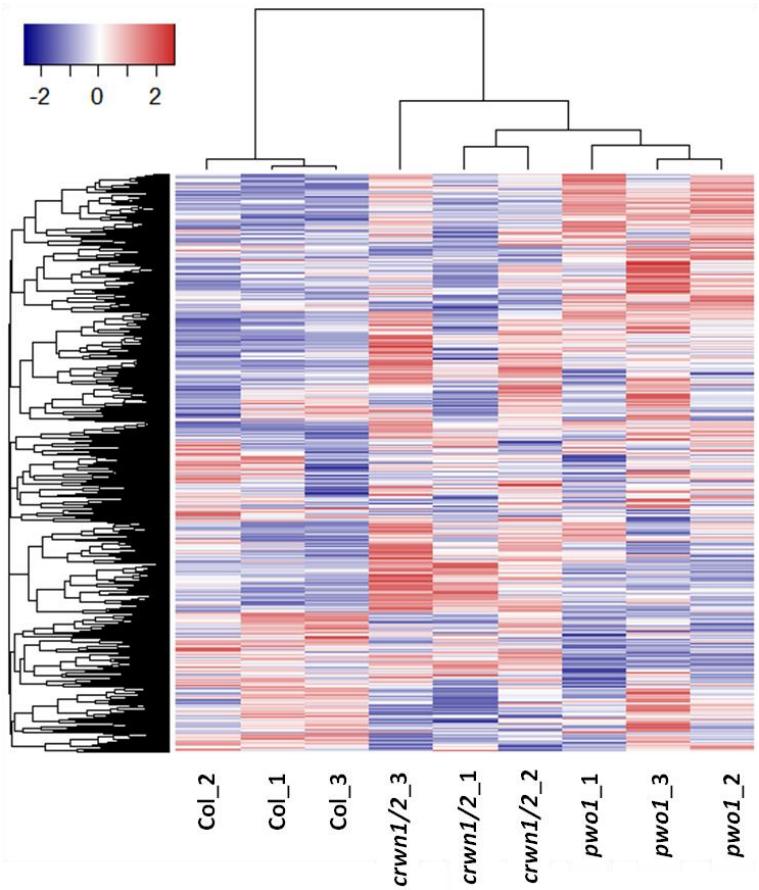



**Fig.S8. Correlation between samples in RNA-seq experiment.** A matrix containing "counts per million" (cpm) values for TAIR10 protein-coding genes was created. Genes containing less than 1 cpm were removed and not included in further analyses. The matrix served as input for heatmap generation in Heatmap3 v1.1.1 R-package. Default methods for distance computing and dendrogram re-odering were used. The figure shows three biological replicates ("_1","_2","_3") from 3 genotypes: wildtype/Col-0 ("Col"), *pwo1* ("*pwo1")* and *crwn1 crwn2 ("crwn1/2").*

**Table S2. Oligonucleotides used in the study.**

| purpose | locus | forward | reverse |
|---|---|---|---|
| **RT-qPCR** | | | |
| | AT1G66080 | CTTCGTCGGTGAAGCTTATGAT | GCTGATCCTGGAGATTGAACAT |
| | AT4G29770 | CATGAAAGGAGAAGTGATTGAGG | CCATCGTAAATACCCTCTCCA |
| | AT3G12900 | CTTACGTGTGAGGCTACCCA | GTTCACCACCTGGAAGAAGC |
| | AT4G19690 | CTATCATTAAGTGTTTCGCCTCC | CACGGGTTCTCTTCAAGACA |
| | AT4G12545 | CCTATACAGATCAGTACTTGTGCC | CGAGCAACATGGCTTTACAG |

### 3.2. Manuscript II

Given the PWO1 presence in PRC2-mediated repression pathway, the profile of H3K27me3, a typical PRC2 histone mark, was studied in *pwo1* mutant. Namely, a chromatin immunoprecipitation method coupled with RT-qPCR or sequencing (ChIP-seq) was used to identify H3K27me3 differentially-bound target genes in *pwo1*. Strikingly, the results revealed small number of genes with PWO1-dependent H3K27me3 abundance, with both, gained and lost levels of this histone mark in *pwo1*.

The observations were summarized as an addendum in Manuscript II: 'PWWP INTERACTOR OF POLYCOMBS 1 (PWO1) impacts H3K27me3 occupancy at specific target genes'.

**Manuscript II**

# Addendum: PWWP INTERACTOR OF POLYCOMBS 1 (PWO1) impacts H3K27me3 occupancy at specific target genes

**Pawel Mikulski1, Dimitrios Zisis2, Pawel Krajewski2, Daniel Schubert1a**

1 Institute for Biology, Free University Berlin, Germany

2 Institute of Plant Genetics, Polish Academy of Sciences, Poznan, Poland

a corresponding author: dan.schubert@fu-berlin.de

Keywords: H3K27me3, Polycomb, PRC2, PWO1, ChIP, ChIP-seq

Author contributions:

PM and DS designed the research. PM, DZ, PK performed ChIP-seq raw reads' analysis (mapping, read counting and differential occupancy analysis). PM performed chromatin immunoprecipitations, RT-qPCR experiments, sequencing library preparation, secondary bioinformatic analyses and sequencing quality checks. The manuscript was written by PM and revised by DS.

### 3.2.1. Abstract

POLYCOMB REPRESSIVE COMPLEX 2 (PRC2) mediates gene repression via tri-methylation of lysine-27 on histone H3 (H3K27me3) in species of all eukaryotic kingdoms. In recent years, numerous PRC2-associated proteins were identified to act as mediators of various steps in the PRC2-pathway. PWWP INTERACTOR OF POLYCOMBS 1 (PWO1) was shown previously to be a PRC2-associated protein in *Arabidopsis* as it interacts physically with a number of PRC2 components in *Arabidopsis thaliana* and control expression of several PRC2 target genes in vegetative tissues. However, PWO1 position in PRC2 pathway is unknown. Here, we used target-directed and genome-wide approaches to study PWO1 role in controlling H3K27me3 levels. We show that PWO1 has a mild effect on global levels of this histone mark, but *pwo1* loss-of-function mutant displays differential binding of H3K27me3 on a small number of genes. Overall, our results led to the identification of a number of indirect/direct PWO1-targets, highlighted redundancy within PWO gene family and a potential PWO1 function downstream to H3K27me3 deposition.

### 3.2.2. Results

PWO1 was previously reported to be a PRC2-associated protein in *Arabidopsis* due to its physical interaction with several PRC2 components (Hohenstatt, 2012), epistasis with CLF (Hohenstatt, 2012) and transcriptomic control of a similar set of target genes as PRC2 (manuscript I). However, the exact function of PWO1 in PRC2-pathway remains unknown. As H3K27me3 is a prominent histone mark for PRC2-mediated gene repression, we sought to investigate whether H3K27me3 levels are PWO1-dependent using genome-wide and target-directed approaches.

#### Target-directed H3K27me3 differential binding

In order to study H3K27me3 level dependence on PWO1, we performed a chromatin immunoprecipitation (ChIP) experiment on 2 week-old seedlings from wildtype control (Col-0), *pwo1* single mutant and *pwo1 pwo3* double mutant. Anti-H3K27me3 antibody and rabbit IgG, for control of antibody non-specific binding, were used. Differential H3K27me3 binding between the wildtype and PWO family mutants was assessed by RT-qPCR on several key

flowering genes that are known to be PRC2 targets during in the vegetative stage: AGAMOUS (AG), FLOWERING LOCUS C (FLC), FUSCA 3 (FUS3), SEPALLATA 3 (SEP3). The non-PRC2 target gene, At4g18950, located upstream to AG was included as a negative control locus ("H3K27me3-").The results revealed around 50% reduction in H3K27me3 enrichment on PRC2 targets in PWO family mutants compared to the wildtype (Fig.1). Noteworthy, no substantial differences in histone mark abundance were seen between *pwo1* and *pwo1 pwo3*. Expectedly, we observed H3K27me3 enrichment on the background level in negative IgG samples and on the negative locus (Fig.S1). Overall, we showed that PWO1 is required for full H3K27me3 level on several known PRC2 target loci and act non-additively with PWO3.

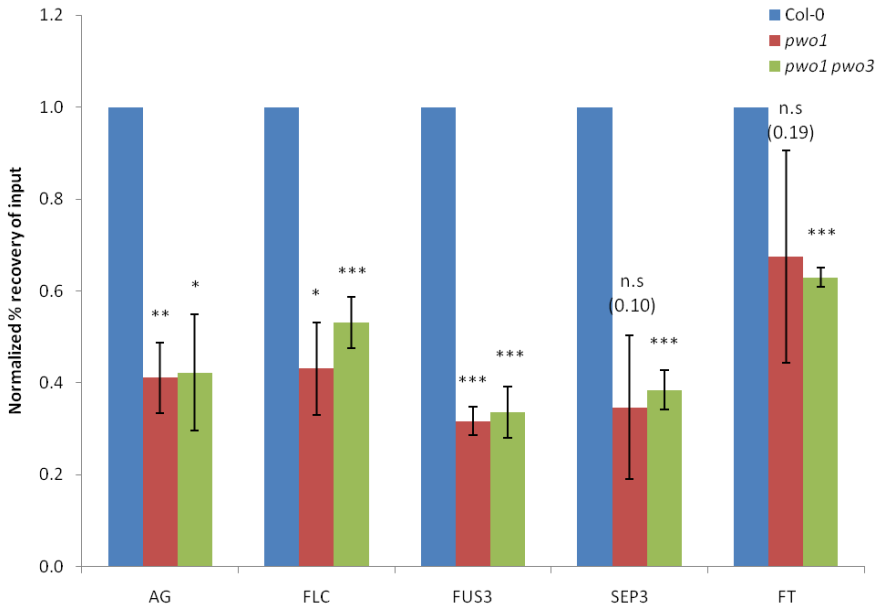



**Fig.1. H3K27me3 occupancy on selected PRC2-targets in *pwo1* and *pwo1/3*.** Chromatin immunoprecipitation experiment was performed using H3K27me3 antibody on chromatin from wildtype (Col-0), *pwo1* and *pwo1/3*. H3K27me3 occupancy is represented as % recovery of input normalized to the wildtype level. Error bars correspond to the standard standard deviation (SD) from 2 biological replicates. Significance was calculated using one sample Student's t-test.

## Genome-wide H3K27me3 differential binding

Given a reduction in H3K27me3 occupancy on several loci seen already in *pwo1* single mutant, we inquired about the abundance of the histone mark on a genome-wide scale. Consequently, a chromatin immunoprecipitation method coupled with next-generation sequencing was performed on wildtype Col-0 and *pwo12* week-old seedlings with anti-H3K27me3 antibody. Immunoprecipitated samples were sequenced in two biological

replicates with an addition of single-replicate input samples from both genotypes. The sequencing yielded in each sample ~ 30 mln reads that passed Q20 quality score giving ~ 22-24x *Arabidopsis* genome coverage. Afterwards, the reads were mapped to the reference genome with an efficiency of ~ 94-95% for input and ~ 64-74% for H3K27me3 samples. A subsequent normalization to the input led to identification of ~ 4900-5400 domains in H3K27me3 samples. Sequencing parameters are collectively presented in Table 1.

**Table 1. ChIP-seq parameters**

| Name | Reads | Q20 | Genome coverage | Mapping efficiency | Domain number |
|---|---|---|---|---|---|
| input Col | 33,753,912 | 0.98 | 24.62 | 95.18% | NA |
| input *pwo1* | 33,781,246 | 0.99 | 24.68 | 94.29% | NA |
| H3K27me3 IP1 Col | 30,277,024 | 0.98 | 21.96 | 68.20% | 5406 |
| H3K27me3 IP1 *pwo1* | 33,460,158 | 0.98 | 24.34 | 71.82% | 4911 |
| H3K27me3 IP2 Col | 31,527,030 | 0.98 | 22.90 | 63.89% | 4920 |
| H3K27me3 IP2 *pwo1* | 29,329,998 | 0.98 | 21.36 | 74.40% | 4953 |

The columns correspond to following (from the left to the right): samples names, total number of reads, fraction of reads passing Q20 quality score threshold, Bowtie mapping efficiency and H3K27me3 domain number identified after normalization to the input.

In order to assess consistency between our dataset and published results, we annotated H3K27me3 domains in the wildtype control and cross-compared resulting gene list with three reference H3K27me3 targets (Bouyer et al., 2011; Oh et al., 2008; Zhang et al., 2007). As a result, we detected a substantial overlap (~ 93%) between our H3K27me3 target list in Col-0 and published datasets, confirming the consistency with the references (Fig.2).

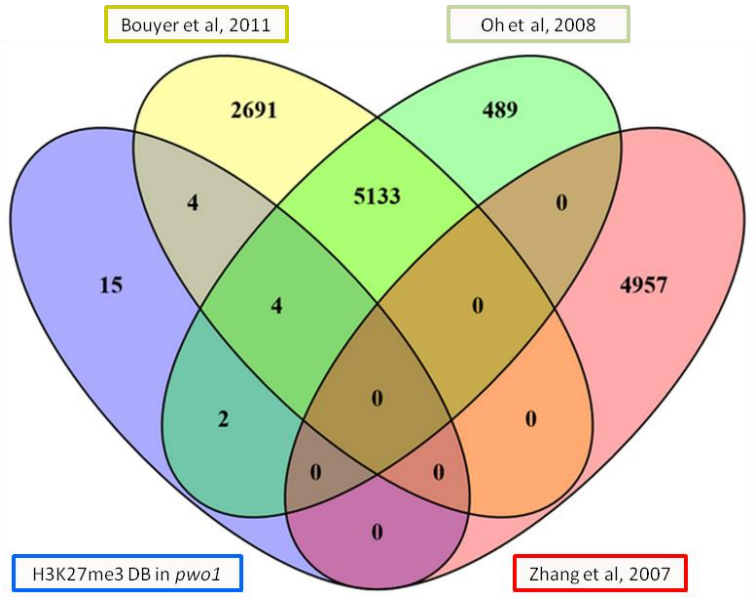



**Fig.2. The overlap of the reference H3K27me3 target lists and differentially-bound genes with elevated H3K27me3 level in *pwo1*.** The reference lists consist of genes covered by H3K27me3 in the wildtype (Col-0) seedlings shown in: (Bouyer et al., 2011; Oh et al., 2008; Zhang et al., 2007).

Next, we performed differential binding analysis for H3K27me3 between *pwo1* and the wildtype. Specifically, we used DiffBind-pipeline so the affinity scores of the reads were calculated and compared between consensus H3K27me3 domains, shared by at least 2 biological samples from either of the genotypes. Afterwards, the domains were annotated by intersection with the reference TAIR10 protein-coding gene list. As a result, the analysis revealed 60 H3K27me3 differentially-bound (DB) genes, with 35 genes having lower H3K27me3 levels and 25 genes with higher H3K27me3 abundance (Fig.3, Table S1).

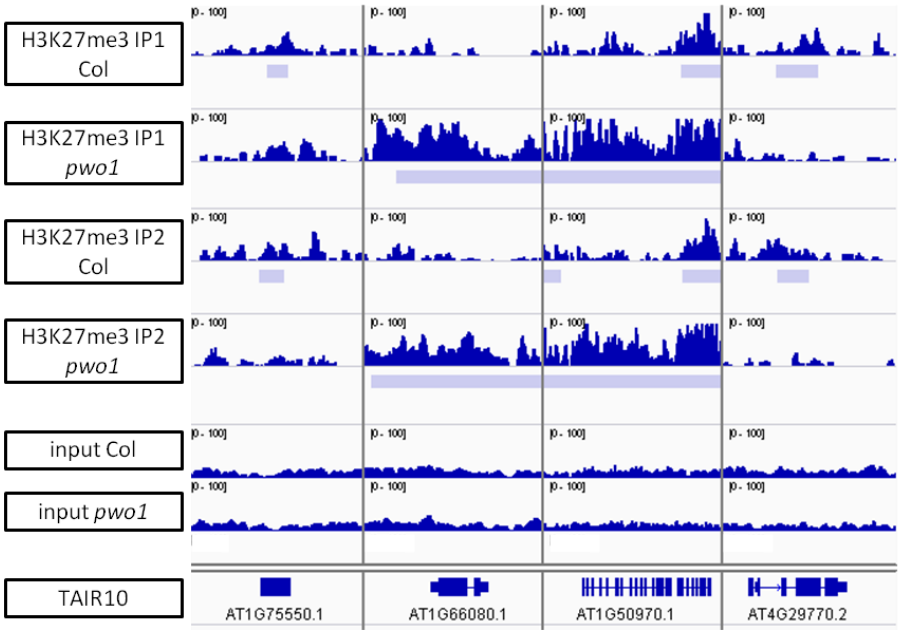



**Fig.3. H3K27me3-domain occupancy over selected H3K27me3 differentially-bound genes.** Bars below H3K27me3 tracks correspond to domain presence identified by MACS2 peak caller with input used for normalization. Visualization was performed in IGV v2.3 software.

The analysis led to two striking observations. Firstly, lack of PWO1 causes both, reduction and elevation, of H3K27me3 abundance, depending on the target gene. Interestingly, the gain of the histone mark for 15 out of 25 genes with elevated H3K27me3 abundance was not only quantitative, but also qualitative, as those targets were not reported in the references to be covered with H3K27me3 during seedling stage (Fig.4). Secondly, changes of H3K27me3 abundance in *pwo1* were generally very mild in comparison to the wildtype. Using target-directed approach we showed that disruption of PWO1 does not lead to the full depletion of H3K27me3, but rather ~ 50% decrease from wildtype level (Fig.1). However, our target-directed approach was limited to the RT-qPCR-amplification of selected regions of the genes and it does not reflect an abundance on the whole gene-body. Direct comparison of ChIP-seq-

based and ChIP-RT-qPCR-based approaches on the same target loci revealed that fluctuations of H3K27me3 levels between *pwo1* and the wildtype are local and the occupancy on whole gene-body remains similar between the genotypes (Fig.S3).

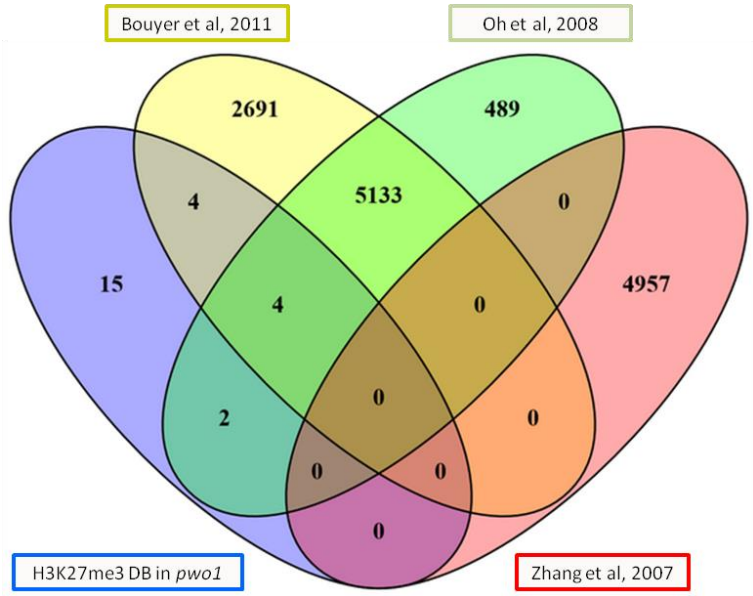



**Fig.4. The overlap of the reference H3K27me3 target lists and differentially-bound genes with elevated H3K27me3 level in *pwo1*.** The reference lists consist of genes covered by H3K27me3 in the wildtype (Col-0) seedlings shown in: (Bouyer et al., 2011; Oh et al., 2008; Zhang et al., 2007).

Noteworthy, a low number of H3K27me3 differentially-bound regions in *pwo1* might come from the stringency of the analysis and uneven quality of sequencing reads between the samples. We detected unexpectedly similar normalized count-based correlation values between Col-0 IP2 sample and *pwo1* replicates as between wildtype samples (Fig.S4). Such issue might be explained by poorer mapping efficiency in H3K27me3 Col-0 IP2 sample and lower number of H3K27me3 domains identified in input-normalized H3K27me3 Col-0 IP2, in comparison to H3K27me3 Col-0 IP1 sample (Table 1). The possible causes concern DNA contamination and amplification artifacts during library preparation, rather than technical problems with sequencing itself as Phred quality scores of the nucleotides in trimmed reads were optimal (data not shown). Technical issues were addressed by setting stringent conditions for peak calling (duplicated reads removed) and differential binding analysis (number of minimal overlapping peaksets set to 2) to ensure reliability of output list of differentially-bound genes. In addition, the results were confirmed for selected targets by ChIP-RT-qPCR experiment on independent biological replicate and samples used to generate ChIP-seq data (Fig.5). Overall, despite technical discrepancies between ChIP-seq samples,

stringency of the analysis allowed to eliminate false positive targets, giving a small and reliable set of differentially-bound genes. However, we see a possibility that our ChIP-seq data used here contains false negatives and that the analysis on improved ChIP-seq data might yield a larger list of targets.

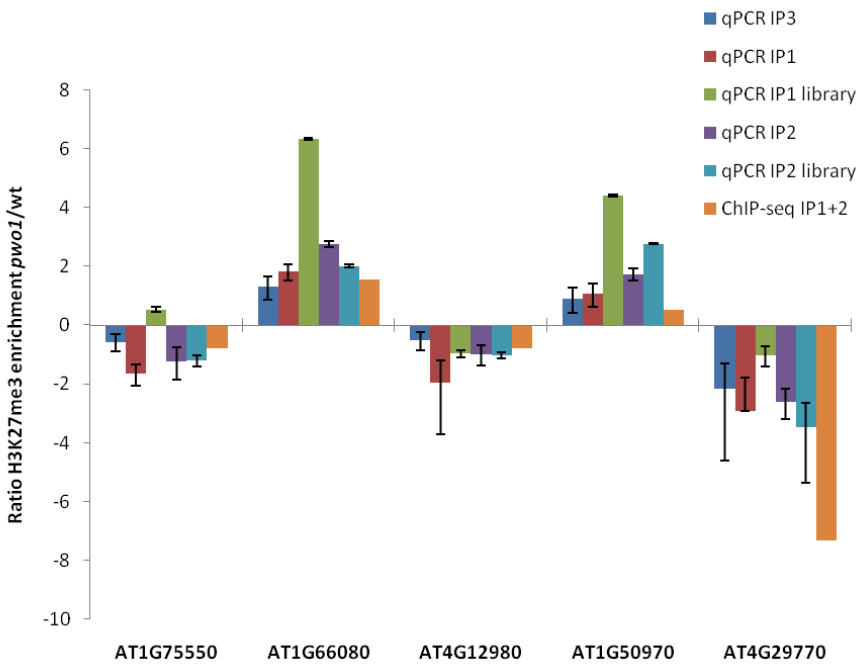



**Fig.5. RT-qPCR-based validation of ChIP-seq.** PCR amplification of target gene fragments was done on samples from two biological replicates before ChIP-seq library preparation (IP1, IP2), after ChIP-seq libary preparation (IP1 library, IP2 library) and independent biological replicate not used for sequencing (IP3). The bars represent relative log2 fold ratio *pwo1*/wildtype. For comparison, average log2 fold enrichment from two sequenced samples calculated in Diffbind v2.0.7 during ChIP-seq analysis is shown (ChIP-seq IP1+2). Error bars correspond to log-transformed standard error between three technical RT-qPCR replicates (does not apply to ChIP-seq data).

## Characterization of H3K27me3 differentially-bound targets in *pwo1*

In order to further analyze H3K27me3 differentially-bound (DB) targets in *pwo1*, we focused on their functional classification and relation to transcriptomic activity. Unfortunately, 15 out of 60 DB genes correspond to unknown proteins, lacking proper annotation. Nevertheless, we performed gene ontology (GO) analysis on two DB gene groups of either higher or lower H3K27me3 abundance in *pwo1*. The results revealed significant (FDR=0.029) GO enrichment in protein-disulfide reductase activity in targets with reduced H3K27me3 in *pwo1*, but none among genes with elevated H3K27me3 in *pwo1* (Fig.6). This enriched GO term is represented by 3 targets belonging to the same family of cysteine/histidine-rich C1

domain proteins: AT5G42840, AT5G45730, AT4G13992. C1 domain binds a secondary messenger diacyloglycerol (DAG) and analogous phorbol esters, making the C1 domain-containing proteins the important intermediates in signalling pathways controlling metabolism, development and stress responses (Dong et al., 2012; Hurley et al., 1997). However, due to the low gene number of this GO term in our dataset and high number of poorly annotated H3K27me3 DB targets, our functional classification results remain currently inconclusive and require further validation.

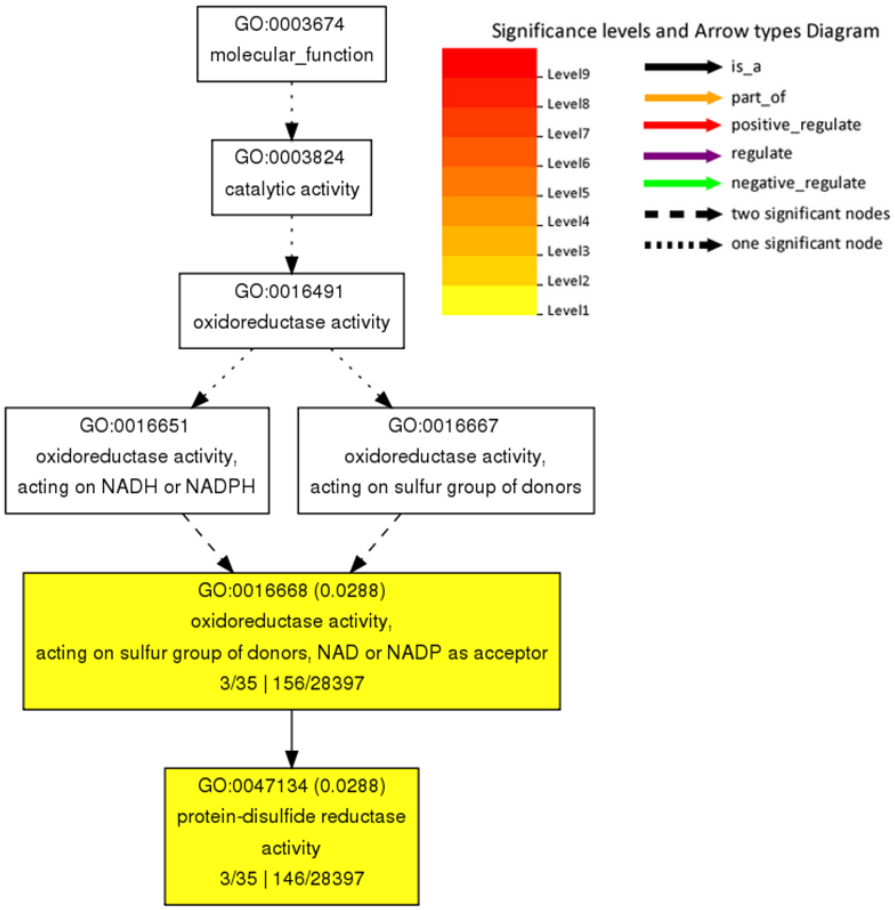



**Fig.6. Gene ontology analysis.** The figure corresponds to the differentially-bound genes with reduced H3K27me3 levels in *pwo1* compared to the wildtype. Colour coding represents significance level (1-9) below adjusted p-value threshold (0.05). Adjusted p-values are shown next to GO IDs for each GO term. Relationships between GO terms are reflected in the arrow types. Numbers below subcategory name correspond to: GO subcategory gene number in query/total gene number in query | GO subcategory gene number in reference TAIR10/total TAIR10 protein-coding gene number.

Next, given a canonical, repressive function of H3K27me3, we inquired about the expression status of DB genes in *pwo1*. We cross-compared H3K27me3 DB gene list to a set of differentially expressed genes (DEGs) in *pwo1* (manuscript I), as both dataset were generated

from similar material (2 week-old seedlings grown in long-day conditions). Ultimately, we detected a significant 11% overlap (7 out of 60 DB genes) between both datasets (Fig.7A). The overlapping targets display expected inverse correlation, with a loss of H3K27me3 accompanied by elevated transciptomic activity and *vice versa* (Fig.7B). In conclusion, large number of DEGs non-overlapping with H3K27me3 DB targets suggests mostly the indirect transcriptomic changes dependent on PWO1 function, but irrespective to H3K27me3 status and corresponds partially to the potential PcG-independent role of PWO1. On the other side, large number of H3K27me3 DB tagets non-overlapping with DEGs disagrees with canonical relationship between H3K27me3 occupancy and gene expression, but is in agreement with recent reports about non-straightforward influence of this histone mark on target transcriptomic status (see discussion). Alternatively, for several targets with small, but nonetheless significant H3K27me3 differential binding, the impact on the expression might depend only on H3K27me3 occupancy changes of greater fold.

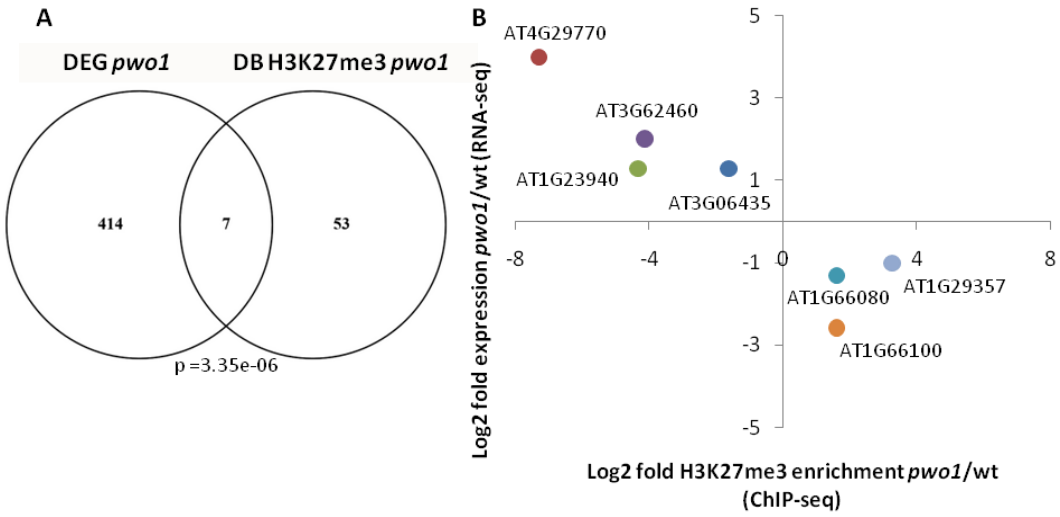



**Fig.7. Relationship between PWO1-dependent changes in H3K27me3 occupancy and gene expression.** (A) Overlap between genes: differentially expressed in *pwo1* (DEG *pwo1*) and H3K27me3 differentially-bound in *pwo1* (DB H3K27me3 *pwo1*). DEG *pwo1* correspond to misexpressed targets inferred from RNA-seq analysis done using two methods: Tuxedo protocol (Cufflinks) and edgeR. Only the targets that showed significant (FDR<0.05) 2-fold changes in *pwo1* compared to the wildtype (Col-0) and significant (FDR<0.05) misexpression in edgeR output, were included in the comparison to DB H3K27me3 *pwo1*. DB H3K27me3 in *pwo1* correspond to genes showing significant H3K27me3 occupancy changes in *pwo1* comparing to the wildype (Col-0). Significance level inferred from hypergeometric test of overlapped gene number: p-value = 3.35e-06. (B) Inverse correlation between PWO1-dependent H3K27me3 occupancy changes and misexpression levels on overlapped set of genes from (A).

### 3.2.3. Discussion

Here we reported that PWO1 impacts H3K27me3 abundance on a specific set of genes using target-directed and genome-wide approaches. Our results led to several conclusions.

Firstly, we observed an overall mild effect on H3K27me3 abundance dependent on PWO1 presence, which is in agreement to *pwo1* whole-plant phenotype resemblance to the wildtype. As single and double mutants within PWO family show similar whole-plant phenotypes (Hohenstatt, 2012), it is likely that such redundancy occurs also on the H3K27me3 level. In addition, our target-directed approach showed similar local reduction of H3K27me3 in *pwo1* and *pwo1 pwo3*, suggesting that PWO1 and PWO3 act in a non-additive manner and PWO2 alone ensures partial H3K27me3 enrichment on the target genes. However, such effect is probably not specific to PWO2 as similar whole-plant phenotypes of double mutants within PWO family (Hohenstatt, 2012) suggest that none of the members have a dominant function. However, we cannot exclude that the phenotypes of PWO family mutants correspond at least partially to non-PcG-related mechanisms or that specific functional differences between the functions of PWO family genes exist. Further work on H3K27me3 in different loss-of-function mutant combinations or in systems of inducible expression of PWO family proteins in triple mutant background should dissect such potential functional differences within PWO family.

Moreover, further characterization of PWO1 function in PcG-mediated pathway should concern identification of PWO1direct targets. Despite pinpointing genes showing PWO1-dependent H3K27me3 occupancy in this study, it remains to be elucidated whether PWO1 binds to their chromatin and how exactly PWO1 affects H3K27me3 abundance. Given PWO1 association with nuclear lamina (chapter II), it is possible that small number of H3K27me3 DB genes in *pwo1* correspond to indirect targets and PWO1 function concerns rather spatial organization of PcG targets. In that way, PWO1 role would be downstream to the catalysis of H3K27me3, analogous to the other architectural factors, AtMORC1 and AtMORC6, involved in heterochromatin condensation downstream to H3K9me3 deposition and DNA methylation (Moissiard et al., 2012).

Furthermore, we observed lack of straightforward correlation between H3K27me3 occupancy and transcriptomic activity. Despite canonical repressive role of PRC2, recent studies reported that changes in H3K27me3 status are not fully linked to the misexpression, especially concerning individual genes, instead of the global level (Arthur et al., 2014; Sani et

al., 2013; Zhang et al., 2007). The cause of such discrepancy is currently unknown, but the possible explanations include interplay with the other histone marks, as seen for the bivalent domains in vertebrates (Voigt et al., 2013), and the coordination with downstream chromatin remodelers and transcription factors.

Bivalent domains containing both, H3K27me3 and H3K4me4, marks are predominantly associated with mammalian cells; however they were identified also for a number of target genes in *Arabidopsis* (Jiang et al., 2008; Saleh et al., 2007). Assessing the abundance of H3K4me3 and the other histone marks in *pwo1* in further studies will decipher if PWO1-dependent H3K27me3 changes relate also to the other chromatin modifications.

The PRC2-pathway is known to be associated with a number of downstream factors and H3K27me3 function is dependent on the presence of reader proteins (Derkacheva and Hennig, 2014; Sauvageau and Sauvageau, 2010), instead of directly influencing the expression of its targets. Therefore, it is possible that slight changes in H3K27me3 abundance are not a sufficient factor to trigger response in downstream pathways or downstream factors are regulated also in H3K27me3-independent manner and potentially not activated/expressed under certain conditions. Interestingly, PWO1 was found previously to interact with numerous chromatin-associated proteins. Scoring their function and abundance in *pwo1* mutant should unravel an impact of putative PWO1-downstream factors on gene expression status.

### 3.2.4. Materials & methods

#### Plant material

*A. thaliana* seedlings were grown on ½ MS plates in long day conditions (22°C). For knock-out mutations, following transgenic lines were used: *pwo1-1* (SAIL_342_C09), *pwo1-1 pwo3-2* (SAIL_342_C09, Sail_828_A07). Oligonucleotides used for genotyping are listed in Table S2.

**Chromatin immunoprecipitation**

ChIP experiments from 1g of 14-day old seedlings were carried out as described in Universal Plant ChIP-seq Kit manual (Diagenode). Antibodies against H3K27me3 (Diagenode, #C15410195) and IgG (Diagenode, #C15410206), conjugated with protein A-coated magnetic beads (Thermo Fisher Scientific, #88846) were used. De-crosslinked DNA was purified by standard phenol-chloroform method and taken for: 1) ChIP-seq library preparation or 2) RT-qPCR.1) Sequencing library was prepared following manual from MicroPlex Library Preparation Kit (Diagenode, #C05010010) with multiplexing according to Low-Level Pooling guidelines (Illumina, #1005361) and gel-free DNA purification on SPRI beads (AMPure XP, Becman Coulter, #A63880). The library was sequenced on HiSeq 2000 (Illumina). 2) RT-qPCR reactions were done using KAPA SYBR FAST Master Mix (KapaBiosystems, #KK4600) and run on QuantStudio 5 Real-Time PCR System (Applied Biosystems). For final results, Ct values for IP samples were calculated to % enrichment to Input. Oligonucleotides used for RT-qPCR are listed in Table S2.

**ChIP-seq analysis**

Qualtity of 2x 100bp paired-end raw sequencing reads was assessed using FastQC v0.10. The reads were trimmed according to their quality in Bowtie v2.2.5 and aligned to Arabdiopsis reference genome TAIR10 with minimal (-I) and maximal (-X) fragment length set to 200 and 600, respectively; minimal alignment (--score-min) as L,0,-0.25 and allowing a mixed mode for read-pairing. Reads mapping to multiple locations were removed in R Studio v0.99. Subsequently, uniquely mapped reads had duplicates removed in Samtools v1.1 by rmdup command and used for peak calling in MACS v2.1.1 with broad domain settings and q-value = 0.05. As rmdup command in Samtools v1.1 does not work for unpaired reads, additional stringency was ensured by keeping one tag per location in MACS v2.1.1 (--keep-dup = 1). Differential H3K27me3 binding was scored in DiffBind v2.0.7 with minimal overlap in read counting (minOverlap) set to 2 peaksets. Differentially bound regions were selected based on significance threshold FDR = 0.05 and intersected with TAIR10 gene annotation by Bedtools v2.26. Alignment files were converted to wig format in PeakRanger v1.18, then to tdf format in IGVtools and visualized in genome browser IGV v2.3.

**Secondary bioinformatic analyses**

Correlation between biological replicates was assessed by counting number of H3K27me3 reads normalized to the input in 300bp-long bins and performing R-squared regression on the values obtained.

Gene ontology was inferred by using Singular Enrichment Analysis on AgriGO server (Du et al., 2010) against a complete GO list. Protein annotations were extracted from UniProtKB database and used as reference. To select significantly enriched GO terms, Fisher test with Yekutieli adjustment was used a statistical method. Adjusted p-value (FDR) threshold was set to 0.05 and minimal number of entries kept at 3.

Venn diagrams were generated in Venny v2.1 software. Statistical significance of overlapping gene number was calculated in R using hypergeometric test.

### 3.2.5. References


Arthur, R. K., Ma, L., Slattery, M., Spokony, R. F., Ostapenko, A., Negre, N., et al. (2014). Evolution of H3K27me3-marked chromatin is linked to gene expression evolution and to patterns of gene duplication and diversification. *Genome Res.* 24, 1115–1124. doi:10.1101/gr.162008.113.

Bouyer, D., Roudier, F., Heese, M., Andersen, E. D., Gey, D., Nowack, M. K., et al. (2011). Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. *PLoS Genet.* 7, e1002014. doi:10.1371/journal.pgen.1002014.

Derkacheva, M., and Hennig, L. (2014). Variations on a theme: Polycomb group proteins in plants. *J. Exp. Bot.* 65, 2769–2784. doi:10.1093/jxb/ert410.

Dong, W., Lv, H., Xia, G., and Wang, M. (2012). Does diacylglycerol serve as a signaling molecule in plants? *Plant Signal. Behav.* 7, 472–5. doi:10.4161/psb.19644.

Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: A GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* 38. doi:10.1093/nar/gkq310.

Hohenstatt, M. L. (2012). Functional analysis of SCI1 – A PWWP domain protein involved in Polycomb group mediated gene regulation in Arabidopsis.

Hurley, J. H., Newton, A. C., Parker, P. J., Blumberg, P. M., and Nishizuka, Y. (1997). Taxonomy and function of C1 protein kinase C homology domains. *Protein Sci.* 6, 477–80. doi:10.1002/pro.5560060228.

Jiang, D., Wang, Y., Wang, Y., He, Y., Turck, F., Fornara, F., et al. (2008). Repression of FLOWERING

LOCUS C and FLOWERING LOCUS T by the Arabidopsis Polycomb Repressive Complex 2 Components. *PLoS One* 3, e3404. doi:10.1371/journal.pone.0003404.

Moissiard, G., Cokus, S. J., Cary, J., Feng, S., Billi, A. C., Stroud, H., et al. (2012). MORC Family ATPases Required for Heterochromatin Condensation and Gene Silencing. *Science (80-. ).* 336.

Oh, S., Park, S., and van Nocker, S. (2008). Genic and global functions for Paf1C in chromatin modification and gene expression in Arabidopsis. *PLoS Genet.* 4, e1000077. doi:10.1371/journal.pgen.1000077.

Saleh, A., Al-Abdallat, A., Ndamukong, I., Alvarez-Venegas, R., and Avramova, Z. (2007). The Arabidopsis homologs of trithorax (ATX1) and enhancer of zeste (CLF) establish "bivalent chromatin marks" at the silent AGAMOUS locus. *Nucleic Acids Res.* 35, 6290–6. doi:10.1093/nar/gkm464.

Sani, E., Herzyk, P., Perrella, G., Colot, V., and Amtmann, A. (2013). Hyperosmotic priming of Arabidopsis seedlings establishes a long-term somatic memory accompanied by specific changes of the epigenome. *Genome Biol.* 14, R59. doi:10.1186/gb-2013-14-6-r59.

Sauvageau, M., and Sauvageau, G. (2010). Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. *Cell Stem Cell* 7, 299–313. doi:10.1016/j.stem.2010.08.002.

Voigt, P., Tee, W.-W., and Reinberg, D. (2013). A double take on bivalent promoters. *Genes Dev.* 27, 1318–38. doi:10.1101/gad.219626.113.

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y. V, Pellegrini, M., Goodrich, J., et al. (2007). Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. *PLoS Biol.* 5, e129. doi:10.1371/journal.pbio.0050129.

### 3.2.6. Supplementary data

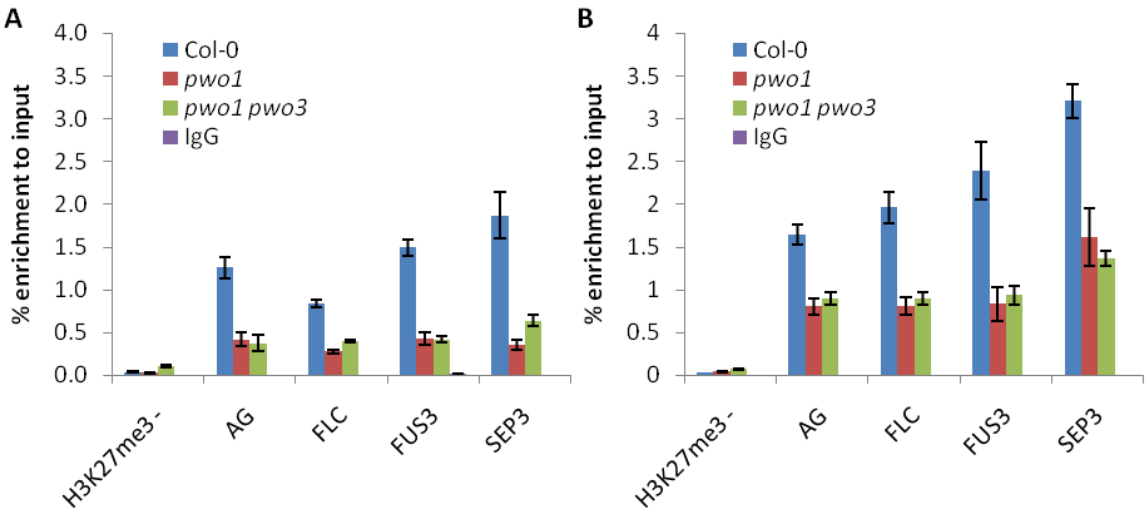



**Fig.S1. H3K27me3 occupancy on selected PRC2-targets in *pwo1* and *pwo1/3*.** Chromatin immunoprecipitation experiment was performed using H3K27me3 antibody on chromatin from wildtype (Col-0), *pwo1* and *pwo1/3* biological replicate #1 (A) and #2 (B). H3K27me3 occupancy is represented as % recovery of input normalized to wildtype level. Error bars correspond to standard error from 3 PCR technical replicates.

**Table S1. H3K27me3 differentially-bound genes in *pwo1*.**

| AGI | Gene model description | Fold |
|---|---|---|
| AT4G29770 | Target of trans acting-siR480/255. | -7.32 |
| AT2G32680 | RECEPTOR LIKE PROTEIN 23 (RLP23) / receptor like protein 23 (RLP23) | -7.24 |
| AT5G42840 | Cysteine/Histidine-rich C1 domain family protein | -7.20 |
| AT5G28640 | ANGUSTIFOLIA 3 (AN3) / Involved in cell proliferation during leaf and flower development | -6.08 |
| AT2G08986 | unknown protein | -5.41 |
| AT1G28190 | unknown protein | -5.36 |
| AT5G48400 | GLUTAMATE RECEPTOR 1.2 (ATGLR1.2) / member of Putative ligand-gated ion channel subunit family | -5.24 |
| AT1G05540 | Protein of unknown function (DUF295) | -4.37 |
| AT1G05550 | Protein of unknown function (DUF295) | -4.37 |
| AT2G11010 | unknown protein | -4.35 |
| AT1G23940 | ARM repeat superfamily protein | -4.30 |
| AT1G23935 | CONTAINS InterPro DOMAIN/s: Apoptosis inhibitory 5 | -4.30 |
| AT3G62460 | Putative endonuclease or glycosyl hydrolase | -4.15 |
| AT5G33390 | glycine-rich protein | -3.88 |
| AT5G08000 | GLUCAN ENDO-1,3-BETA-GLUCOSIDASE-LIKE PROTEIN 3 (E13L3) / Encodes a member of the X8-GPI family of proteins | -3.50 |
| AT3G43960 | Encodes a putative cysteine proteinase. Mutants exhibit shorter root hairs under phosphate-deficient conditions. | -3.44 |
| AT5G54062 | unknown protein | -2.99 |

| | | |
|---|---|---|
| AT5G26900 | CELL DIVISION CYCLE 20.4 (CDC20.4) / No expression of gene detected yet. | -2.98 |
| AT1G44130 | Eukaryotic aspartyl protease family protein | -2.91 |
| AT5G03510 | C2H2-type zinc finger family protein | -2.42 |
| AT5G27810 | MADS-box transcription factor family protein | -2.41 |
| AT1G64830 | Eukaryotic aspartyl protease family protein | -2.33 |
| AT1G35730 | PUMILIO 9 (PUM9) / Contains PUF domain that regulate both mRNA stability and translation | -2.24 |
| AT3G11350 | Pentatricopeptide repeat (PPR) superfamily protein | -2.22 |
| AT3G11340 | UDP-DEPENDENT GLYCOSYLTRANSFERASE 76B1 (UGT76B1) / Modulates plant defense and senescence. | -2.22 |
| AT2G41470 | unknown protein | -2.19 |
| AT5G45730 | Cysteine/Histidine-rich C1 domain family protein | -2.17 |
| AT4G13992 | Cysteine/Histidine-rich C1 domain family protein | -1.97 |
| AT5G18160 | F-box and associated interaction domains-containing protein | -1.86 |
| AT5G35525 | PLAC8 family protein | -1.82 |
| AT3G10080 | RmlC-like cupins superfamily protein | -1.64 |
| AT3G06435 | unknown protein | -1.64 |
| AT5G09730 | BETA-XYLOSIDASE 3 (BXL3) / Encodes a protein similar to a beta-xylosidase located in the extracellular matrix | -1.63 |
| AT2G05330 | BTB/POZ domain-containing protein | -1.58 |
| AT1G19460 | Galactose oxidase/kelch repeat superfamily protein | -1.52 |
| AT1G66090 | Disease resistance protein (TIR-NBS class) | 1.57 |
| AT1G66110 | Family of unknown function (DUF577) | 1.57 |
| AT1G66100 | Predicted to encode a PR (pathogenesis-related) protein | 1.57 |
| AT1G66070 | Translation initiation factor eIF3 subunit | 1.57 |
| AT1G66080 | unknown protein | 1.57 |
| AT1G02405 | proline-rich family protein | 2.05 |
| AT1G02450 | NIM1-INTERACTING 1 (NIMIN1) / NIMIN1 modulates PR gene expression | 2.15 |
| AT5G46490 | Disease resistance protein (TIR-NBS-LRR class) family | 2.25 |
| AT5G63145 | pre-tRNA | 2.64 |
| AT5G63140 | PURPLE ACID PHOSPHATASE 29 (PAP29) | 2.64 |
| AT5G63150 | unknown protein | 2.64 |
| AT1G28600 | GDSL-like Lipase/Acylhydrolase superfamily protein | 2.79 |
| AT3G32940 | RNA-binding KH domain-containing protein | 2.96 |
| AT3G23380 | ROP-INTERACTIVE CRIB MOTIF-CONTAINING PROTEIN 5 (RIC5) / Interacts with GTP-bound Rop1 (Rho GTPase) | 3.19 |
| AT4G12620 | ORIGIN OF REPLICATION COMPLEX 1B (ORC1B) / Involved in the initiation of DNA replication | 3.21 |
| AT1G29357 | Potential natural antisense gene, locus overlaps with AT1G29355 | 3.31 |
| AT1G29355 | unknown protein | 3.31 |
| AT1G40104 | unknown protein | 3.84 |
| AT1G22130 | AGAMOUS-LIKE 104 (AGL104) / Encodes a member of the MIKC family of transcriptional regulators | 4.08 |
| AT1G22140 | unknown protein | 4.08 |
| AT3G44805 | TRAF-like superfamily protein | 6.56 |
| AT3G53940 | Mitochondrial substrate carrier family protein | 7.27 |

| | | |
|---|---|---|
| AT5G10950 | Tudor/PWWP/MBT superfamily protein | 9.27 |
| AT4G35520 | MUTL PROTEIN HOMOLOG 3 (MLH3) / DNA mismatch repair protein similar to MutL. Required for normal levels of crossovers | 10.16 |
| AT4G35519 | unknown protein | 10.16 |

The list shows H3K72me3 differentially-bound genes from comparison between wildtype (Col-0) and *pwo1*, with significance level: FDR < 0.05. Column labels correspond to AGI locus identifier, TAIR10 gene model description and log2 ratio *pwo1*/wildtype.

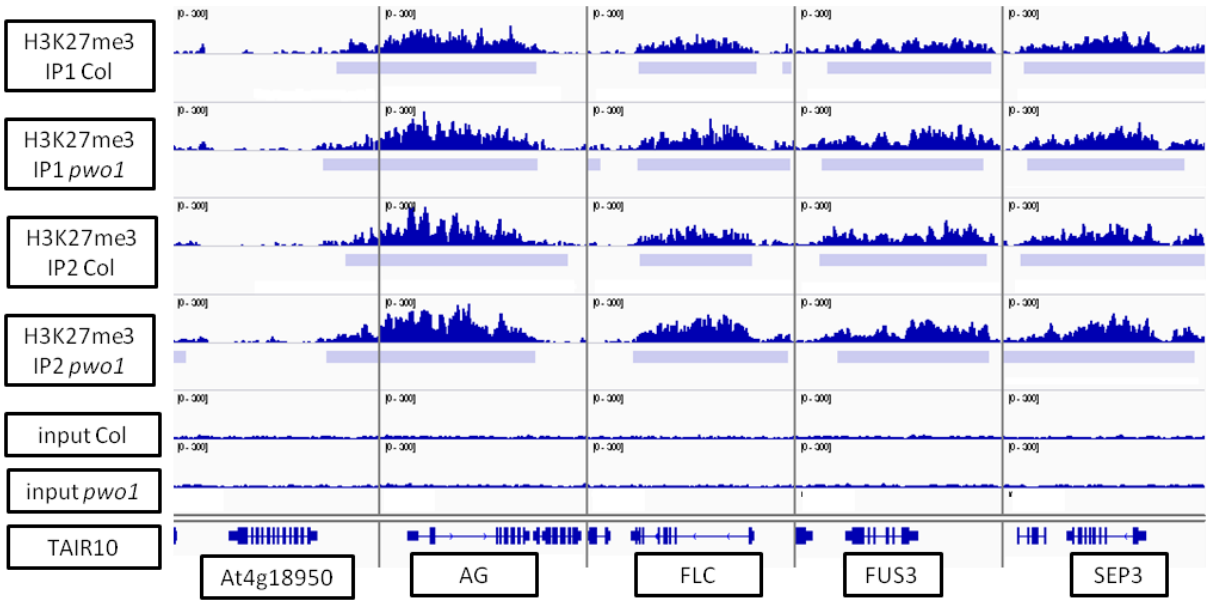



**Fig.S3. The H3K27me3 levels over PRC2 targets selected from ChIP-RT-qPCR analysis.** Bars below H3K27me3 tracks correspond to domains identified by MACS2 peak caller with input used for normalization. Visualization was performed in IGV v2.3 software.

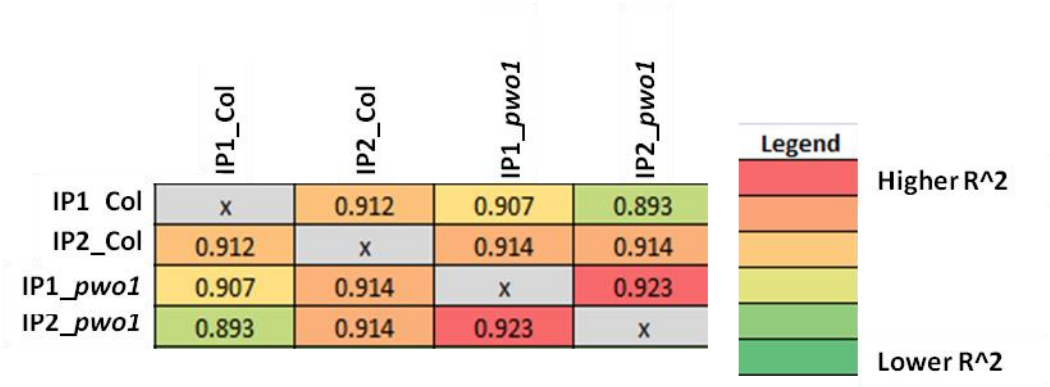



**Fig.S4. Correlation between samples in ChIP-seq experiment.** Heatmap represents correlation matrix based on R-squared (R^2) values computed from H3K27me3 read counts in 300bp-long genome bins from wildtype ("Col") and *pwo1* genotypes in two biological replicates (IP1 and IP2).

**Table S2. Oligonucleotides used in the study.**

| purpose | locus | forward | reverse |
|---|---|---|---|
| **genotyping (wildtype allele)** | | | |
| | *pwo1-1* | GCAGCACAAGGGAAAAAGAA | GTGCCCTTGTCTTTTGGCTA |
| | *pwo3-2* | ATGGGTAGTAGTGATGAGCGAAACTG | GACGGAGAGCAATCCCTGCTACT |
| **genotyping (transgene)** | | | |
| | *pwo1-1* | GCCTTTTCAGAAATGGATAAATAGCCTT GCTTCC | GTGCCCTTGTCTTTTGGCTA |
| | *pwo3-2* | ATGGGTAGTAGTGATGAGCGAAACTG | GCCTTTTCAGAAATGGATAAATAGC CTTGCTTCC |
| **ChIP-qPCR** | | | |
| | At4g18950 | CTCAGATGAGGCCATTGACA | GGCTGCTCAAATCCCAATTA |
| | AG | GGGAAACAAATTGGGGAGAG | CAACAATGGAGGATGGATGA |
| | FLC | GACCAGGCTGGAGAGATGAC | GTTTCCAGTGGCCTTTTCAA |
| | FUS3 | GTGGCAAGTGTTGATCATGG | AGTTGGCACGTGGGAAATAG |
| | SEP3 | GGGTTTCCAATTTTGGGTTT | GATGAATCCCATCCCCAAGT |
| | AT1G75550 | TCTTCATTGGTCGTTATCTTCCTC | GATCATTGGAACGGTCATGC |
| | AT1G66080 | GTGGTGTTGATGGAAGTAAATTGG | GGATCACGTTTAGCTTTCTCCT |
| | AT4G12980 | AACATCGTCTCTTACAGCTCTC | ATTCGTAAAGCTCCACCGTC |
| | AT1G50970 | AAAGATTTGTGACGAGATTCGG | GCTCTAGTTCCGGAGTTACC |
| | AT4G29770 | CCTCAGTATTTGGTGGATAACG | AGTAACTAAGGAGGTTTCCGAC |

### 3.3. Manuscript III

Phylogenetic conservation of Polycomb group proteins and H3K27me3 is extensively studied in higher plants and animals. However, the abundance of PcG homologs and function of H3K27me3 in lower plant branches remain largely unknown. In order to investigate PcG-mediated chromatin repression in lower plants, the following work concerned *in silico* identification PRC1 and PRC2 homologs in several representative species from lower eukaryotic branches. After the biochemical confirmation of H3K27me3 presence, the profile of this histone mark was determined for a selected species, *Cyanidioschyzon merolae*, using ChIP-seq and RNA-seq assays. In concordance to the other works, the results revealed overall high conservation of core PRC2 members and canonical role of H3K27me3 in chromatin repression in *C. merolae*. Interestingly, *C. merolae* H3K27me3 was found also to, i.e. silence predominantly repetitive elements, bind subtelomeric and telomeric regions and associate with the genes involved in protein splicing.

The results are described in Manuscript III: 'Characterization of PcG-mediated repression in unicellular algae'.

**Manuscript III**

## Characterization of PcG-mediated repression in unicellular algae

Pawel Mikulski[1,4], Olga Komarynets[2,4], Fabio Fachinelli[3], Andreas P.M. Weber[3], Daniel Schubert[1,4]

1 Institute of Biology, Free University, Berlin, Germany

2 Faculty of Medicine, University of Geneva, Geneva, Switzerland

3 Institute of Plant Biochemistry, Heinrich Heine University, Düsseldorf, Germany

4 Institute of Genetics, Heinrich-Heine-University, Düsseldorf, Germany

Keywords: Polycomb, PcG, H3K27me3, *Cyanidioschyzon merolae*, algae, intein, protein splicing, telomere

Author contributions:

PM, OK and DS designed the research. FF and APMW designed *Cyanidioschyzon merolae* growth scheme. FF handled *Cyanidioschyzon merolae* cell cultures. PM performed bioinformatic analyses, chromatin immunoprecipitations and sequencing library preparations. OK performed western blot experiments. PM and DS performed homology searches for H3K27me3 targets used in ChIP-RT-qPCR. The manuscript was written by PM and revised by DS.

### 3.3.1. Abstract

Polycomb Group (PcG) proteins mediate chromatin repression in plants and animals by catalyzing H3K27 methylation and H2AK118/119 monoubiquitination through the activity of the Polycomb repressive complexes 2 (PRC2) and PRC1, respectively. PcG proteins were studied extensively in higher plants, but their function and target genes in unicellular branches of the green lineage remain largely unknown. To shed light on PcG function and *modus operandi* in a broad evolutionary context, we demonstrate a phylogenetic relationship of core PRC1 and PRC2 proteins and H3K27me3 biochemical presence in several unicellular algae of different phylogenetic subclades. We focus then on one of the species, the model red alga, *Cyanidioschizon merolae*, and show that H3K27me3 occupies both, genes and repetitive elements, and mediates the strength of repression depending on the differential occupancy over gene bodies. Furthermore, we report that H3K27me3 in *C. merolae* is enriched in telomeric and subtelomeric regions of the chromosomes and has unique preferential binding toward intein-containing genes involved in protein splicing. Thus, our study gives important insights for Polycomb-mediated repression in lower eukaryotes, with previously unknown putative link between H3K27me3 targets and protein splicing.

### 3.3.2. Introduction

In the eukaryotic cells, transcription state is dependent on the underlying chromatin state. The fundamental structure of chromatin is based on a nucleosome, a complex of 147bp-long fragments of DNA wrapped around the core histone proteins (H2A, H2B, H3, H4). Chromatin state can be influenced by post-translational modifications deposited on the histones, either by direct structural changes in the nucleosomes or recruitment/displacement of secondary proteins involved in chromatin remodelling or transcription. The presence of particular histone modifications often determines the type of the chromatin and correlates with transcriptional activity of the target DNA. Among those, trimethylation of lysine 27 on histone H3 (H3K27me3) and monoubiquitination of histone H2AK118/H2AK119 (H2AK118ub/H2AK119ub) are commonly associated with transcriptionally-silent facultative heterochromatin.

Deposition of H3K27me3 and H2AK118ub/H2AK119ub is mediated by Polycomb group (PcG) proteins, whose function was initially shown to control developmentally regulated

processes and maintain cell identity (Papp and Müller, 2006; Schuettengruber et al., 2007) in both, animals (reviewed in: (Schwartz and Pirrotta, 2008)) and plants (reviewed in: (Köhler and Villar, 2008)). PcG proteins form distinct complexes, like Polycomb Repressive Complex 2 (PRC2), involved in H3K27 trimethylation (in metazoans Pcl-PRC2 complex (Müller and Verrijzer, 2009; Nekrasov et al., 2007)), and Polycomb Repressive Complex 1 (PRC1), responsible for H2A monoubiquitination. PRC2 in *Drosophila melanogaster*, where it was initially discovered, consists of four core subunits: Enhancer of zeste (E(z)), a catalytic component needed for methylation of H3K27; Extra Sex Combs (ESC), a WD40 motif-containing protein that scaffolds interactions within the complex; Suppressor of zeste (Su(z)12), a Zinc Finger subunit essential for binding of PRC2 to nucleosomes and p55, a nucleosome remodelling factor (Schwartz and Pirrotta, 2013). In turn, PRC1 is formed by: dRING/Sex Combs Extra (Sce); Posterior Sex Combs (Psc), both responsible for mono-ubiquitination activity; Polyhomeotic (Ph), essential for maintaining protein-protein interactions; and Polycomb (Pc),involved in recruitment of the complex to chromatin; and Sex comb on midleg (Scm), important for spreading of PcG silencing (Schwartz and Pirrotta, 2013). Both complexes are functionally connected with each other. In the canonical hierarchical model, initially introduced H3K27me3 mark is recognized by PRC1 and followed by H2A monoubiquitination to maintain the repression. However, it was shown that the mechanism of PcG-mediated repression can happen as well in the opposite order, with PRC1 introducing its modification prior to PRC2 activity (Del Prete et al., 2015; Yang et al., 2013).

Polycomb group complexes were identified also in the plant kingdom. Core PRC1 complex in model flowering plant *Arabidopsis thaliana* consist of AtRING1a/b (equivalent to dRING/Sce), AtBMI1a/b/c and EMF1 (equivalent to Psc), and LHP1 (equivalent to Pc). Although a wide range of PRC1-associated proteins exists (Yong et al., 2016), it is unclear how the combinations of different subunits depend on specific temporal and spatial conditions. In contrast to PRC1, composition of the *Arabidopsis* PRC2 complex is better characterized. *Arabidopsis* PRC2 subunits underwent gene duplication, resulting in the presence of different, partially redundant or independently acting PRC2 complexes (Derkacheva and Hennig, 2014). In the *Arabidopsis* genome the following PRC2 components can be identified: three E(z)-homologs – CURLY LEAF (CLF), SWINGER (SWN), MEDEA (MEA); three Su(z)12-homologs – EMBRYONIC FLOWER 2 (EMF2), VERNALIZATION 2 (VRN2), FERTILIZATION INDEPENDENT SEED 2 (FIS2); as well as single gene

subunits: ESC-homolog – FERTILIZATION INDEPENDENT ENDOSPERM (FIE) and p55-homologs – MULTICOPY SUPPRESSOR OF IRA 1-5 (MSI1-5).

One of the aspects of PRC2 biology concerns its origins and the abundance in lower eukaryotes. Due to the absence of PRC2 in the model unicellular species: *Saccharomyces cerevisiae* and *Saccharomyces pombe* (Lachner et al., 2004; Veerappan et al., 2008), the PRC2 appearance was previously thought to co-occur with the emergence of multicellularity. However, high conservation of PRC2 subunits in both, animal and plant lineages implies that PcG genes appeared already in the last common unicellular ancestor (Bowman et al., 2007), before those two kingdoms diverged. In an elegant study, Shaver *et al* (Shaver et al., 2010) identified the novel computational PRC2-homologs in several unicellular species and showed that E(z)-homolog in unicellular green algae *Chlamydomonas rheinhardtii* is responsible for mono- and dimethylation of lysine 27 on histone H3. As no trimethylation activity of E(z) in *Chlamydomonas* was found, this study highlights important differences in the functional conservation of PRC2 components. For instance, although H3K27 trimethylation is considered to be a prominent histone mark in PRC2-mediated repression, there is a range of H3K27 methylation levels catalyzed by PRC2 that varies between plant and *Drosophila* or mammals. In *Arabidopsis*, PRC2 controls H3K27 trimethylation and H3K27 dimethylation in the euchromatin (Bastow et al., 2004; Lindroth et al., 2004; Sung et al., 2006), with mono-methylation being catalyzed by ATXR5 and ATXR6 (Jacob and Michaels, 2009; Jacob et al., 2009). However, in metazoans no homologues of ATXR5/6 were not found and their PRC2 complexes mediate H3K27methylation in all contexts (Ebert et al., 2004; Ferrari et al., 2014; Müller et al., 2002). Another difference between species is the type of DNA elements targeted by PRC2. *Drosophila* PRC2 targets both, genes and transposable elements/repetitive sequences (Yin et al., 2011), whereas in *Arabidopsis* the H3K27me3 mark is excluded from the majority of transposable elements/repetitive sequences (Deleris et al., 2012; Lafos et al., 2011; Park et al., 2012).

Even though PRC2-mediated repression has been extensively studied in higher plants and animals, its characterization in lower eukaryotes gathered considerably less attention. Existence of the H3K27me3 in lower eukaryotes is largely unknown (Shaver et al., 2010) and distribution of the mark in the genome and its function were reported only for a handful of species, such as *Neurospora crassa*, P*haeodactylum tricornutum* or *Tetrahymena thermophila*. In *N. crassa* H3K27me3 arranges into broad domains covering 774 genes connected with a full spectrum of functions (Jamieson et al., 2013). In *T. thermophila*, the

mark associates with the developmentally regulated genome rearrangements as it occupies sequences eliminated during differentiation of the macronucleus (Liu et al., 2007). In *P. tricornutum*, H3K27me3 covers transposable elements and genes, and its targets are involved in i.e. signal transduction, development and cell cycle control (Veluchamy et al., 2015).

Given the distinct degrees of phylogenetic conservation of PRC2 and PRC1, and poor characterization of H3K27me3 in lower eukaryotes, we sought to investigate the presence of Polycomb group homologs and H3K27me3 abundance in several representative unicellular, photosynthetically active eukaryotes. Our scope included the representatives of red algae, green algae and Glaucophyta that vary in genome size, genome architecture, ecological niche and metabolism (Matsuzaki et al., 2004; Merchant et al., 2010; Palenik et al., 2007; Price et al., 2012; Worden et al., 2009). After an initial small-scale screen on H3K27me3 presence, we focused on one of the species, *Cyanidioschyzon merolae*, to study genome-wide location of H3K27me3 and characterize its targets. *Cyanidioschyzon merolae* is a unicellular red alga, living in highly acidic environment with high temperatures. It contains small (16Mb) genome which was fully sequenced as the first algal genome (Matsuzaki et al., 2004) and assembled as first 100% complete eukaryotic genome (Nozaki et al., 2007). Interestingly, the genome shows extremely simplified structure that contains almost exclusively intron-lacking genes (only 26 genes contain an intron), very low percentage (0,7%) of transposable elements and a novel class of a repetitive element, corresponding to the truncated ORF from White spot syndrome virus (WSV repeat) (Nozaki et al., 2007). Given the unicellularity, evolutionary ancestry, primitive architecture of the genome and availability of sequencing data, we argue that *C. merolae* is a suitable model for studying chromatin repression in the evolutionary context.

Our results confirmed high conservation of PRC2 core members in lower eukaryotes and widespread presence of H3K27me3 modification. We report that in *C. merolae* H3K27me3 mark targets both, genes and repetitive elements and is anti-correlated with transcriptional activity. We show that H3K27me3 distribution over its targets is not uniform, but can be grouped into several different clusters that correlate with different levels of repression. Furthermore, we demonstrate that H3K27me3 has a preferential chromosomal localization toward telomeric and subtelomeric regions. For the first time, we also reveal that H3K27me3 target genes are enriched in the functional class of intein-mediated protein splicing. Moreover, by deposition of RNA- and ChIP-sequencing data, we provide a resource for studies on the chromatin and transcriptome in the model red alga *C. merolae*.

### 3.3.3. Results

**Conservation of PRC1 and PRC2 homologs in in green lineage**

In order to assess evolutionary conservation of Polycomb complexes, we performed phylogenetic analysis in the species from various groups of lower plants. We focused on the representatives from subclades:

Chlorophyta (green algae): *C. rheinhardtii, M. pusilla, O tauri*

Rhodophyta (red algae): *C. merolae*

Glaucophyta: *C. paradoxa*

For comparison, a representative from Metazoa, *Drosophila melanogaster*, was included as well. Using reciprocal BLAST searches with full length proteins from *Arabidopsis thaliana* or *Drosophila melanogaster*, we identified homologs of PRC1 and PRC2 members in several species (Table1).

**Table 1. Number of identified homologs of PRC2 and PRC1 components in selected species**

| | E(z) | ESC | Suz 12 | P55 | BMI1 /Psc | RING1A / Sce | LHP1 / Pc | EMF1 | Scm | H3 |
|---|---|---|---|---|---|---|---|---|---|---|
| *D. melanogaster* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | ~100 |
| *A. thaliana* | 3 | 1 | 3 | 6 | 3 | 2 | 1 | 1 | 0 | 15 |
| *C. reinhardtii* | 1 | 2 | 0 | 2 | 1 | 1 | 0 | 0 | 0 | 35 |
| *M. pusilla* | 1 | 1 | 1 | 2 | 0 | 1 | 0 | 0 | 0 | 5 |
| *O. tauri* | 1 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 4 |
| *C. merolae* | 1 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 4 |
| *C. paradoxa* | 0 | 0 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 12 |

We found a general absence of PRC1 components: Psc/BMI1, Pc/LHP1, EMF1 and Scm in the algal genomes analyzed, with an exception of Psc/BMI1 homolog in *C. reinhardtii*. In

contrast, RING1 subunit is widely conserved, suggesting a monoubiquitylation activity unrelated to PRC1 in algae and protein moonlighting (Jeffery, 2003). In general, an abundance of RING1 homologs in chlorophytes, but the absence of other PRC1 components in lower branches of Archeaplastida used here is consistent with the previous studies (Berke and Snel, 2015; Hennig and Derkacheva, 2009; Yong et al., 2016) and agrees with the notion of lower PRC1 conservation and potential loss of the complex in several phylogenetic branches.

Homology searches on PRC2 revealed broad distribution of complex components: E(z), ESC and p55, consistently with results published elsewhere (Butenko and Ohad, 2011; Kim et al., 2013; Shaver et al., 2010). Importantly, p55 in higher eukaryotes has various functions and participates also in other complexes than PRC2 (i.e. chromatin assembly factor 1 (CAF1)). Therefore it still remains to be proven whether a role of p55 homologues is Polycomb-related in lower plants. Moreover, we observed an absence of Suz12 from *C. rheinhardtii* and *C. paradoxa*. Similarly to the other studies (Shaver et al., 2010), the lower abundance of this subunit shows that Suz12 was lost in several species, despite the presence of the other PRC2 components. Interestingly, we did not detect any core PRC2 members in *Cyanophora* representative. Given that E(z) and ESC homologs were present in the representatives of Rhodophyta and Chlorophyta, it suggests the existence of PRC2 in the common algal ancestor and subsequent loss of the complex from *Cyanophora*. However, we could not exclude technical issues coming from incomplete annotation and incorrectly predicated protein models in the *Cyanophora* genome. Apart from that, we identified mostly one homolog per complex component in all screened species, apart from *Arabidopsis*, which adds another piece of evidence to frequent gene duplication occurring in flowering plants.

Overall, our results agree with the ancient presence of core PRC2 components and the frequent losses of the Suz12 subunit and PRC1 complex members.

**Phylogenetic relationship of PRC2 homologs**

Phylogenetic analyses on the conserved domains or full sequences of core PRC2 homologs in lower and higher plants with the creation of Neighbour-Joining (NJ) trees revealed grouping into several distinct classes (Huang et al., 2016; Kim et al., 2013; Shaver et al., 2010). In a different approach, we extracted amino acid sequences from the most abundant PRC2

members, aligned them and automatically selected only the conserved blocks in the multiple alignment (see material and methods). With such prepared sequences we created Bayesian trees for E(z), ESC and Su(z)12 homologs. Consistently with the published data, the homologs formed defined clades. Bayesian trees for E(z) and ESC sequences generally resolved the evolutionary distance between groups (Figure 1A,B). In those cases, we found separation of representatives of Rhodophyta and Viridiplantae, with species from Chlorophyta and Embryophyta forming separate subclades within the latter. The exceptions included one of the ESC homologs from *C. rheinhardtii* (CrESC.1), which clustered together with the *C. merolae* homolog, rather than the other green algae. In turn, our analysis on Su(z)12 detected distinct clades for the organisms from Embryophyta, Chlorophyta and Rhodophyta, without grouping of *Arabidopsis* and green algae sequences (Figure 1C). Moreover, we noted a phylogenetic distance between *C. merolae* and any other species for all the three proteins.

In summary, our results on Bayesian phylogenetic trees on conserved sequence blocks are in agreement with the published Neighbour-Joining trees using the full length (Kim et al., 2013) or domain-only (Shaver et al., 2010) sequences. Given high conservation of domain architecture in core PRC2 components between the subclades (Kim et al., 2013; Shaver et al., 2010), we conclude that PRC2 is widely distributed and already evolved in a common unicellular ancestor.

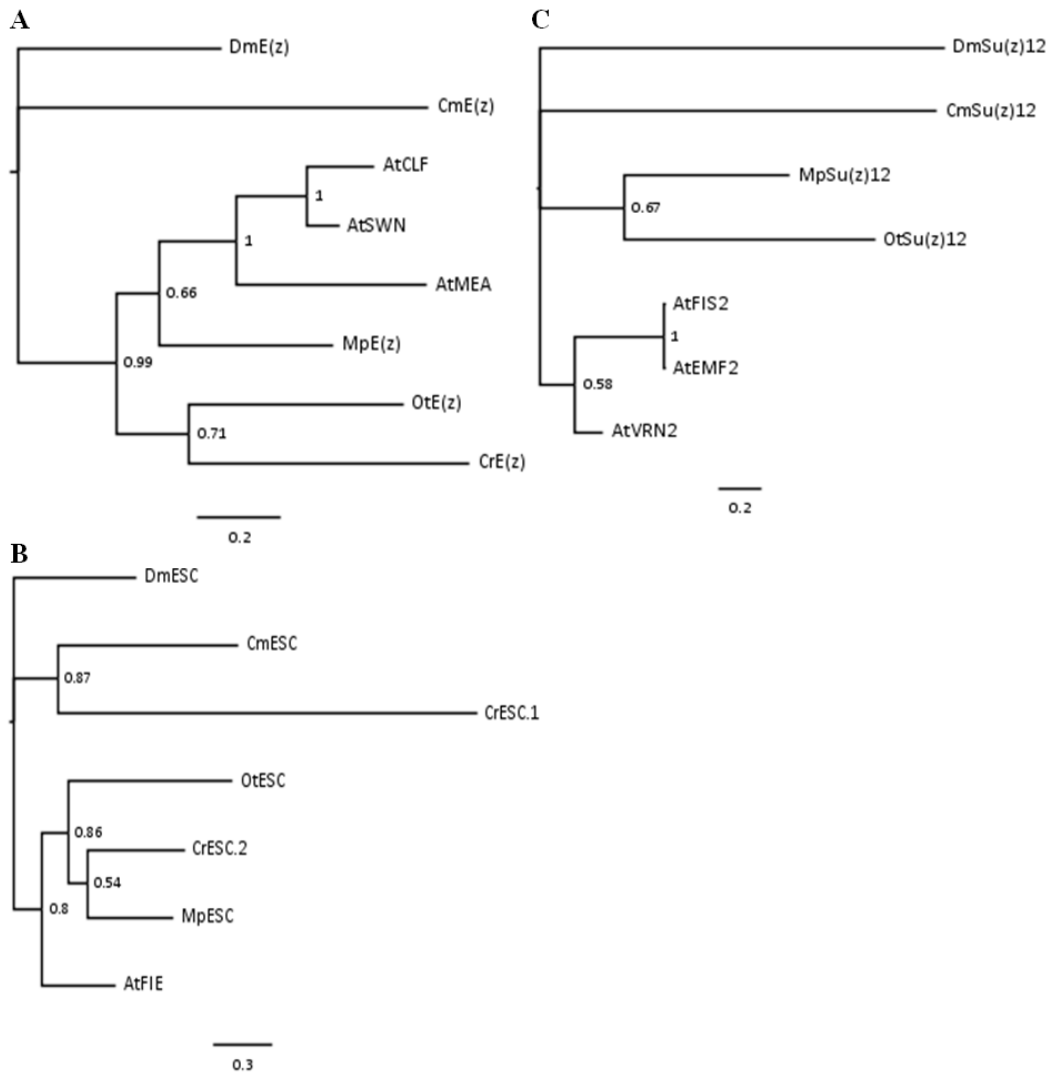



**Figure 1. Bayesian-inferred phylogenetic trees on conserved blocks from the homologs of core PRC2 components.** Consensus trees were generated by MrBayes v2 and visualized in FigTree v1.4. Clade credibility values are depicted at the nodes. Branch length corresponds to the rate of expected mutations per site in the protein sequence, according to relevant scale bar below each tree. Sequences from *D. melanogaster* were selected as outgroups. Two-letter abbreviations represent species' names: Dm – *D. melanogaster*; Cm – *C. merolae*; At – *A. thaliana*; Mp – *M. pusilla*; Ot – *O. tauri*; Cr – *C. reinhardtii*. (A) Phylogenetic tree on E(z) homologs. Conserved blocks selected from multiple alignment on following homologs (UniProtKB ID or protein/transcript name from resource databases (see material and methods) indicated: DmE(z) - P42124; CmE(z) - CMQ156C; AtCLF - AT2G23380; AtSWN - AT4G02020; AtMEA - AT1G02580; MpE(z) – 59369; OtE(z) – 6642; CrE(z) - Cre17.g746247.t1.1. (B) Phylogenetic tree on ESC homologs. Homologs included in the analysis: DmESC - Q24338; CmESC - CMK173C; AtFIE - AT3G20740; MpESC – 49065; OtESC – 22117; CrESC.1 - Cre16.g693750.t1.1; CrESC.2 - Cre03.g180050.t1.1. (C) Phylogenetic tree on Su(z)12 homologs. Homologs included in the analysis – DmSu(z)12 - Q9NJG9; CmSu(z)12 - CML082C; AtFIS2 - AT2G35670; AtEMF2 - AT5G51230; AtVRN2 - AT4G16845; MpSu(z)12 – 9357; OtSu(z) – 13623.

## Conservation of histone H3 sequences

As PRC2 has the canonical function to methylate lysine 27 in histone H3, we decided to

analyse protein sequence conservation of histone H3. Amino acid sequences of algal histone H3 were retrieved from the databases of respective genome sequencing projects (see materials and methods) or NCBI, using the canonical histone H3.1 from *Arabidopsis* as a query (Okada et al., 2005). The results revealed a presence of multiple H3 genes in all of the species studied (Table 1). Noteworthy, we detected an equal amount of histone H3 gene copies as in the published reports for *M. pusilla* and *C. reinhardtii* (Cui et al., 2015). The number of H3 gene copies in *Drosophila* was taken from the published results (McKay et al., 2015).

Next, the closest H3 homologs between the species were aligned using ClustalX and visualized with Jalview. The alignment revealed overall high amino acid conservation and a presence of lysine-27 in all studied species (Figure 2). Moreover, we noted that the sequences around lysine-27, including the underlying motif ARKS, does not show any sequence divergence with an exception of *C. rheinhardtii*, which contains threonine-28 instead of serine-28. The results suggested that there is the potential for introduction of H3K27me3 modification in the most of species analyzed. However, in *C. rheinhardtii* the presence of threonine-28 next to lysine-27might not permit detection using commercial antibodies against the methylated H3K27. Moreover, a mass spectrometry analysis on *Chlamydomonas* histone H3 did not reveal the presence of H3K27me3 modification (Shaver et al., 2010).

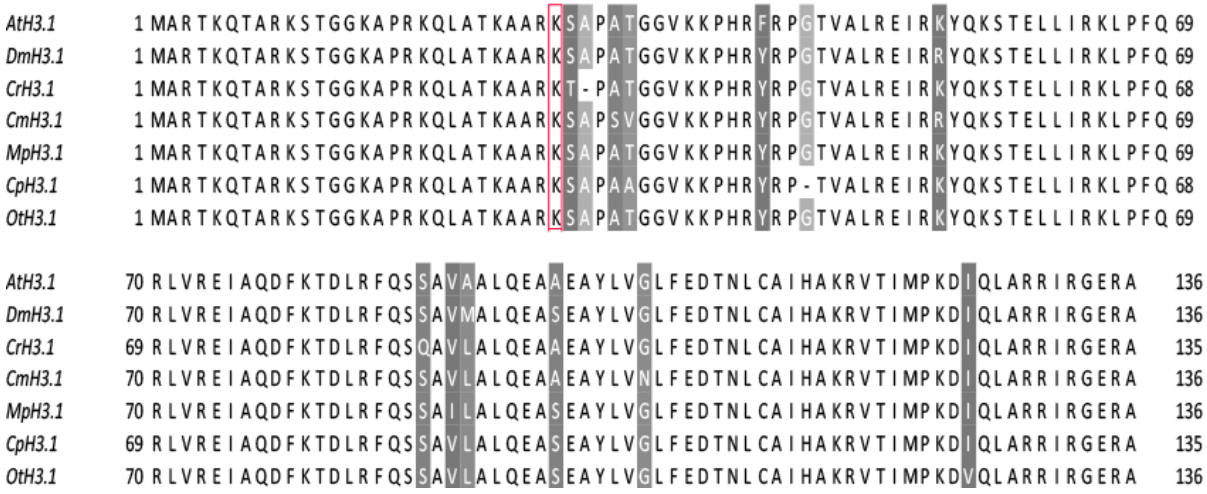



**Figure 2. H3 sequence conservation. AtH3.1 was used a query for BLAST searches against databases of respective species.** The closest homologs were selected and used for multiple alignment in ClustalX, visualized afterwards in Jalview v2.9. Colour-shading marks non-conserved residues. Red frame correspond to lysine-27, a residue typically modified by PRC2.

**Detection of H3K27me3 modification**

Given the phylogenetic conservation of PRC2 core components and high similarity of amino acid composition of histone H3 in the vast majority of the species, we sought to examine H3K27me3 presence in total protein extracts from selected organisms. Isolated proteins from crude extract were separated on a SDS-PAGE gel and detected using two independently raised anti-H3K27me3 antibodies. Our results revealed the presence of the H3K27me3 modification in *M. pusilla*, *O. tauri*, *C. paradoxa* and *C. merolae* (Figure 3A). In order to decipher H3K27me3 relative abundance, we performed western blotting for histone H3 and calculated H3K27me3/H3 ratios based on band intensity quantification. We detected different relative amounts of the modification in the studied species (Figure 3B), with the lowest H3K27me3/H3 ratios for *M. pusilla* and *O. tauri*, intermediate for *C. merolae* and the highest for *C. paradoxa*.

Due to the fact that one of the antibodies detected proteins in the sizes not corresponding to histones, we performed protein competition assay for the confirmation of specific band reactivity. After preincubation with H3K27me3 peptide, we did not detect any band corresponding to this modification for a commercial histone extract from calf thymus (positive control), as well as for *M. pusilla*, *O. tauri* and *C. merolae* (Figure 3C,D), suggesting a specificity of the antibody. Intensity decrease of the band corresponding to H3K27me3 for *C. paradoxa* was accompanied by overall signal loss, including the bands of higher molecular weight proteins. Therefore, the presence of H3K27me3 in *C. paradoxa* remains inconclusive.

In summary, we were able to biochemically identify and confirm a presence of H3K27me3 in *M. pusilla*, *O. tauri* and *C. merolae*. This result is consistent with our phylogenetic analyses and suggests that the PRC2 complex is active in these species. For *C. paradoxa*, the biochemical results are unclear: potential absence of H3K27me3 mark would correspond to the lack of PRC2 homologues identified in its genome. Differences in relative H3K27me3 amounts between species imply, e.g. a distinct abundance of genomic targets of H3K27me3 or different H3K27me3 enrichment levels per genomic region. Further work should decipher an impact of such differences in the evolutionary context.

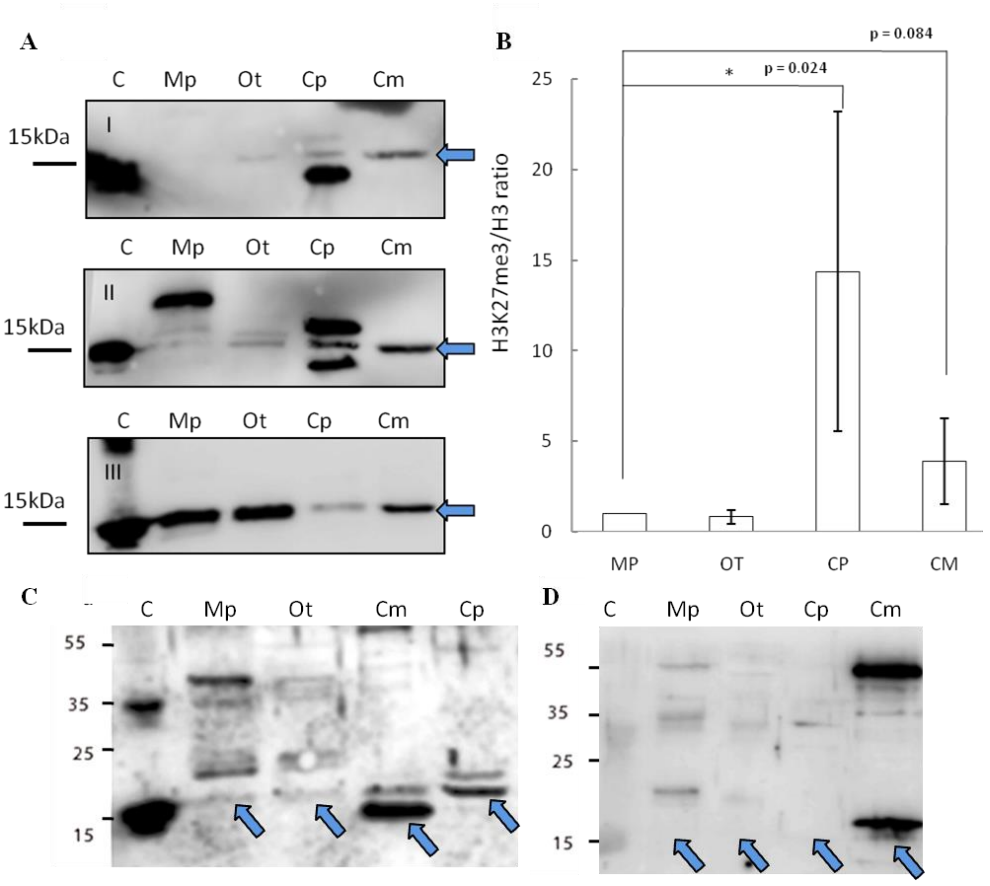



**Figure 3. Biochemical detection of H3K27me3 histone mark.** (A) Western blot analysis of H3K27me3 in different algae (C-histone powder control, Mp – *M. pusilla*, Ot – *O. tauri*, Cp - *C. paradoxa*, Cm - *C. merolae*). Membranes were incubated with antibodies against H3K27me3 (Millipore, N07-449 (I) and Diagenode, C15410206 (II)) and H3 (Abcam, ab1791-100 (III)). The arrows indicate the band of interest. (B) Quantification of western blotting. Intensity of the bands was obtained using ImageJ. H3K27me3/H3 protein level ratio of three independent experiments is shown. P-values were calculated using Student t-test. Error bars represent standard deviation. Figures (C) and (D) corresponds to the competition assay. (C) Western blotting with primary antibodies against H3K27me3 (Millipore, N07-449). (D) Western blotting with antibodies pre-incubated with methylated-H3K27 (Intavis) peptide. The arrows indicate a band of interest corresponding to H3K27me3.

## Characterization of H3K27me3 target genes

Based on the broad distribution of PRC2 genes, high sequence conservation of histone H3 and specific signal detected with the anti-H3K27me3 antibody on a protein blot, we selected one of the species, *Cyanidioschyzon merolae*, to investigate the H3K27me3 abundance further. Firstly, we asked whether the modification has similar targets as in the other organisms. Reciprocal BLAST searches between protein databases of *C. merolae* and *A. thaliana* were used to detect homologous genes of known *A. thaliana* PRC2 targets. (Figure S1). We selected a MADS box-containing gene CMA095C (CmMADS) as nearly all

*Arabidopsis* MADS box genes carry H3K27me3. For negative controls, estimated not to be targeted by H3K27me3, homologous genes for EUKARYOTIC ELONGATION FACTOR EIF4A (CMK028C, CmEIF4A) and 60S RIBOSOMAL PROTEIN L23 (CMS262C, Cm60S) were taken.

Next, we performed Chromatin-immunoprecipitation (ChIP) using anti-H3K27me3 antibody and analysed the expression of candidate genes by RT-PCR. As a result, we were able to show significant enrichment of H3K27me3 on CmMADS and low abundance of the mark on negative targets: CmEIF4A and Cm60S (Figure 4). Overall, we were able to identify positive target of H3K27me3 and show high enrichment of the mark comparing to negative loci.

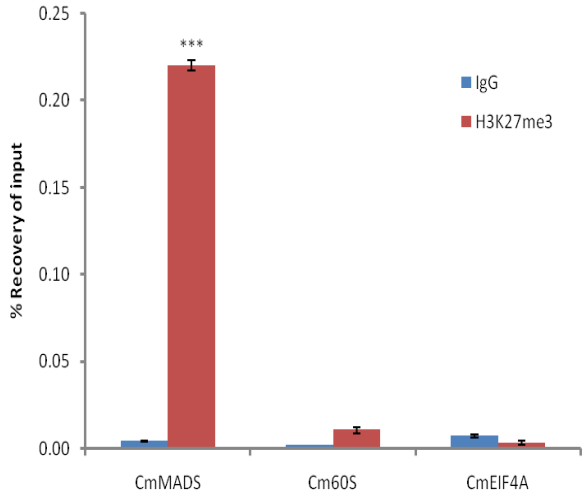



**Figure 4. H3K27me3 abundance on selected genes from *C. merolae*.** Chromatin immunoprecipitation was performed using H3K27me3 antibody (#C15410195, Diagenode) and IgG (#C15410206, Diagenode) for negative control, followed by RT-PCR analysis on a putative H3K27me3 target (CmMADS) and housekeeping genes as the putative non-H3K27me3 targets (Cm60S, CmEIF4A). Selected genes relate to following locus IDs from *C. merolae* genome project: CmMADS – CMA095C, Cm60S - CMS262C, CmEIF4A - CMK028C. Error bars correspond to the standard deviation from three biological replicates. Significance was calculated using Student's t-test for the comparison between CmMADS and either, Cm60S or CmEIF4A.

## H3K27me3 genome-wide distribution – peak identification

In order to characterize the targets of H3K27me3 and distribution of H3K27me3 peaks in a genome-wide scale, we created a sequencing library and performed chromatin-immunoprecipitation coupled with sequencing (ChIP-seq) for the samples: H3K27me3-bound DNA, H3-bound DNA (both in triplicate) and input.

Cleaning raw sequencing reads, mapping to the reference genome and peak calling with input or H3 sample for normalization, let us identify more than 1300 H3K27me3 peaks per replicate (Figure 5A). The peaks cover 14% of total nuclear genome length, with an average peak size of 1792bp.

Next, we compared H3K27me3-peak coordinates with the loci in *C. merolae* reference genome. We applied a threshold of 50% overlap of the locus length and successfully annotated ca. 600 peaks ('high overlap annotated peaks') (Figure 5A). The remaining peaks corresponded to the reference loci covered by H3K27me3 in less than 50% of their length ('low overlap annotated peaks') and to the unannotated peaks without any overlap with reference loci. We concluded that the unannoted peaks are the intergenic regions and/or uncharacterized genomic features missing in the current annotation. Considering the second possibility, we used several strategies to find *de novo* annotation of unknown peaks.

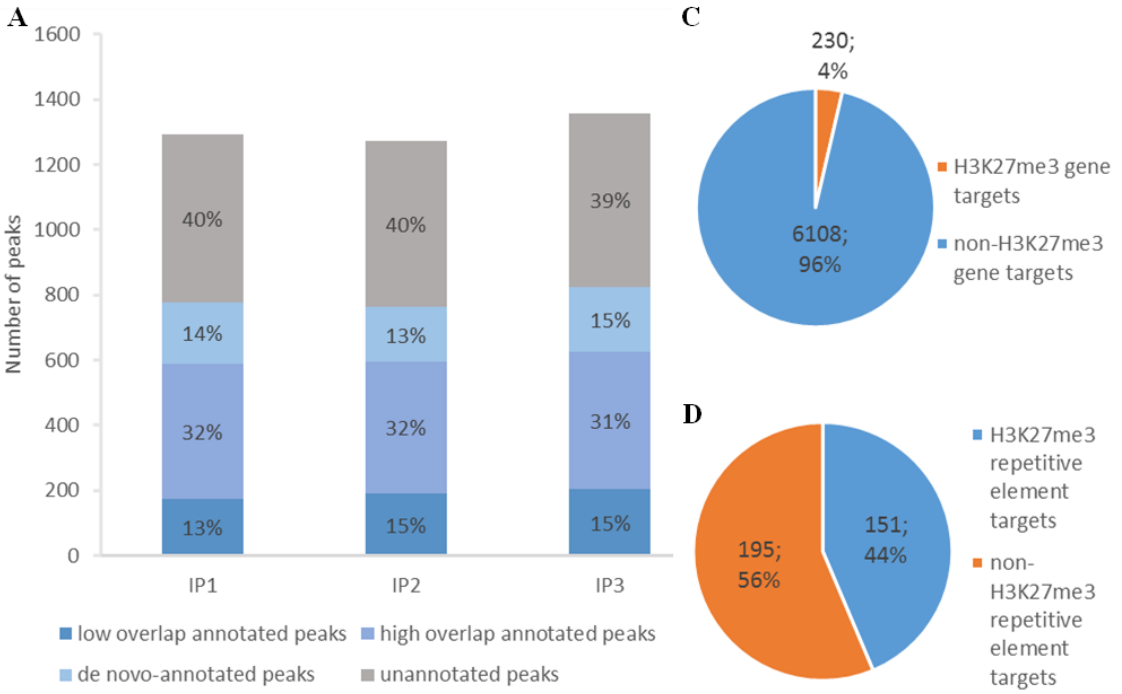
Firstly, we *in silico* translated sequences from the unannotated peaks and searched for open reading frames (ORFs) within them. We found ORFs for all of the unannotated peaks and used them for BLASTP/Pfam searches against the NCBI non-redundant protein sequence database (nr database) or Pfam protein database. Using a specific alignment threshold (alignment length > 50%; E-value < 0.5), we could annotate further 178, 164 and 195 peaks for biological replicate 1, 2 and 3, respectively (Figure 5B). In the second approach, we extracted DNA sequences of the unknown peaks and used BLASTN searches against the NCBI nucleotide collection (nr/nt database) and NCBI transcript reference sequences (refseq_rna database). The hits surpassing the alignment threshold (alignment length > 50%, E-value < 0.5) were compared to BLASTP/Pfam output, which allowed to annotate next 4 to 8 peaks, depending on the biological replicate (Figure 5B). In general, our sequence homology searches helped to detect *de novo* annotation of 13 to 15% peaks from their total number, depending on the biological replicate (Figure 5A).

We reasoned that the remaining unknown peaks after *de novo* annotation correspond to the regions containing promoters and the other regulatory elements of neighbouring annotated elements. Given the high density of genes along the chromosomes in *C. merolae* (Matsuzaki et al., 2004), such regions would reside close to 5' or 3' gene ends, but would not be present in the annotation. In order to check this idea, we calculated average distances between gene locations and unannotated peaks. We compared distributions between peaks successfully aligned to BLASTP database (annotated peaks) and the ones showing no significant alignment (unannotated peaks), assuming that BLASTP-aligned peaks truly correspond to novel genes missing in our reference, rather than uncharacterized *cis* elements. However, we did not observe major differences between distances calculated for annotated and unannotated peaks (Table S1), suggesting that unannotated peaks do not exclusively correspond to *cis* regulatory elements of known genomic elements. Further development of

*C. merolae* genome annotation should deepen our understanding about these remaining uncharacterized H3K27me3 peaks.

**H3K27me3 genome-wide distribution – genomic feature**

Next, we sought to investigate the genomic feature profile underlying H3K27me3 peaks. We picked a stringent group of high overlap annotated targets, removed duplicates (two peaks spanning one genomic element) for each replicate and selected a consensus set of features that appear in any two out of three replicates. As a result, we found that H3K27me3 covers 230 genes, which amounts to 4% of *C. merolae* total gene number (Figure 5C). On the other hand, we also detected 170 H3K27me3-bound repetitive elements. Given very low repetitive element occupancy in the *C. merolae* genome (0,7% transposable elements + 5% WSV repeats) (Nozaki et al., 2007), H3K27me3 seems to be predominantly present on repetitive elements, covering as much as 44% of their total number (Figure 5D). We noted that the enrichment on both, genes and repetitive elements, is in concordance with the studies in *Drosophila* (Yin et al., 2011). Interestingly, the predominant enrichment of H3K27me3 on repetitive elements was found also in diatom *P. tricornutum* (Veluchamy et al., 2015), suggesting an ancestral role of Polycomb-mediated gene regulation in guarding the genome.




B

| | IP1 | IP2 | IP3 |
|---|---|---|---|
| Total number of peaks identified | 1293 | 1273 | 1356 |
| Low overlap annotated peaks | 174 | 191 | 203 |
| High overlap annotated peaks | 415 | 403 | 423 |
| BlastP/Pfam-annotated peaks | 178 | 164 | 195 |
| BlastN(nr)-annotated peaks | 8 | 5 | 4 |
| BlastN(refseq)-annotated peaks | 1 | 0 | 0 |
| Remaining unannotated peaks | 526 | 515 | 535 |

**Figure 5. H3K27me3 genome-wide distribution and peak annotation.** (A) and (B) - Number of peaks identified in 3 different biological replicates from chromatin immunoprecipitation experiment (IP1-3) in 'high overlap annotated peaks', 'low overlap annotated peaks' category, '*de novo*-annotated' and 'unannotated' categories (see main text). (C) and (D) - Characterization of H3K27me3 peaks according to genomic element type. High overlap H3K27me3 peaks cover 230 genes, which amounts to 4% of total gene number (C) or 151 repetitive elements, 44% of total repetitive element number (D). Both, loci of hypothetical proteins and transcripts, were included in total gene number.

## H3K27me3 is a silencing mark in *C. merolae*

Next, we wanted to characterize the correlation of H3K27me3 enrichment with the expression of its targets. In order to decipher expression level on a genome-wide scale, we performed a RNA-seq experiment on reverse-transcribed RNA extracted from *C. merolae* cultures. The sequencing reads were mapped to the reference genome and expression level was assessed based on the FPKM values. We intersected the FPKM data with H3K27me3-binding profile and distinguished genes and repetitive elements for the analyses.

We found that both, H3K27me3-covered genes and H3K27me3-covered repetitive elements are on average significantly less expressed than non-H3K27me3 targets (Figure 6), confirming that the H3K27me3 modification in *C. merolae* is highly correlated with gene repression. Moreover, we noted that the general level of repetitive element expression is lower than gene expression, irrespectively of the H3K27me3 status. These results suggest the existence of additional mechanisms involved in repression of the repetitive elements.

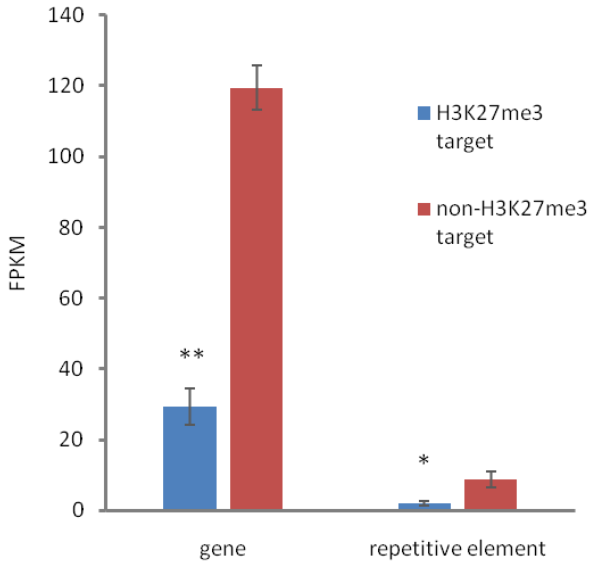



**Figure 6. Relationship between expression and H3K27me3 presence.** H3K27me3 occupancy is anti-correlated with expression level in both, genes and TEs. Expression was estimated based on FPKM values obtained from RNA-seq experiment. Student t-test was used to calculate significance. Double asterisk: p-value < 0.01, single asterisk: p-value < 0.05. Error bars represent standard error (SE).

## H3K27me3 gene-body distribution

H3K27me3 enrichment has a broad distribution over gene bodies in plants (Yang et al., 2014; Zhang et al., 2007) or gene bodies and flanking regions in animals (Cerase et al., 2014; Pauler et al., 2009) and diatoms (Veluchamy et al., 2015). To inspect the distribution of the mark in *C. merolae*, we calculated average H3K27me3 enrichment over genic coordinates and observed that H3K27me3 covers gene bodies, 0.5kb upstream and 0.5kb downstream regions (Figure S2). This general approach obscured whether all targets show similar, broad distribution of H3K27me3 or whether the differential enrichment subgroups are present among the targets. We therefore performed k-means clustering on the results. Our analysis identified 4 clusters with the preferential presence of H3K27me3 at gene bodies (cluster 1), the 0.5kb downstream region (cluster 2) and the 0.5kb upstream region (cluster 3) (Figure 7A). Cluster 4 contains non-H3K27me3 targets.

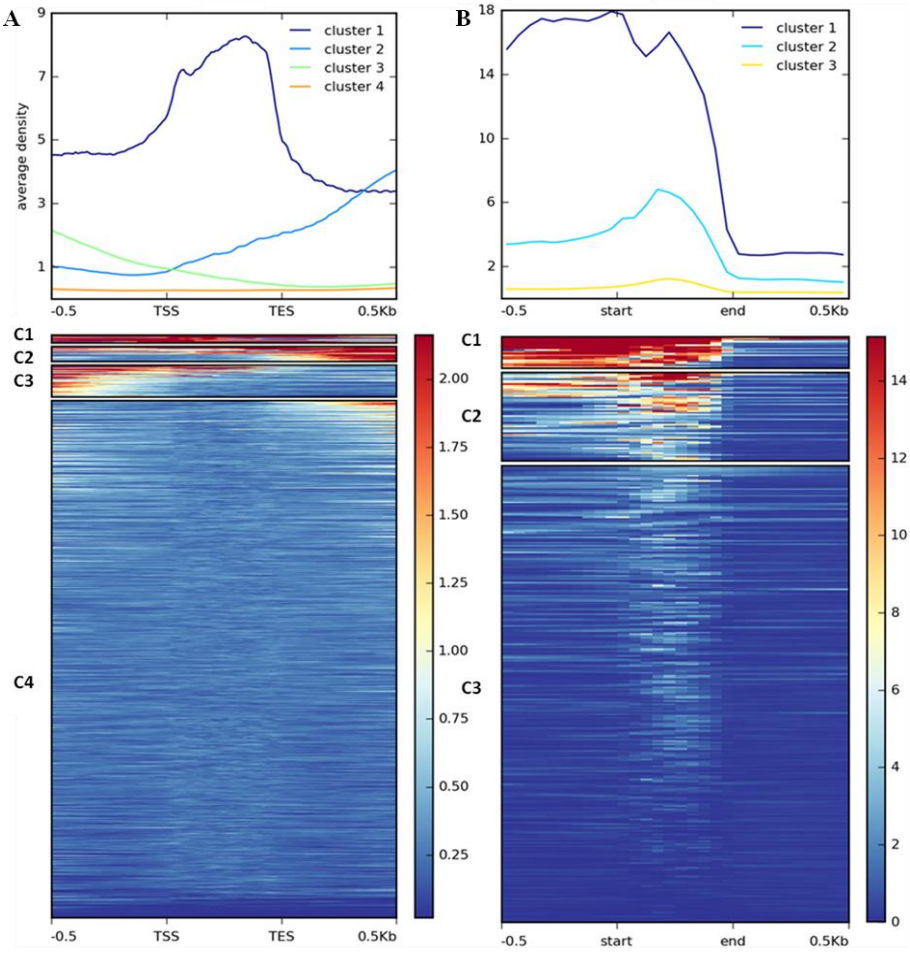



**Figure 7. Clustering of H3K27me3 gene body enrichment.** Coverage of ChIP-seq H3K27me3 reads normalized to input was calculated and used to generate matrixwith gene (A) or (B) repetitive element locations as reference. The reads were scaled to 500bp windows and flanking regions set to 500bp. Matrix files served as an input for heatmap generation and K-means clustering in deepTools2.

As distinct H3K27me3 enrichment clusters impact target gene activity differently in some other species (Veluchamy et al., 2015; Young et al., 2011), we asked whether gene expression level varies between the H3K27me3 clusters in *C. merolae*. H3K27me3 differential enrichment compared with RNA-seq data revealed that clusters 1-3 correlate with the gene repression, albeit to an uneven extent. Cluster 1 correlated with the strongest repression level, whereas clusters 2 and 3 correlated with a similar, milder repression (Figure 8A). Thus, the strongest silencing effect was seen for H3K27me3 distribution at the gene bodies, similarly to the diatom *P. tricornutum* (Veluchamy et al., 2015), and the H3K27me3 enrichment at the flanking regions is correlated with the milder repression.

We performed similar clustering analysis to characterize H3K27me3 enrichment also on the repetitive elements. We could distinguish 2 clusters with histone mark presence on 0.5kb

upstream region and gene bodies (cluster 1) or gene bodies only (cluster 2) (Figure 7B). Cluster 3 corresponds to non-H3K27me3 targets. In contrast to the genes, we found that cluster 1 and 2 correlate with reduced gene expression to the same extent (Figure 8B), suggesting that gene body H3K27me3 distribution has a predominant effect on gene repression. On the other hand, clusters present at repetitive elements are not fully analogous to those at genes as for the former we did not identify a cluster with enrichment exclusively at the flanking regions.

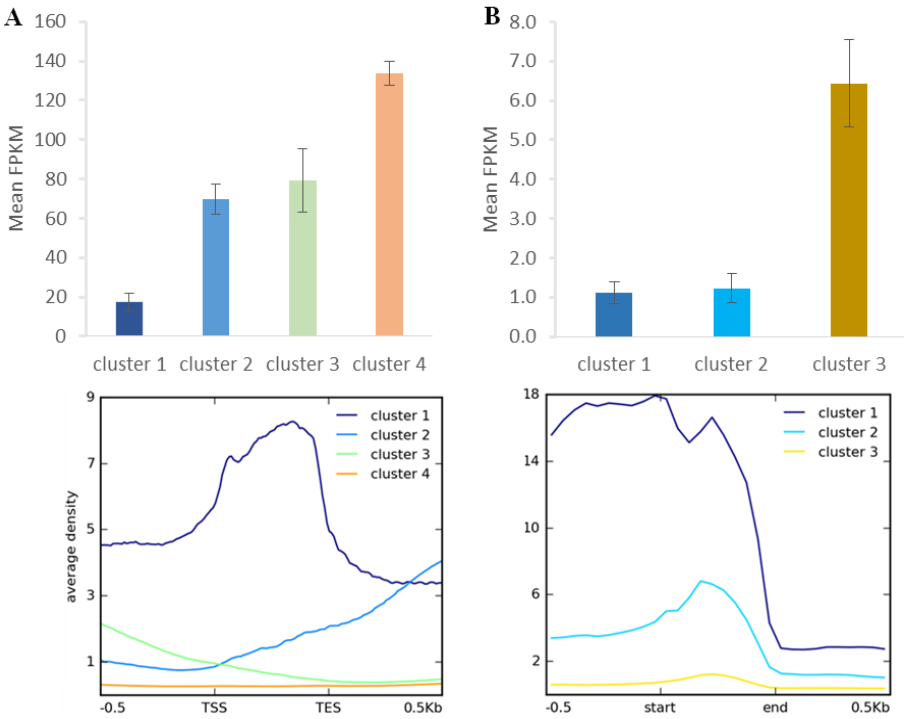



**Figure 8. Relationship between H3K27me3 occupancy and expression after clustering analysis.** Identified clusters (Figure 7) were correlated with expression level measured as FPKM in RNA-seq experiment. Expression level was determined for H3K27me3 occupancy clusters from genes (A) and repetitive elements (B). Error bars represent standard error.

## H3K27me3 location on chromosomes

Looking at the whole-chromosome level of H3K27me3 distribution, we found a preferential enrichment of the mark on the chromosome ends (Figure 9). Such observation concerns 36 out 40 ends from 20 chromosomes. The exceptions, in which the closest H3K27me3 domain was detected only 4 to 5 kb away from the chromosome end, include: 5' end of chromosome 3 and 3' end of chromosomes 7, 14, 16 (Table S2). Telomeric repeats in plants span regions from 0.5 kb in *Chlorella vulgaris* to 150kb in *Nicotiana tabacum* (Fajkus et al., 1995;

Higashiyama et al., 1995). *C. merolae* telomere length is relatively short and varies from 400bp to 700bp (Nozaki et al., 2007). Hence, the H3K27me3-domains at the chromosome ends detected in *C. merolae* cover the regions of telomeres and subtelomeres, similarly to what has been shown for the fungi: *Neurospora crassa* (Jamieson et al., 2013) and *Cryptococcus neoformans* (Dumesic et al., 2014).

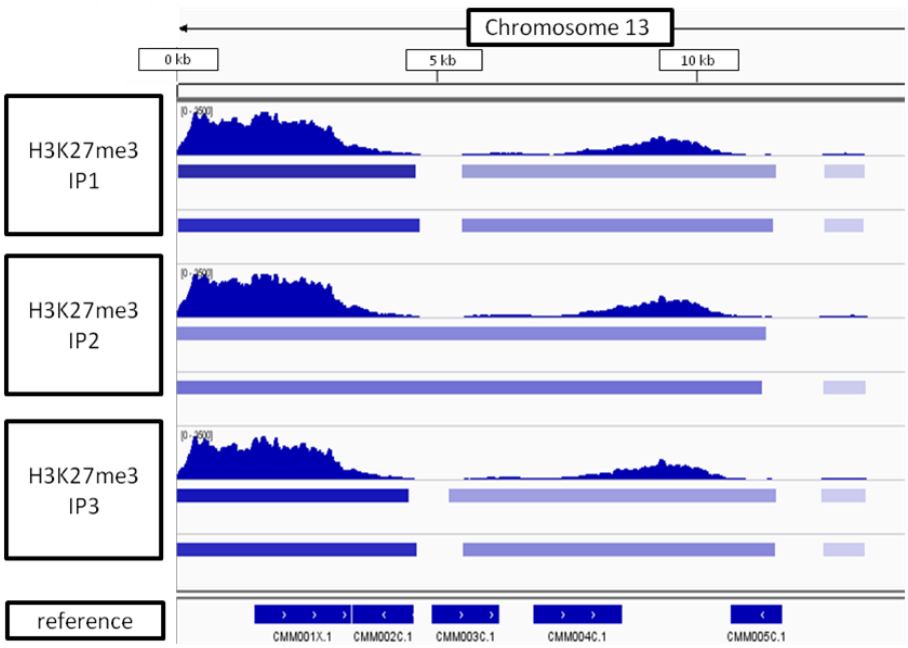



**Figure 9. H3K27me3 domain enrichment on chromosome ends.** 5' end of chromosome 13 is used as an example. Results from three biological replicates (IP1-3) taken for sequencing are shown. Bars below H3K27me3 tracks correspond to peaks scored by MACS2 with input (top bar) or H3 (bottom bar) used for normalization. The lowest track correspond to *C. merolae* genome reference. The data was visualized in IGV v2.3.

## H3K27me3-target gene ontology

In higher eukaryotes, H3K27me3 is known to target i.e. developmental, tissue-specific or stress-responsive genes (Boyer et al., 2006; Bracken et al., 2006; Kleinmanns and Schubert, 2014; Lafos et al., 2011; Zhang et al., 2007). We sought to investigate whether similar functional association can be found in *C. merolae*, a unicellular organism with primitive developmental program.

In order to decipher functional classification of H3K27me3 targets, we performed gene ontology analysis with Singular Enrichment Analysis from the AgriGO toolkit (Du et al., 2010). GO classes of H3K27me3-targets and full lists of *C. merolae* genes used as a

reference were extracted from the UniProtKB-database. We found a GO annotation for 3085 genes (50,5% from total number of 6108, including hypothetical proteins and transcripts) in the reference with 144 GO-annotated genes among H3K27me3-targets (62,6% from total number of 230). Statistical significance was determined based on the adjusted p-value with Yekiuteli test for FDR. For statistical tests, we selected only those GO classes that were represented by at least 5 entries (see material and methods). All three different GO sub-ontologies were taken into account: molecular function (GO:0003674), biological process (GO::0008150) and cellular component (GO::0008372).

We did not reveal any significantly (adjusted p-value < 0.05) enriched GO terms in cellular component sub-ontology. In contrast, we detected three GO term branches in biological process and one in molecular function sub-ontologies (Figure 10). Consistently with the known H3K27me3 function in the repression of developmental programs, we detected significantly enriched GO terms related to organismal process and development. Surprisingly, among enriched terms we found also those corresponding to protein maturation and intein-mediated protein splicing.

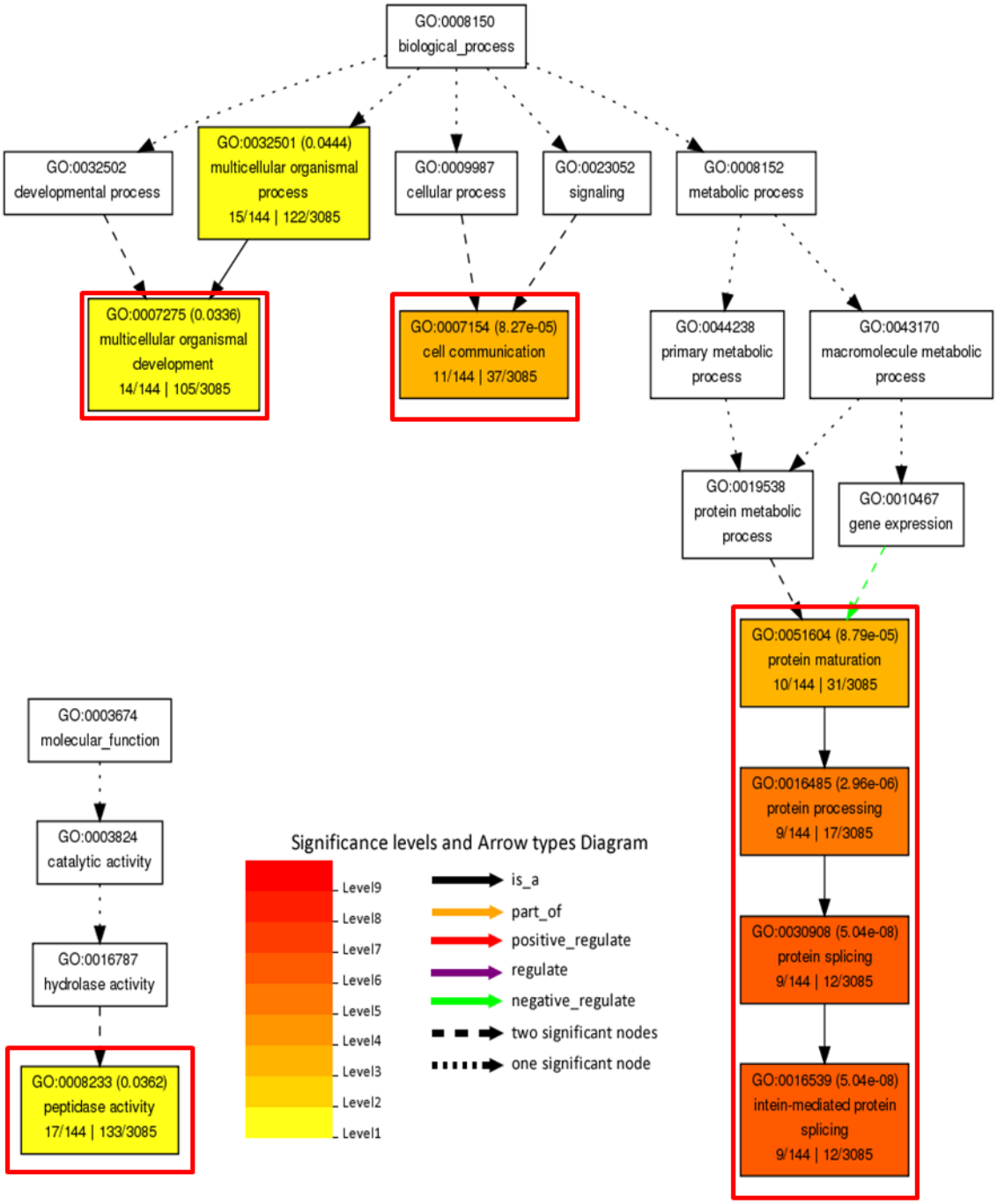



**Figure 10. Gene ontology analysis on H3K27me3 target genes.** Significantly enriched terms from molecular function and biological process subontologies (no significant hits found for cellular component subontology). Colour coding represents significance level (1-9) below adjusted p-value threshold ($p<0.05$). Adjusted p-values are shown next to GO IDs for each GO term. Relationships between GO terms are reflected in the arrow types. Red frames highlight GO terms connected with intein-mediated protein splicing. Numbers below subcategory name correspond to: GO subcategory gene number in query/total gene number in query | GO subcategory gene number in reference (TAIR10)/total gene number in reference.

Protein splicing is a protein maturation mechanism based on the excision of a protein fragment (intein) from a precursor and ligation of the flanking polypeptides (exteins).

Comparative analyses of amino acid sequences revealed a homology of protein splice-junction motif from inteins to the C-terminal Hog domains in Hedgehog proteins, the secretory proteins controlling developmental processes in Metazoa (Jiang and Hui, 2008).

In concordance to the protein splicing mechanism and the function of intein-containing genes, we observed that the majority of protein entries present in significantly enriched GO terms (organismal process and development, cell communication and peptidase function) overlap with the protein splicing group. Moreover, we found that all of the H3K27me3 targets present in the protein splicing term indeed possess a Hog domain as the Hedgehog proteins.

Interestingly, the genes encoding Hedgehog proteins in *C. merolae* reside in the telomeric and subtelomeric regions of the chromosomes (Nozaki et al., 2007), which is in agreement to our results on the preferential H3K27me3 enrichment at the chromosome ends (Figure 9). Moreover, comparing with clustering data, we observed that Hedgehog loci were enriched in cluster 3 (Table S3), bound by H3K27me3 preferentially to the upstream regions and moderately repressed.

### 3.3.4. Discussion

Polycomb group mediated regulation of gene expression in multicellular organisms is intensively studied in plants and has important roles in stress responses and developmental phase transitions (Kleinmanns and Schubert, 2014; Köhler and Villar, 2008). However, a model for Pc-G function in unicellular, photosynthetic organisms is currently lacking. We therefore performed homology searches and Bayesian phylogenetic tree analyses on the conserved alignment blocks. Our results confirmed a frequent absence of the PRC1 complex, ancient origin of PRC2 complex and the widespread distribution of core components of PRC2 in the green lineage (see also (Kim et al., 2013; Shaver et al., 2010)). We detected the PcG mark H3K27me3 in prasinophytes and red algae, arguing for the functionality of PRC2 in these organisms. Subsequent analyses of H3K27me3 occupancy in the model red alga, *C. merolae*, and correlative analyses with transcriptomic data revealed several important observations in this unicellular organism.

We observed that H3K27me3 in *C. merolae* is present on both, genes and repetitive elements, covering 4 % of total gene number and 44 % of total repetitive element number. The

occupancy on both genomic types is in the agreement with the results from *Drosophila* and *Phaeodactylum*, suggesting an ancestral function of PcG in repetitive element silencing. As *Arabidopsis* H3K27me3 occupies preferentially genes, our results suggest also a partial divergence from PcG-mediated repetitive element repression in higher plants.

In *Arabidopsis* endosperm, the vicinity of repetitive elements was shown to impact H3K27me3-mediated gene imprinting (Weinhofer et al., 2010; Wolff et al., 2011). In contrast, the H3K27me3-covered genes in *C. merolae* are not in the vicinity of repetitive elements. However, we cannot exclude the possibility that the genome annotation is incomplete, with unknown types of repetitive elements being not reported.

We showed thatH3K27me3 occupancy anti-correlates with the expression of its targets, consistently with its function in the multicellular organisms. As the other lysine-27 methylation marks, H3K27me1 and H3K27me2, have a repressive function in higher plants (Bastow et al., 2004; Jacob et al., 2009), similarly to the H3K27me3, the future steps should include profiling of those modifications in *C. merolae* genome. We also observed that the degree of repression is tightly correlated with the H3K27me3 profile at the genic level. The lowest expression was found for the genes covered by H3K27me3 over the gene body, whereas the targets with H3K27me3 occupancy on 5' and 3' flanking regions associate with intermediate repression.

Furthermore, H3K27me3 shows an enrichment on chromosome ends and covers a broad region including telomeric repeats and subtelomeric genes in *C. merolae*. Interestingly, several other species were reported to show an enrichment of H3K27me3 at the subtelomeric and telomeric regions (Smith et al., 2008; Vaquero-sedas et al., 2012). Moreover, Polycomb proteins were proved to be associated with telomere-binding factors (Zhou et al., 2013) and telobox motifs demonstrated to be enriched at PcG-peaks (Deng et al., 2013), arguing for a functional connection between PcG-mediated repression and chromosomal ends.

Our gene ontology analysis showed that H3K27me3 covers genes associated with intein-mediated protein splicing. Inteins are considered as ancient mobile elements present in all three domains of life (eukaryotes, archaea, bacteria) and viruses (Novikova et al., 2014; Pietrokovski, 2001). Hence, our results suggest that the H3K27me3-mediated repression might function in guarding the *C. merolae* genome. Such notion would be consistent with the predominant enrichment of H3K27me3 on the repetitive elements (Figure 5D). On the other hand, inteins may also develop to become post-translational regulatory elements in the course

of evolution, as seen for the conditional protein splicing of Hog domains in animals (Shah and Muir, 2014). Therefore, it is of special importance to decipher the function of intein-containing Hedgehog genes and the conditions for their de-repression in *C. merolae.*

Interestingly, we also observed that the intein-containing genes reside in the subtelomeric regions and are moderately repressed by H3K27me3 that occupies upstream regions from their TSS. It is currently unclear how protein splicing, moderate H3K27me3-mediated repression and subtelomeric localization are inter-connected. It is feasible that the integration of intein-containing genes appeared relatively recently and that these 'young' H3K27me3 targets display H3K27me3 enrichment at regions upstream of their TSS, in agreement with their moderate repression. Comparative genomics between *C. merolae* and related species should be used to validate this idea. Moreover, subtelomeres are dynamic regions and the objects of ectopic recombination events, which facilitate gene diversification, phenotypic diversity and adaptation to different environments, exemplified by an olfactory receptor gene family in humans (Linardopoulou et al., 2001), contingency systems in various pathogens (Barry et al., 2003) and carbon-source metabolism genes in *S. cerevisiae* (Brown et al., 2010; Carlson and Botstein, 1983; Louis et al., 1994). Hence, it is possible that the intein-containing subtelomeric H3K27me3-covered genes are the targets of rapid adaptive evolution. Characterization of their duplication frequency and diversification within their gene families should validate such notion.

Last but not the least, our results on H3K27me3 presence in various algae and genome-wide characterization of expression status and H3K27me3 occupancy in the model red alga *C. merolae* provide a resource for chromatin and transcriptomic studies. An existence of genetic tools, optimized growth conditions and the sequenced genome in *C. merolae* offer a possibility to decipher genome regulation in a broad evolutionary context, including the organisms with super-reduced genomes.

### 3.3.5. Materials and Methods:

#### Phylogenetic analysis and homology search

Sequences of PRC1 and PRC2 components from *Arabidopsis thaliana* and *Drosophila melanogaster* were selected from TAIR10 and NCBI databases, respectively, and used as the queries. Homology searches were done using BLAST against organism-specific databases at: https://phytozome.jgi.doe.gov/pz/portal.html (*Chlamydomonas reinhardtii* v5.5 and *Micromonas pisulla* CCMP1545 v3.0), http://merolae.biol.s.u-tokyo.ac.jp/ (*Cyanidioschyzon merolae*), http://genome.jgi.doe.gov/Ostta4/Ostta4.home.html (*Ostreococcus tauri* v2.0), http://cyanophora.rutgers.edu/cyanophora/home.php (*Cyanophora paradoxa*).

The highest scoring candidate proteins were used for BLAST searches against the TAIR10 genome annotation to confirm a reciprocal match to the protein used as an initial query. The sequences were aligned using ClustalXv2.0 (Larkin et al., 2007) and conserved blocks were selected by gBlocks v0.91b (Talavera and Castresana, 2007). For substitution model estimation, ProtTestv3.4 was employed (Abascal et al., 2005). Phylogenetic trees were created using a Bayesian approach in Mr Bayesv3.2 (Ronquist et al., 2012), self-compiled developer version r1067 with implemented LG model. The runs were done with 20 mln generations. Sequences from *D. melanogaster* were selected as an out group. Consensus trees and alignments were visualized with Figtree v1.4.2 (Rambaut, 2009)and Jalviewv2 (Waterhouse et al., 2009), respectively.

#### Algae growth

*Cyanidioschyzon merolae* cells were grown at 42°C with shaking at 150 rpm. The cultures were kept under continuous light, with light intensity of 70 µmol.

#### Western blot

Around 50.000 algae cells were diluted in Laemmli's sample buffer, and 3-7 µl of total protein/lane was subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) using 12% SDS-polyacrylamide gel. The resolved proteins were transferred

onto polyvinylidenedifluoride (PVDF) membranes (Immobilon-P membrane, Milipore) using a Vertical Electrophoresis Cell (Bio-Rad). Following transfer, membranes were washed with phosphate buffered saline containing 0.05% Tween-20 (PBST) and then blocked in 5%milk diluted in PBST for 1 h at room temperature. The membranes were then incubated overnight with anti_H3K27me3 (N 07-449, Millipore or C15410195, Diagenode) and anti H3 (ab1791-100, Abcam) rabbit polyclonal antibodies 1:2000 in PBST containing 5% milk at 4°C. After washing three times for 10 min in PBST, membranes were incubated with peroxidase-labeled secondary antibody for 1 h at room temperature. The membranes were washed three times for 10 min in PBST, incubated with SuperSignal West Femto Chemiluminescence Substrate (Thermo Scientific) for 1 min, and exposed for 1-2 min. Commercial histone extract from calf thymus (Sigma, #9064-47-5) was used as a positive control for H3K27me3 detection.

**Competition assay**

Pre-incubation of anti-H3K27me3 antibody (C15410195, Diagenode) and H3K27me3 peptide (Intavis) was done for two hours in RT in PBST, with occasional mixing. The peptide-to-antibody molar ratio was 50:1. A control solution - PBST with just antibody at 2X final concentration - was used at the same time. At the end of the pre-incubation 4% BSA blocking solution was added to the peptide/antibody mixture until a final concentration of 3% BSA, mixed briefly and added to the membranes. After washing three times for 10 min in PBST, membranes were incubated with peroxidase-labeled secondary antibody for 1 h at room temperature. The membranes were washed three times for 10 min in PBST, incubated with SuperSignal West Femto Chemiluminescence Substrate (Thermo Scientific) for 1 min, and exposed for 1-2 min.

**Chromatin immunoprecipitation**

*Cyanidioschyzon merolae* cells were grown until late log-phase. 40ml samples were fixed (Formaldehyde, 1% (v/v)) for 10 min, until addition of glycine (to final concentration of 125mM, 5 min incubation). Superfluous formaldehyde was removed by three washes with ice cold PBS buffer and the remaining cell pellet was resuspended in 4 ml of Extraction Buffer (50mM Tris-HCl pH 8, 10mM EDTA, 1% SDS) with Protease Inhibitor Cocktail (Roche). The samples were sonicated for 5 min with 30 sec ON/ 30 sec OFF cycle using Bioruptor

Plus device (Diagenode) and cleared by two rounds of centrifugation (13000rpm, 4°C, 10 min). Subsequent steps were performed as in the Plant ChIP-seq kit protocol (Diagenode) with higher volume of sample taken aside as an input (1:5 of chromatin for IP). Immunoprecipitation was done using anti-H3K27me3 Polyclonal Premium antibody (C15410195, Diagenode) and, as a negative control, IgG fraction from rabbit (C15410206, Diagenode). Quality and fragment size of immunoprecipitated DNA and input samples were measured using agarose gel electrophoresis and Bioanalyzer 2100 (Agilent Technologies).

**Real time PCR**

DNA samples obtained from ChIP were used for H3K27me3 enrichment analysis for several target genes by real-time quantitative PCR. Reactions were prepared using KAPA SYBR FAST qPCR Mastermix (KapaBiosystems), according to the manufacturer's protocol, and run on iQ5 detection system (Biorad) using a 2-step programme. Differences in H3K27me3 enrichment on target genes were scored by comparison of % recovery of input and standard error values. 5'->3' sequences of oligonucleotides used for amplification were as follows: CmMADS- forward: GGATGAGAAAGCGAGAAATACGA and reverse: TCACAATGCCGATCTCACAG; CmEIF-4A – forward: TGTACGATATGATCCAGAGAAGAG and reverse: TGTAGATTTGCTCCTTGAAACC; Cm60S – forward: AAGTTTCGCTGTACGCTTGG and reverse: TAACCAGGACCATATCGCCG.

**ChIP-seq**

DNA library was prepared with MicroPlex Library Preparation Kit (Diagenode) and sequenced on HiSeq 2000 sequencer (Illumina Inc.). Quality of paired-end raw reads was assessed using FastQC v0.11.4. The reads were trimmed according to the quality and mapped to the reference in Bowtie v2.2.6, with standard options. Peaks were called using MACS v2.1 with exclusion of clonal reads. The new annotation from the *Cyanidioschyzon merolae* Genome Project v3 (czon.jp/download/annot_list.txt) was employed as a reference. The correlation between ChIP-seq replicates was scored using Pearson correlation and shown in Figure S3A.The sequences were deposited in the Gene Expression Omnibus under the series GSE93913.

**RNA extraction and RNA-seq**

*Cyanidioschyzon merolae* cells were grown until late log-phase. RNA was isolated using RNeasy Plant Mini Kit (Qiagen), following a standard protocol. cDNA synthesis and DNase treatment were performed using RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) with oligo(dT) primers, according to a standard manual. The quality and concentration of samples were measured spectrometrically using Nanodrop 1000 (Thermo Scientific) and electrophoretically using Bioanalyzer 2100 (Agilent Technologies). Libraries were constructed using TruSeq RNA Library Prep kit (Illumina) with gel-free library purification based on Agencourt AMPure XP beads (Beckman Coulter, Inc.). The samples were sequenced on HiSeq 2000 device (Illumina Inc.). Paired-end raw sequencing reads were analyzed using Galaxy implementations (usegalaxy.org) of FastQC programme and Tuxedo protocol (Trapnell et al., 2012). The new annotation from *Cyanidioschyzon merolae* Genome Project v3 (czon.jp/download/annot_list.txt) was employed as a reference. The correlation between replicates in RNA-seq experiment was scored using Pearson correlation and shown in Figure S3B. The sequences were deposited in the Gene Expression Omnibus, under the series GSE93912.

**Bioinformatic secondary analysis of NGS data**

Identified H3K27me3 peaks were annotated by intersection with the reference using Bedtools v2.17 (Quinlan and Hall, 2010). Sequences from the unannotated peaks were translated *in silico* to ORF in HMMER2GO v0.17 software and the longest ORF was used for homology searches by BLASTP/Pfam against respective protein databases. Alternatively, unannotated peak sequences were used for homology searches by BLASTN against NCBI nucleotide databases: nr/nt and ref_seq. Homology searches were performed in BLAST+ stand alone package v2.2.28+ (Camacho et al., 2009). Alignment threshold was set as follows: > 50% alignment length and < 0.5 E-value. Complete genome records in nucleotide databases were used to form negative GI list and excluded from the BLASTN searches. Only the top alignment hit was included in the further analyses. Distance between peak and annotated genomic feature was obtained using ‘closest’ command from Bedtoolsv2.17. Mapping of the reads was visualized in IGV v2.3 (Thorvaldsdóttir et al., 2013). Clustering of H3K27me3 enrichment was done on deepTools2 (Ramirez et al., 2016). Gene ontology was inferred by using Singular Enrichment Analysis on the AgriGO server (Du et al., 2010) against complete

GO list. Protein annotations were extracted from UniProtKB database and used as reference. To select significantly enriched GO terms, Fisher test with Yekuteli adjustment was used a statistical method. P-value was set to 0.05 and minimal number of entries kept at 5.

### 3.3.6. Acknowledgements

We thank Prof. Motomichi Matsuzaki for providing *C. merolae* genome annotation and Dr. Michael Grünstäudl for crucial insights in phylogenetic tree creation. We thank Julia Kleinmanns and Dorota Komar for critical revision of the manuscript. PM, OK and DS were supported by the DFG-funded SFB973, the Marie Curie International Training Network "EpiTRAITS" and by Freie Universität Berlin within the Excellence Initiative of the German Research Foundation.

### 3.3.7. References

Abascal, F., Zardoya, R., Posada, D., Miller, C. J., Lou, H. J., Johnson, A. D., et al. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21, 2104–2105. doi:10.1093/bioinformatics/bti263.

Barry, J. D., Ginger, M. L., Burton, P., and McCulloch, R. (2003). Why are parasite contingency genes often associated with telomeres? *Int. J. Parasitol.* 33, 29–45. doi:10.1016/S0020-7519(02)00247-3.

Bastow, R., Mylne, J. S., Lister, C., Lippman, Z., Martienssen, R. a, and Dean, C. (2004). Vernalization requires epigenetic silencing of FLC by histone methylation. *Nature* 427, 164–167. doi:10.1038/nature02269.

Berke, L., and Snel, B. (2015). The plant Polycomb repressive complex 1 ( PRC1 ) existed in the ancestor of seed plants and has a complex duplication history. 1, 1–10. doi:10.1186/s12862-015-0319-z.

Bowman, J. L., Floyd, S. K., and Sakakibara, K. (2007). Green genes-comparative genomics of the green branch of life. *Cell* 129, 229–34. doi:10.1016/j.cell.2007.04.004.

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441, 349–353. doi:10.1038/nature04733.

Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. 1123–1136. doi:10.1101/gad.381706.present.

Brown, C. A., Murray, A. W., and Verstrepen, K. J. (2010). Rapid Expansion and Functional Divergence of Subtelomeric Gene Families in Yeasts. *Curr. Biol.* 20, 895–903. doi:10.1016/j.cub.2010.04.027.

Butenko, Y., and Ohad, N. (2011). Polycomb-group mediated epigenetic mechanisms through plant evolution. *Biochim. Biophys. Acta* 1809, 395–406. doi:10.1016/j.bbagrm.2011.05.013.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 1–9. doi:10.1186/1471-2105-10-421.

Carlson, M., and Botstein, D. (1983). Organization of the *SUC* gene family in *Saccharomyces*. *Mol. Cell. Biol.*

3, 351–359.

Cerase, A., Smeets, D., Tang, Y. A., Gdula, M., Kraus, F., Spivakov, M., et al. (2014). Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy. *Proc. Natl. Acad. Sci. U. S. A.* 111, 2235–2240. doi:10.1073/pnas.1312951111.

Cui, J., Zhang, Z., Shao, Y., Zhang, K., Leng, P., and Liang, Z. (2015). Genome-wide identification, evolutionary, and expression analyses of histone H3 variants in plants. *Biomed Res. Int.* 2015. doi:10.1155/2015/341598.

Deleris, A., Stroud, H., Bernatavichute, Y., Johnson, E., Klein, G., Schubert, D., et al. (2012). Loss of the DNA methyltransferase MET1 Induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in Arabidopsis thaliana. *PLoS Genet.* 8, e1003062. doi:10.1371/journal.pgen.1003062.

Deng, W., Buzas, D. M., Ying, H., Robertson, M., Taylor, J., Peacock, W. J., et al. (2013). Arabidopsis Polycomb Repressive Complex 2 binding sites contain putative GAGA factor binding motifs within coding regions of genes. *BMC Genomics* 14, 593. doi:10.1186/1471-2164-14-593.

Derkacheva, M., and Hennig, L. (2014). Variations on a theme: Polycomb group proteins in plants. *J. Exp. Bot.* 65, 2769–2784. doi:10.1093/jxb/ert410.

Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: A GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* 38. doi:10.1093/nar/gkq310.

Dumesic, P. A., Homer, C. M., Moresco, J. J., Pack, L. R., Shanle, E. K., Coyle, S. M., et al. (2014). Product Binding Enforces the Genomic Specificity of a Yeast Polycomb Repressive Complex. *Cell*, 1–15. doi:10.1016/j.cell.2014.11.039.

Ebert, A., Schotta, G., Lein, S., Kubicek, S., Krauss, V., Jenuwein, T., et al. (2004). Su(var) genes regulate the balance between euchromatin and heterochromatin in Drosophila. *Genes Dev.* 18, 2973–2983. doi:10.1101/gad.323004.

Fajkus, J., Kovařík, A., mKrálovics, R., and Bezděk, M. (1995). Organization of telomeric and subtelomeric chromatin in the higher plant Nicotiana tabacum. *MGG Mol. Gen. Genet.* 247, 633–638. doi:10.1007/BF00290355.

Ferrari, K. J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., et al. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol. Cell* 53, 49–62. doi:10.1016/j.molcel.2013.10.030.

Hennig, L., and Derkacheva, M. (2009). Diversity of Polycomb group complexes in plants: same rules, different players? *Trends Genet.* 25, 414–23. doi:10.1016/j.tig.2009.07.002.

Higashiyama, T., Maki, S., and Yamada, T. (1995). Molecular organization of Chlorella vulgaris chromosome I: Presence of telomeric repeats that are conserved in higher plants. *Mol. Gen. Genet.* 246, 29–36. doi:10.1007/BF00290130.

Hohenstatt, M. L. (2012). Functional analysis of SCI1 – A PWWP domain protein involved in Polycomb group mediated gene regulation in Arabidopsis.

Huang, Y., Chen, D.-H., Liu, B.-Y., Shen, W.-H., and Ruan, Y. (2016). Conservation and diversification of polycomb repressive complex 2 (PRC2) proteins in the green lineage. *Brief. Funct. Genomics*, elw007–. doi:10.1093/bfgp/elw007.

Jacob, Y., Feng, S., LeBlanc, C. A., Bernatavichute, Y. V, Stroud, H., Cokus, S., et al. (2009). ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. *Nat. Struct. Mol. Biol.* 16, 763–8. doi:10.1038/nsmb.1611.

Jacob, Y., and Michaels, S. D. (2009). H3K27me1 is E(z) in animals, but not in plants. *Epigenetics* 4, 366–369. doi:10.4161/epi.4.6.9713.

Jamieson, K., Rountree, M. R., Lewis, Z. a, Stajich, J. E., and Selker, E. U. (2013). Regional control of histone H3 lysine 27 methylation in Neurospora. *Proc. Natl. Acad. Sci. U. S. A.* 110, 6027–32. doi:10.1073/pnas.1303750110.

Jeffery, C. J. (2003). Moonlighting proteins: Old proteins learning new tricks. *Trends Genet.* 19, 415–417. doi:10.1016/S0168-9525(03)00167-7.

Jiang, J., and Hui, C.-C. (2008). Hedgehog signaling in development and cancer. *Dev. Cell* 15, 801–812. doi:10.1016/j.devcel.2008.11.010.

Kim, E., Ma, X., and Cerutti, H. (2013). Bioresource Technology Gene silencing in microalgae : Mechanisms and biological roles. *Bioresour. Technol.* 184, 23–32. doi:10.1016/j.biortech.2014.10.119.

Kleinmanns, J. A., and Schubert, D. (2014). Polycomb and Trithorax group protein-mediated control of stress responses in plants. *Biol. Chem.* 395, 1291–300. doi:10.1515/hsz-2014-0197.

Köhler, C., and Villar, C. B. R. (2008). Programming of gene expression by Polycomb group proteins. *Trends Cell Biol.* 18, 236–243. doi:10.1016/j.tcb.2008.02.005.

Lachner, M., Sengupta, R., Schotta, G., and Jenuwein, T. (2004). Trilogies of histone lysine methylation as epigenetic landmarks of the eukaryotic genome. in *Cold Spring Harbor Symposia on Quantitative Biology*, 209–218. doi:10.1101/sqb.2004.69.209.

Lafos, M., Kroll, P., Hohenstatt, M. L., Thorpe, F. L., Clarenz, O., and Schubert, D. (2011). Dynamic regulation of H3K27 trimethylation during arabidopsis differentiation. *PLoS Genet.* 7. doi:10.1371/journal.pgen.1002040.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi:10.1093/bioinformatics/btm404.

Linardopoulou, E., Mefford, H. C., Nguyen, O., Friedman, C., van den Engh, G., Farwell, D. G., et al. (2001). Transcriptional activity of multiple copies of a subtelomerically located olfactory receptor gene that is polymorphic in number and location. *Hum. Mol. Genet.* 10, 2373–2383. doi:10.1093/hmg/10.21.2373.

Lindroth, A. M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., et al. (2004). Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. *EMBO J.* 23, 4286–4296. doi:10.1038/sj.emboj.7600430.

Liu, Y., Taverna, S. D., Muratore, T. L., Shabanowitz, J., Hunt, D. F., and Allis, C. D. (2007). RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. 1530–1545. doi:10.1101/gad.1544207.structures.

Louis, E. J., Naumova, E. S., Lee, A., Naumov, G., and Haber, J. E. (1994). The chromosome end in yeast: Its mosaic nature and influence on recombinational dynamics. *Genetics* 136, 789–802.

Matsuzaki, M., Misumi, O., Shin-I, T., Maruyama, S., Takahara, M., Miyagishima, S.-Y., et al. (2004). Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. *Nature* 428, 653–657. doi:10.1038/nature02398.

McKay, D. J., Klusza, S., Penke, T. J. R., Meers, M. P., Curry, K. P., McDaniel, S. L., et al. (2015). Interrogating the function of metazoan histones using engineered gene clusters. *Dev. Cell* 32, 373–386. doi:10.1016/j.devcel.2014.12.025.

Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, J., Witman, G. B., et al. (2010). The Chlamydomonas Genome Reveals the Evolution of Key Animal and Plant Functions. *Science (80-. ).* 318, 245–250. doi:10.1126/science.1143609.The.

Müller, J., Hart, C. M., Francis, N. J., Vargas, M. L., Sengupta, A., Wild, B., et al. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* 111, 197–208. doi:10.1016/S0092-8674(02)00976-5.

Müller, J., and Verrijzer, P. (2009). Biochemical mechanisms of gene regulation by polycomb group protein complexes. *Curr. Opin. Genet. Dev.* 19, 150–8. doi:10.1016/j.gde.2009.03.001.

Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Köcher, T., et al. (2007). Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. *EMBO J.* 26, 4078–4088. doi:10.1038/sj.emboj.7601837.

Novikova, O., Topilina, N., and Belfort, M. (2014). Enigmatic distribution, evolution, and function of inteins. *J. Biol. Chem.* 289, 14490–7. doi:10.1074/jbc.R114.548255.

Nozaki, H., Takano, H., Misumi, O., Terasawa, K., Matsuzaki, M., Maruyama, S., et al. (2007). A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga Cyanidioschyzon merolae. *BMC Biol.* 5, 28. doi:10.1186/1741-7007-5-28.

Oh, S., Park, S., and van Nocker, S. (2008). Genic and global functions for Paf1C in chromatin modification and gene expression in Arabidopsis. *PLoS Genet.* 4, e1000077. doi:10.1371/journal.pgen.1000077.

Okada, T., Endo, M., Singh, M. B., and Bhalla, P. L. (2005). Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant AtMGH3. *Plant J.* 44, 557–68. doi:10.1111/j.1365-313X.2005.02554.x.

Palenik, B., Grimwood, J., Aerts, A., Rouzé, P., Salamov, A., Putnam, N., et al. (2007). The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. *Proc. Natl. Acad. Sci. U. S. A.* 104, 7705–10. doi:10.1073/pnas.0611046104.

Papp, B., and Müller, J. (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev.* 20, 2041–2054. doi:10.1101/gad.388706.

Park, S., Oh, S., and van Nocker, S. (2012). Genomic and gene-level distribution of histone H3 dimethyl lysine-27 (H3K27me2) in Arabidopsis. *PLoS One* 7, e52855. doi:10.1371/journal.pone.0052855.

Pauler, F. M., Sloane, M. A., Huang, R., Regha, K., Koerner, M. V, Tamir, I., et al. (2009). H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. 221–233. doi:10.1101/gr.080861.108.4.

Pietrokovski, S. (2001). Intein spread and extinction in evolution. *Trends Genet.* 17, 465–472. doi:10.1016/S0168-9525(01)02365-4.

Del Prete, S., Mikulski, P., Schubert, D., and Gaudin, V. (2015). One, two, three: Polycomb proteins hit all dimensions of gene regulation. *Genes (Basel).* 6, 520–542. doi:10.3390/genes6030520.

Price, D. C., Chan, C. X., Yoon, H. S., Yang, E. C., Qiu, H., Weber, A. P., et al. (2012). Cyanophora paradoxa genome elucidates origin of photosynthesis in algae and plants. *Science (80-. ).* 335, 843–847.

doi:10.1126/science.1213561.

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi:10.1093/bioinformatics/btq033.

Rambaut, A. (2009). FigTree, a graphical viewer of phylogenetic trees. *Inst. Evol. Biol. Univ. Edinburgh*.

Ramirez, F., Ryan, D. P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., et al. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44, 160–165. doi:10.1093/nar/gkw257.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi:10.1093/sysbio/sys029.

Sakamoto, Y., and Takagi, S. (2013). LITTLE NUCLEI 1 and 4 regulate nuclear morphology in arabidopsis thaliana. *Plant Cell Physiol.* 54, 622–633. doi:10.1093/pcp/pct031.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome Regulation by Polycomb and Trithorax Proteins. *Cell* 128, 735–745. doi:10.1016/j.cell.2007.02.009.

Schwartz, Y. B., and Pirrotta, V. (2008). Polycomb complexes and epigenetic states. *Curr. Opin. Cell Biol.* 20, 266–273. doi:10.1016/j.ceb.2008.03.002.

Schwartz, Y. B., and Pirrotta, V. (2013). A new world of Polycombs: unexpected partnerships and emerging functions. *Nat. Rev. Genet.* 14, 853–64. doi:10.1038/nrg3603.

Shah, N. H., and Muir, T. W. (2014). Inteins: nature's gift to protein chemists. *Chem. Sci.* 5, 446–461. doi:Doi 10.1039/C3sc52951g.

Shaver, S., Casas-Mollano, J. A., Cerny, R. L., and Cerutti, H. (2010). Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga Chlamydomonas. *Epigenetics* 5, 301–312. doi:10.4161/epi.5.4.11608.

Smith, K. M., Kothe, G. O., Matsen, C. B., Khlafallah, T. K., Adhvaryu, K. K., Hemphill, M., et al. (2008). The fungus Neurospora crassa displays telomeric silencing mediated by multiple sirtuins and by methylation of histone H3 lysine 9. *Epigenetics Chromatin* 1, 5. doi:10.1186/1756-8935-1-5.

Sung, S., Schmitz, R. J., and Amasino, R. M. (2006). A PHD finger protein involved in both the vernalization and photoperiod pathways in Arabidopsis. *Genes Dev.* 20, 3244–3248. doi:10.1101/gad.1493306.

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564–77. doi:10.1080/10635150701472164.

Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192. doi:10.1093/bib/bbs017.

Topilina, N. I., and Mills, K. V (2014). Recent advances in in vivo applications of intein-mediated protein splicing. *Mob. DNA* 5, 5. doi:10.1186/1759-8753-5-5.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–78. doi:10.1038/nprot.2012.016.

Vaquero-sedas, I., Luo, C., and Vega-palas, M. A. (2012). Analysis of the epigenetic status of telomeres by

using ChIP-seq data. 1–6. doi:10.1093/nar/gks730.

Veerappan, C. S., Avramova, Z., and Moriyama, E. N. (2008). Evolution of SET-domain protein families in the unicellular and multicellular Ascomycota fungi. *BMC Evol. Biol.* 8, 190. doi:10.1186/1471-2148-8-190.

Veluchamy, A., Rastogi, A., Lin, X., Lombard, B., Murik, O., Thomas, Y., et al. (2015). An integrative analysis of post-translational histone modifications in the marine diatom Phaeodactylum tricornutum. *Genome Biol.* 16, 102. doi:10.1186/s13059-015-0671-8.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi:10.1093/bioinformatics/btp033.

Weinhofer, I., Hehenberger, E., Roszak, P., Hennig, L., and Köhler, C. (2010). H3K27me3 profiling of the endosperm implies exclusion of polycomb group protein targeting by DNA methylation. *PLoS Genet.* 6, 1–14. doi:10.1371/journal.pgen.1001152.

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M. T. A., et al. (2011). High-Resolution analysis of parent-of-origin allelic expression in the Arabidopsis endosperm. *PLoS Genet.* 7. doi:10.1371/journal.pgen.1002126.

Worden, A. Z., Lee, J.-H., Mock, T., Rouzé, P., Simmons, M. P., Aerts, A. L., et al. (2009). Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes Micromonas. *Science* 324, 268–272. doi:10.1126/science.1167222.

Yang, C., Bratzel, F., Hohmann, N., Koch, M., Turck, F., and Calonje, M. (2013). VAL-and AtBMI1-Mediated H2Aub initiate the switch from embryonic to postgerminative growth in arabidopsis. *Curr. Biol.* 23, 1324–1329. doi:10.1016/j.cub.2013.05.050.

Yang, H., Howard, M., and Dean, C. (2014). Antagonistic roles for H3K36me3 and H3K27me3 in the cold-induced epigenetic switch at Arabidopsis FLC. *Curr. Biol.* 24, 1793–7. doi:10.1016/j.cub.2014.06.047.

Yin, H., Sweeney, S., Raha, D., Snyder, M., and Lin, H. (2011). A high-resolution whole-genome map of key chromatin modifications in the adult Drosophila melanogaster. *PLoS Genet.* 7, e1002380. doi:10.1371/journal.pgen.1002380.

Yong, D. C., Ying, H., and Shen, R. W. (2016). The evolutionary landscape of PRC1 core components in green lineage. *Planta* 243, 825–846. doi:10.1007/s00425-015-2451-9.

Young, M. D., Willson, T. A., Wakefield, M. J., Trounson, E., Hilton, D. J., Blewitt, M. E., et al. (2011). ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. 7415–7427. doi:10.1093/nar/gkr416.

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y. V, Pellegrini, M., Goodrich, J., et al. (2007). Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. *PLoS Biol.* 5, e129. doi:10.1371/journal.pbio.0050129.

Zhou, Y., Hartwig, B., James, G. V., Schneeberger, K., and Turck, F. (2013). Complementary Activities of TELOMERE REPEAT BINDING Proteins and Polycomb Group Complexes in Transcriptional Regulation of Target Genes. 87–101. doi:10.1105/tpc.15.00787.

### 3.3.8. Supplementary data

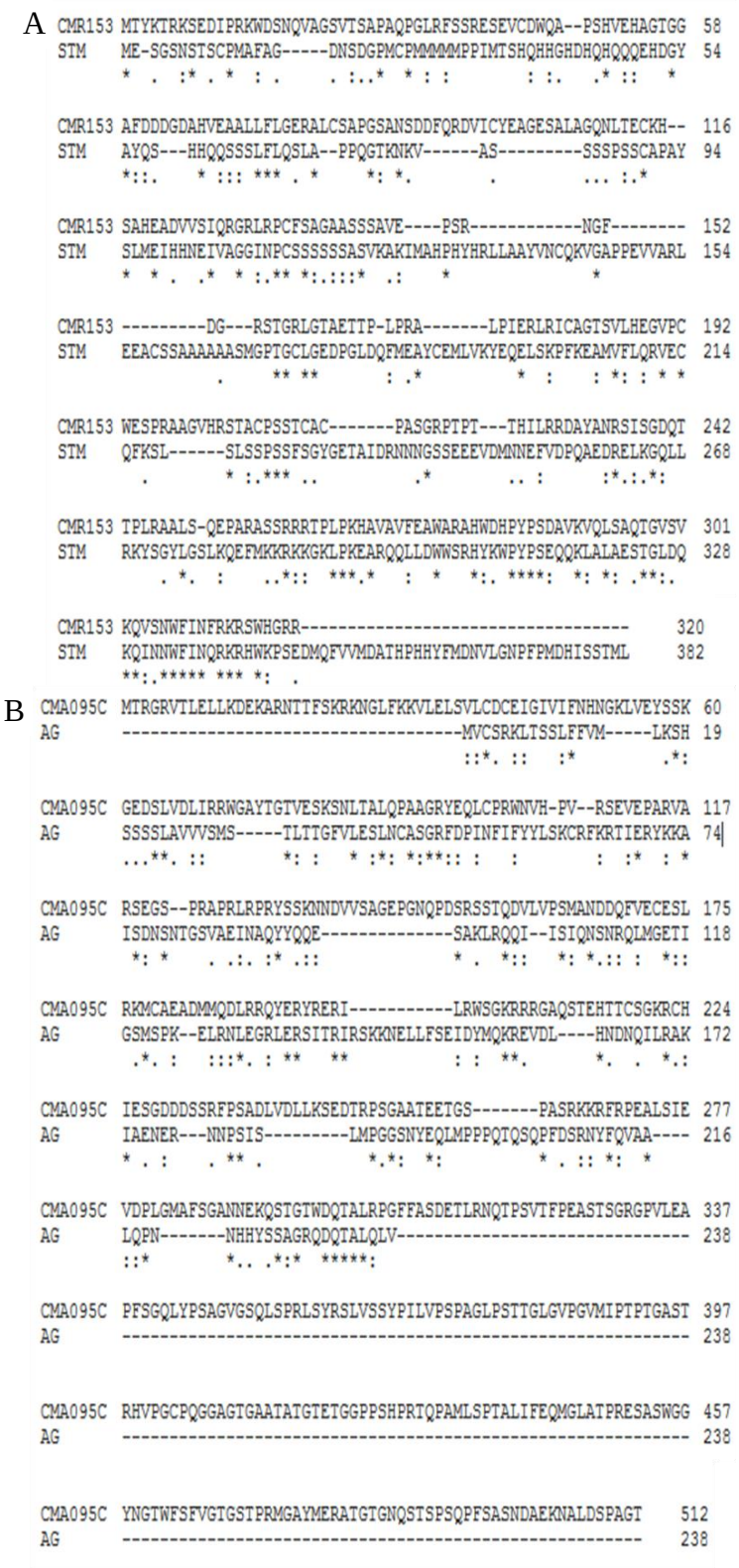



**Figure S1. Alignment of selected H3K27me3 targets from *A. thaliana*:** (A) SHOOT MERISTEMLESS (STM) and (B) AGAMOUS (AG),and their homologs from *C. merolae*.

**Table S1. Distance between H3K27me3 peaks and gene locations**

| | IP1 | | IP2 | | IP3 | |
|---|---|---|---|---|---|---|
| | ann. peak | unann. peak | ann. peak | unann. peak | ann. peak | unann. peak |
| distance to annotated feature (mean) | 542.9 | 543.8 | 539.2 | 520.5 | 504.1 | 514.0 |
| distance to annotated feature (median) | 516.0 | 498.0 | 506.0 | 461.5 | 486.0 | 465.0 |

Distance was obtained using 'closest' command from Bedtools v2.17 for 3 biological replicates (IP1-3) from chromatin immunoprecipitation experiment. Mean and median were calculated from distance values in base pairs for each peak. Peaks were divided into annotated ('ann.peak') and unannotated ('unannot.peak') subgroups. See main text for the details.

**Table S2. Distance of H3K27me3 domain to the chromosome ends.**

| | distance [bp] | | | distance [bp] | |
|---|---|---|---|---|---|
| chromosome | to 5' end | to 3' end | chromosome | to 5' end | to 3' end |
| 1 | 777 | 1321 | 11 | 144 | 44 |
| 2 | 19 | 247 | 12 | 126 | 28 |
| 3 | 3977 | 46 | 13 | 33 | 23 |
| 4 | 21 | 773 | 14 | 28 | 4734 |
| 5 | 41 | 453 | 15 | 18 | 18 |
| 6 | 216 | 24 | 16 | 27 | 3955 |
| 7 | 157 | 4756 | 17 | 74 | 36 |
| 8 | 149 | 31 | 18 | 236 | 22 |
| 9 | 22 | 28 | 19 | 97 | 987 |
| 10 | 1601 | 29 | 20 | 99 | 90 |

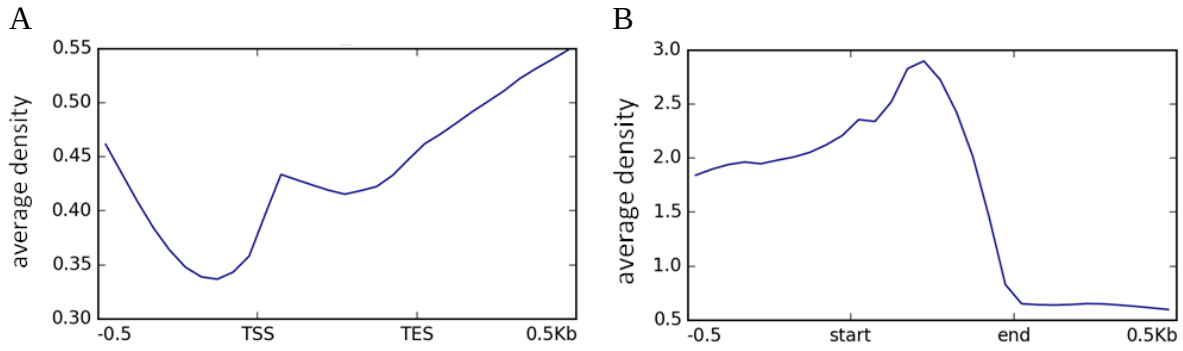



**Figure S2. Average H3K27me3 occupancy over genes (A) and repetitive elements (B).** The H3K27me3 reads were scaled to 500bp windows and flanking regions set to 0.5kb. DeepTools2 package was used to computed an abundance of the reads and plot creation.

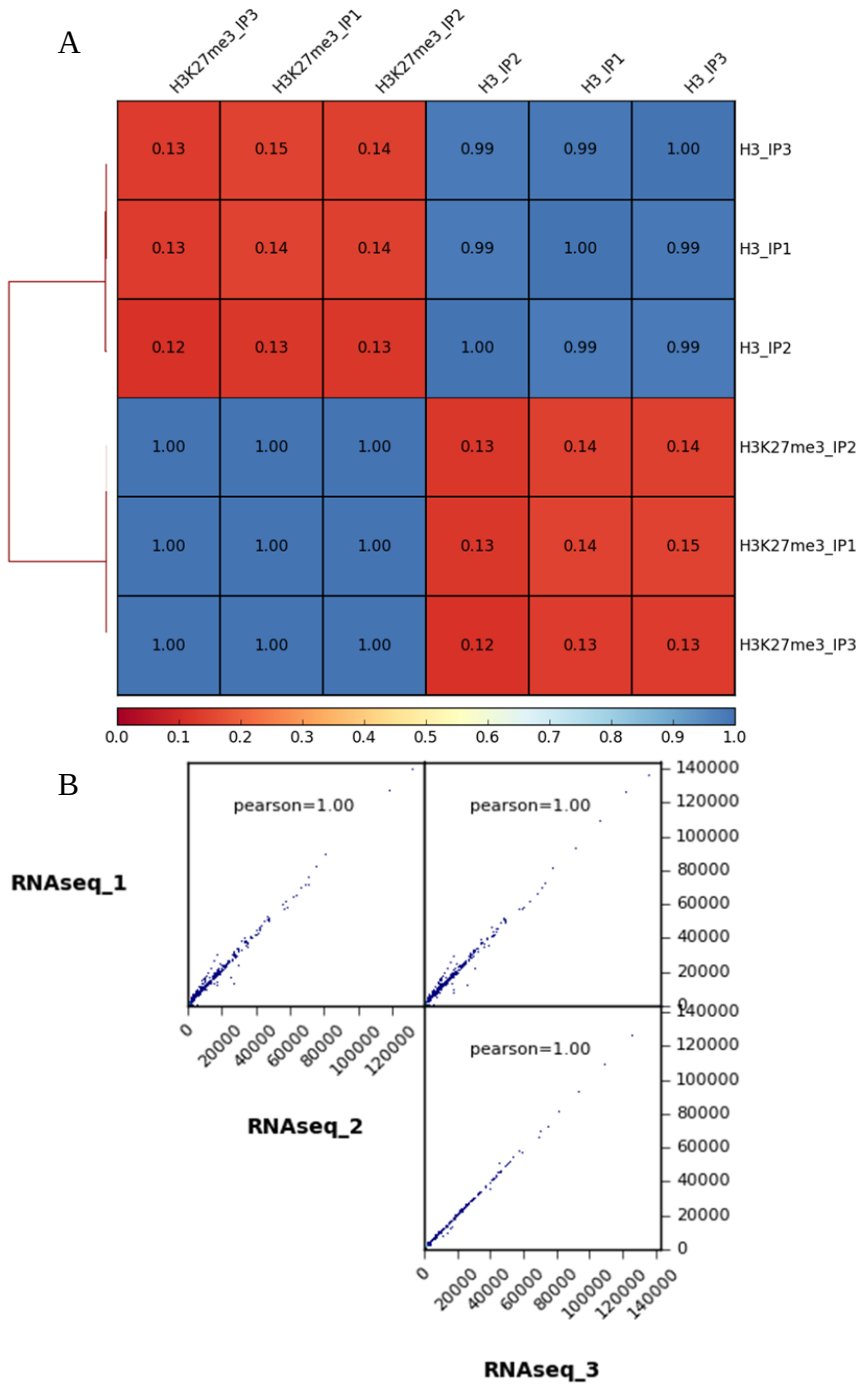



**Figure S3. Correlation between replicates in NGS experiments.** (A) Correlation for ChIP-seq experiment (replicates IP1-3). Values inside cells represent correlation coefficients calculated using Pearson method. Matrix for heatmap was computed in deepTools2 with bin size set to 1kb. Bins with zero and large counts were removed. Left side of the image shows hierarchical clustering between samples. (B) Correlation for RNA-seq experiment (replicates RNAseq_1-3). Matrix for the scatterplots was generated in deepTools2 with bin size set to 0.2kb. Correlation was calculated using Pearson method. Bins with large counts were removed. X- and Y-axes show a number of fragments in respective files.

**Table S4. Gene ontology analysis on clustered H3K27me3 target genes.**

| clu-ster | p-value | genes in GO term | genes in cluster (total) | common genes | term ID | GO sub-ontology | term name |
|---|---|---|---|---|---|---|---|
| 3 | 0.0049 | 10 | 136 | 5 | GO:0030908 | BP | protein splicing |
| 3 | 0.0049 | 10 | 136 | 5 | GO:0016539 | BP | intein-mediated protein splicing |
| 4 | 0.000146 | 402 | 2978 | 369 | GO:0050896 | BP | response to stimulus |
| 4 | 4.35E-05 | 267 | 2978 | 251 | GO:0006950 | BP | response to stress |
| 4 | 0.0351 | 911 | 2978 | 798 | GO:0046483 | BP | heterocycle metabolic process |
| 4 | 0.0206 | 899 | 2978 | 789 | GO:0006725 | BP | cellular aromatic compound metabolic process |
| 4 | 0.00497 | 940 | 2978 | 827 | GO:1901360 | BP | organic cyclic compound metabolic process |
| 4 | 0.0302 | 812 | 2978 | 714 | GO:0006139 | BP | nucleobase-containing compound metabolic process |
| 4 | 0.000306 | 617 | 2978 | 555 | GO:0090304 | BP | nucleic acid metabolic process |
| 4 | 0.0122 | 2994 | 2978 | 2540 | GO:0003674 | MF | molecular_function |
| 4 | 0.0148 | 1828 | 2978 | 1574 | GO:0003824 | MF | catalytic activity |
| 4 | 0.0144 | 614 | 2978 | 546 | GO:0016787 | MF | hydrolase activity |
| 4 | 0.00121 | 145 | 2978 | 139 | GO:0016788 | MF | hydrolase activity, acting on ester bonds |

No significant hits were found for cluster 1 and 2. GO analysis was performed in gProfiler with FDR threshold 0.05. FDR was calculated with Benjamin-Hochberg method.

## 4. Concluding discussion – connecting the dots

Collectively, this work presents important insights about H3K27me3-mediated chromatin repression - its evolutionary history in green lineage and novel layers of function facilitated by a PcG-associated protein, PWO1.

Regarding PWO1, this study showed that it interacts physically with nuclear lamina components and is also functional connected with one of the NL members, CRWN1, exemplified by transcriptomic control of similar set of target genes and a similar impact on the nuclear size in the corresponding mutants. Moreover, PWO1 does not influence global H3K27me3 level, but alters occupancy of the mark on selected genes in a direct or indirect way. Overall, the results presented here suggest that PWO1 might be a PRC2-associated protein acting downstream of H3K27me3 deposition possibly by influencing spatial distribution of its targets under specific conditions, i.e. stress-related. Such notion is supported by several features seen in the *pwo1* mutant: high number of transriptionally upregulated stress-related PRC2 targets; largely unchanged H3K27me3 occupancy, even despite its effect on the expression status; a mild whole-plant mutant phenotype (Hohenstatt, 2012) and a reduced nuclear area, a feature shown to impact the frequency of interchromosomal interactions. Furthermore, PWO1 contains a PWWP domain commonly implicated in binding of DNA (Laguri et al., 2008) or chromatin modifications like H3K36me3 (Dhayalan et al., 2010; Vermeulen et al., 2010). Consequently, the PWWP motif from PWO1 was shown to bind *in vitro* unmodified histone H3 and a range of H3 modifications, including H3K27me3, suggesting a role for PWO1 as a histone "reader" protein.

The effectors in the PcG-pathway downstream to H3K27me3 deposition are largely unknown. Up to date, several proteins, recognizing the H3K27me3 modification through their BAH, WD40 or chromodomain motifs, were identified (Musselman et al., 2012; Zhao et al., 2016). These include: EED, a subunit of PRC2 complex (Margueron et al., 2009); Pc, a subunit of the animal PRC1 complex (Fischle et al., 2003; Kaustov et al., 2011) or LHP1, a subunit of plant PRC1 complex (Turck et al., 2007) and the vertebrate-specific BAHD1 protein (Zhao et al., 2016). Such effectors are responsible for chromatin condensation (Pc) (Grau et al., 2011) or tethering other heterochromatin-associated factors (BAHD1, LHP1 or EED) (Bierne et al., 2009; Cao et al., 2014; Derkacheva and Hennig, 2014). Alternatively, they function in H3K27me3 propagation (EED, LHP1) (Margueron et al., 2009; Veluchamy

et al., 2016), making them only partially downstream to H3K27me3 deposition. Importantly, none of the currently known H3K27me3-downstream effectors was reported to regulate subnuclear compartmentalization of target genes, providing a possibility that PWO1 represents a long-awaited intermediate in the PcG pathway. In order to confirm such hypotheses, it is crucial to identify direct targets of PWO1 and find loci with PWO1-dependent localization within the nucleus. Unfortunately, chromatin immunoprecipitations with PWO1 bait (data not shown) and fluorescent in situ hybridizations for selected PWO1 targets (data not shown) performed so far experienced technical difficulties unsolvable within the timeframe of PhD project presented here. Further optimization of such techniques should resolve whether PWO1 indeed has a PcG-related spatial function downstream to H3K27me3 deposition.

In addition, this study investigated a phylogenetic conservation of PRC1/PRC2 components in the lower plant branches of plants and characterized H3K27me3 profile in the representative red alga, *Cyanidioschyzon merolae*. The results revealed that H3K27me3 in *C. merolae*targets predominantly repetitive elements, shows differential occupancy over gene bodies and their flanking regions, enriches at telomers and subtelomers, as well as represses genes involved in intein-mediated protein splicing. Overall, it can be speculated that H3K27me3-mediated chromatin repression fulfils the role of a 'genome guardian' in *C. merolae*, preventing potentially deleterious effects of repetitive elements' expression. The enrichment of H3K27me3 at repetitive elements in *C. merolae* and in the diatom *P. tricornutum* (Veluchamy et al., 2015), as well as a reactivation of dispersed TOC1 retrotransposon upon depletion of a EZH2 homolog in *Chlamydomonas reinhardtii* (Shaver et al., 2010) suggest an ancient role of H3K27me3 in repressing selfish genetic elements. It remains to be elucidated whether PRC2-mediated repression of repetitive elements evolved earlier than H3K27me3-mediated gene silencing or both mechanisms co-occurred in the common ancestor of the plant and animal kingdoms.

H3K27me3-mediated heterochromatinization of repetitive elements, evolutionary conservation of PRC2-mediated repression and PcG-dependent spatial organization of chromatin provide an interesting perspective for future studies. Despite the fact that 3D chromatin studies focus mostly on higher eukaryotes, several studies reported on the cohesin-mediated, spatial interaction between/within chromosomes in *Saccharomyces cerevisae* (Duan et al., 2010) and *Schizosaccharomyces pombe* (Mizuguchi et al., 2014). Interestingly, in *S. pombe,* proper cohesion loading and demarcation of chromosome territories is

dependent on Clr4, a H3K9-methyltransferase responsible for heterochromatin assembly (Cam et al., 2005). Moreover, the repressive chromatin domains in *S. pombe* and *S. cerevisae* were found to be enriched at the nuclear periphery (Steglich et al., 2013).

Altogether, these results suggest that chromatin in unicellular lower eukaryotes forms defined spatial organization, similarly to higher evolutionary branches, and that its regulation is dependent on heterochromatic modifications. Given that the *C. merolae* genome contains homologs of architectural proteins, including cohesin (Matsuzaki et al., 2004), and has a H3K27me3 machinery, it would be interesting to elaborate whether *C. merolae* H3K27me3 is important for heterochromatin compartmentalization and long-range 3D chromosomal interactions, as observed in several higher eukaryotes (Bantignies and Cavalli, 2011; Denholtz et al., 2013; Vieux-Rochas et al., 2015). However, such mechanism would be independent of PWO1 as its homologs were not found in *C. merolae*. In fact, the BLAST searches for PWO1 revealed an emergence of its homologs only in flowering plants (angiosperms), but an absence in lower plant branches (Fig.A2). Comparative analyses between species of different phylogenetic branches would elucidate core mechanisms of spatial arrangement of chromatin across the eukaryotic kingdoms.

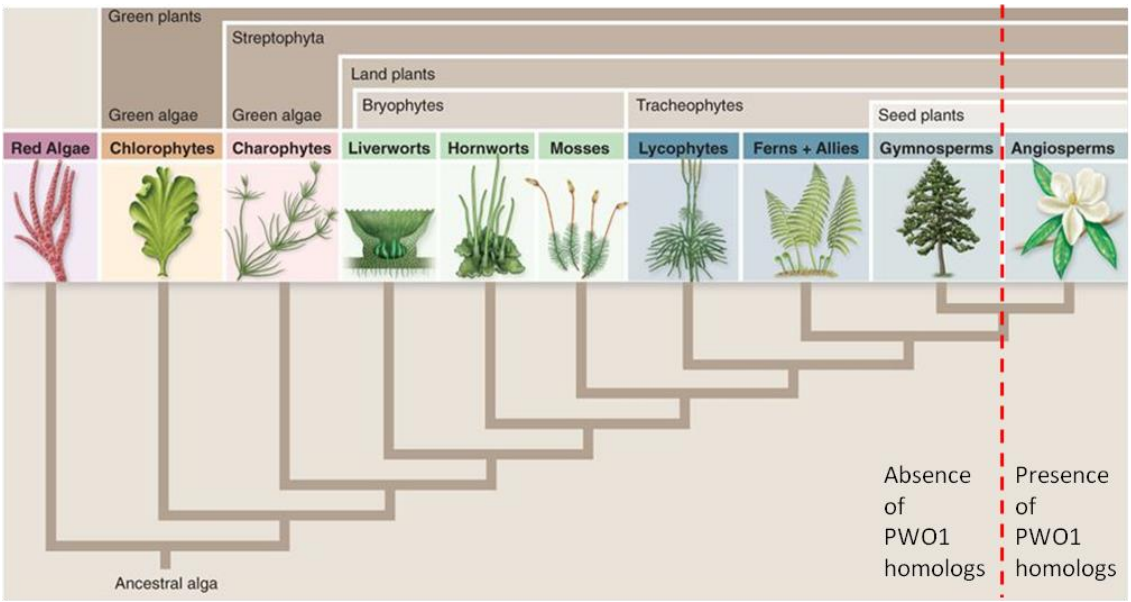



**Fig.A2. Conservation of PWO1 in plant species.** The figure was modified from: (Raven, 2011). The red dashed line demarcates an emergence of PWO1 homologs.

# 5. Abstract

## 5.1. Abstract

The widely conserved Polycomb group (PcG) proteins facilitate chromatin-mediated gene repression by modification of histone tails. One of the PcG protein complexes, Polycomb Repressive Complex 2 (PRC2), promotes repressive chromatin formation via trimethylation of lysine-27 on histone H3 (H3K27me3). Important aspects of the PRC2-pathway concern the influence on spatial organization of chromatin in the nucleus and evolutionary conservation in lower phylogenetic branches, with both themes largely unknown especially in the green lineage.

This study presents a functional association between PRC2-mediated repression and the nuclear lamina (NL), a peripheral protein mesh located underneath the inner nuclear membrane, in *Arabidopsis thaliana*. Namely, this study demonstrates that a novel PRC2-associated protein, PWWP INTERACTOR OF POLYCOMBS 1 (PWO1), associates physically with a several NL components, interacts genetically with one of the NL members, CROWDED NUCLEI 1 (CRWN1), and controls expression of a similar set of genes as CRWN1. Moreover, it is shown that PWO1 localizes partially to speckle-like foci at the subnuclear periphery. Interestingly, PWO1 does not influence the global H3K27me3 status, suggesting a redundancy within PWO family or a placement of PWO1 in the PRC2-pathway downstream to the H3K27 methylation mechanism.

In addition, this study investigates the evolutionary conservation of PRC2 components in unicellular representatives of lower plant branchesand the H3K27me3 features in the model red alga, *Cyanidioschyzon merolae*. The results demonstrate that PRC2 core proteins are widely distributed across phylogenetic subclades and *C. merolae* H3K27me3 has a canonical repressive function. Interestingly, *C. merolae* H3K27me3 targets predominantly repetitive elements in the genome and shows preferential localization at telomeric and subtelomeric chromosomal regions. Moreover, this work indicates that *C. merolae*H3K27me3 displays differential enrichment over gene body and is enriched at targets functionally connected with protein splicing.

In conclusion, this study deepens the understanding about PRC2 involvement in spatial organization of the chromatin in the nucleus and deciphers evolutionary conservation of PRC2-pathway in simple unicellular eukaryotes.

## 5.2. Zusammenfassung

Polycomb-Gruppen (PcG) Proteine erleichtern die Chromatin-vermittelte Genrepression durch Modifikation von Histonen in allen eukaryotischen Königreichen. Einer der PcG-Proteinkomplexe, Polycomb Repressive Complex 2 (PRC2), fördert die Bildung von reprimiertem Chromatindurch Tri-Methylierung von Lysin-27 an Histon H3 (H3K27me3). Zwei wichtige Aspekte der Regulation durchPRC2 sind der Einfluss auf die räumliche Organisation von Chromatin im Zellkern und die evolutionäre Konservierung in niederen phylogenetischen Kladen. Beide sindin Pflanzen weitgehend unbekannt.

Diese Studie in *Arabidopsis thaliana* präsentiert eine funktionelle Assoziation zwischen PRC2-vermittelter Repression und der Kern-Lamina (KL), einem peripheren Proteinnetz, das sich im Kernlumen an der inneren Kernmembran befindet. Diese Studie zeigt, dass das neuartige PRC2-assoziierte Protein PWWP-INTERACTOR OF POLYCOMBS 1 (PWO1) physisch mit einer Anzahl von KL-Komponenten interagiert und mit einem der KL-Mitglieder, CROWDED NUCLEI 1 (CRWN1), genetisch interagiert. Desweiteren regulieren PWO1 und CRWN1 die gleichen Gene. PWO1 lokalisiert teilweise in „Speckle“-Strukturen an der Peripherie des Kerns. Interessanterweise beeinflusst PWO1 nicht den globalen H3K27me3-Status, was auf eine Redundanz mit seinen Homologen in *Arabidopsis* oder auf eine Rollein PRC2-vermittelter Genregulation unterhalb des H3K27-Methylierungsmechanismus hindeutet.

Darüber hinaus wird in dieser Studie die evolutionäre Konservierung von PRC2-Komponenten bei einzelligen Repräsentanten von niederen Pflanzenarten und H3K27me3 in der Rotalge *Cyanidioschyzon merolae* untersucht. Die Ergebnisse zeigen, dass PRC2-Proteine weit über phylogenetische Subkladen verteilt sind und H3K27me3 in *C. merolae* eine kanonisch repressive Funktion hat. Interessanterweise findet man H3K27me3 überwiegend an repetitiven Elementenin Telomeren und Subtelomerenim Genom von *C. merolae*. Außerdem zeigt diese Arbeit, dass H3K27me3 in *C. merolae* differenziell über Gene verteilt ist und Gene reguliert, die Funktionen im Protein-splicing haben.

Zusammengefasstvertieft diese Studie das Verständnis über die Funktion von PRC2 in der räumlichen Organisation von Chromatin im Kern und trägt zur Entschlüsselung der evolutionären Konservierung des PRC2-Weges in einfachen einzelligen Eukaryoten bei.

## 6. References


Baker, M. (2011). Making sense of chromatin states. *Nat Methods* 8, 717–722. doi:10.1038/nmeth.1673.

Bantignies, F., and Cavalli, G. (2011). Polycomb group proteins: Repression in 3D. *Trends Genet.* 27, 454–464. doi:10.1016/j.tig.2011.06.008.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., et al. (2007). High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* 129, 823–837. doi:10.1016/j.cell.2007.05.009.

Berke, L., and Snel, B. (2015). The plant Polycomb repressive complex 1 ( PRC1 ) existed in the ancestor of seed plants and has a complex duplication history. 1, 1–10. doi:10.1186/s12862-015-0319-z.

Bierne, H., Tham, T. N., Batsche, E., Dumay, A., Leguillou, M., Kernéis-Golsteyn, S., et al. (2009). Human BAHD1 promotes heterochromatic gene silencing. *Proc. Natl. Acad. Sci. U. S. A.* 106, 13826–31. doi:10.1073/pnas.0901259106.

Bouyer, D., Roudier, F., Heese, M., Andersen, E. D., Gey, D., Nowack, M. K., et al. (2011). Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. *PLoS Genet.* 7. doi:10.1371/journal.pgen.1002014.

Bowman, J. L., Floyd, S. K., and Sakakibara, K. (2007). Green genes-comparative genomics of the green branch of life. *Cell* 129, 229–34. doi:10.1016/j.cell.2007.04.004.

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441, 349–353. doi:10.1038/nature04733.

Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. 1123–1136. doi:10.1101/gad.381706.present.

Breiling, A., Turner, B. M., Bianchi, M. E., and Orlando, V. (2001). General transcription factors bind promoters repressed by Polycomb group proteins. *Nature* 412, 651–655. doi:10.1038/35088090.

Butenko, Y., and Ohad, N. (2011). Polycomb-group mediated epigenetic mechanisms through plant evolution. *Biochim. Biophys. Acta* 1809, 395–406. doi:10.1016/j.bbagrm.2011.05.013.

Calonje, M., Sanchez, R., Chen, L., and Sung, Z. R. (2008). EMBRYONIC FLOWER1 participates in polycomb group-mediated AG gene silencing in Arabidopsis. *Plant Cell* 20, 277–291. doi:10.1105/tpc.106.049957.

Cam, H. P., Sugiyama, T., Chen, E. S., Chen, X., FitzGerald, P. C., and Grewal, S. I. S. (2005). Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nat. Genet.* 37, 809–819. doi:10.1038/ng1602.

Cao, Q., Wang, X., Zhao, M., Yang, R., Malik, R., Qiao, Y., et al. (2014). The central role of EED in the orchestration of polycomb group complexes. *Nat. Commun.* 5, 343–349. doi:10.1038/ncomms4127.

Deleris, A., Stroud, H., Bernatavichute, Y., Johnson, E., Klein, G., Schubert, D., et al. (2012). Loss of the DNA methyltransferase MET1 Induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in Arabidopsis thaliana. *PLoS Genet.* 8, e1003062. doi:10.1371/journal.pgen.1003062.

Denholtz, M., Bonora, G., Chronis, C., Splinter, E., de Laat, W., Ernst, J., et al. (2013). Long-Range Chromatin Contacts in Embryonic Stem Cells Reveal a Role for Pluripotency Factors and Polycomb Proteins in Genome Organization. *Cell Stem Cell* 13, 602–616. doi:10.1016/j.stem.2013.08.013.

Derkacheva, M., and Hennig, L. (2014). Variations on a theme: Polycomb group proteins in plants. *J. Exp. Bot.* 65, 2769–2784. doi:10.1093/jxb/ert410.

Derkacheva, M., Steinbach, Y., Wildhaber, T., Mozgová, I., Mahrez, W., Nanni, P., et al. (2013). Arabidopsis MSI1 connects LHP1 to PRC2 complexes. *EMBO J.* 32, 2073–85. doi:10.1038/emboj.2013.145.

Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R. Z., Ragozin, S., et al. (2010). The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.* 285, 26114–26120. doi:10.1074/jbc.M109.089433.

Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y. J., Lee, C., et al. (2010). A three-dimensional model of the yeast genome. *Nature* 465, 363–367. doi:10.1038/nature08973.

Dumbliauskas, E., Lechner, E., Jaciubek, M., Berr, A., Pazhouhandeh, M., Alioua, M., et al. (2011). The Arabidopsis CUL4-DDB1 complex interacts with MSI1 and is required to maintain MEDEA parental imprinting. *EMBO J.* 30, 731–43. doi:10.1038/emboj.2010.359.

Dumesic, P. A., Homer, C. M., Moresco, J. J., Pack, L. R., Shanle, E. K., Coyle, S. M., et al. (2015). Product binding enforces the genomic specificity of a yeast Polycomb repressive complex. *Cell* 160, 204–218. doi:10.1016/j.cell.2014.11.039.

Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43–9. doi:10.1038/nature09906.

Fang, L., Wuptra, K., Chen, D., Li, H., Huang, S.-K., Jin, C., et al. (2014). Environmental-stress-induced Chromatin Regulation and its Heritability. *J. Carcinog. Mutagen.* 5, 1–19. doi:10.4172/2157-2518.1000156.

Finch, J. T., Lutter, L. C., Rhodes, D., Brown, R. S., Rushton, B., Levitt, M., et al. (1977). Structure of nucleosome core particles of chromatin. *Nature* 269, 29–36. doi:10.1038/269029a0.

Fischle, W., Wang, Y., Jacobs, S. A., Kim, Y., Allis, C. D., and Khorasanizadeh, S. (2003). Molecular basis for

the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* 17, 1870–81. doi:10.1101/gad.1110503.

Gaudin, V., Libault, M., Pouteau, S., Juul, T., Zhao, G., Lefebvre, D., et al. (2001). Mutations in LIKE HETEROCHROMATIN PROTEIN 1 affect flowering time and plant architecture in Arabidopsis. *Development* 128, 4847–58.

Grau, D. J., Chapman, B. a., Garlick, J. D., Borowsky, M., Francis, N. J., and Kingston, R. E. (2011). Compaction of chromatin by diverse polycomb group proteins requires localized regions of high charge. *Genes Dev.* 25, 2210–2221. doi:10.1101/gad.17288211.

Hansen, K. H., Bracken, A. P., Pasini, D., Dietrich, N., Gehani, S. S., Monrad, A., et al. (2008). A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.* 10, 1291–1300. doi:10.1038/ncb1787.

Hohenstatt, M. L. (2012). Functional analysis of SCI1 – A PWWP domain protein involved in Polycomb group mediated gene regulation in Arabidopsis.

Ishak, C. A., Marshall, A. E., Passos, D. T., White, C. R., Kim, S. J., Cecchini, M. J., et al. (2016). An RB-EZH2 Complex Mediates Silencing of Repetitive DNA Sequences. *Mol. Cell* 64, 1074–1087. doi:10.1016/j.molcel.2016.10.021.

Jeong, C. W., Roh, H., Dang, T. V., Choi, Y. Do, Fischer, R. L., Lee, J. S., et al. (2011). An E3 ligase complex regulates SET-domain polycomb group protein activity in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U. S. A.* 108, 8036–41. doi:10.1073/pnas.1104232108.

Jullien, P. E., Mosquna, A., Ingouff, M., Sakata, T., Ohad, N., Berger, F., et al. (2008). Retinoblastoma and Its Binding Partner MSI1 Control Imprinting in Arabidopsis. *PLoS Biol.* 6, e194. doi:10.1371/journal.pbio.0060194.

Kalushkova, A., Fryknäs, M., Lemaire, M., Fristedt, C., Agarwal, P., Eriksson, M., et al. (2010). Polycomb target genes are silenced in multiple myeloma. *PLoS One* 5. doi:10.1371/journal.pone.0011483.

Kaustov, L., Ouyang, H., Amaya, M., Lemak, A., Nady, N., Duan, S., et al. (2011). Recognition and specificity determinants of the human cbx chromodomains. *J. Biol. Chem.* 286, 521–9. doi:10.1074/jbc.M110.191411.

Kim, S. Y., Lee, J., Eshed-Williams, L., Zilberman, D., and Sung, Z. R. (2012). EMF1 and PRC2 cooperate to repress key regulators of arabidopsis development. *PLoS Genet.* 8. doi:10.1371/journal.pgen.1002512.

Kleinmanns, J. A., and Schubert, D. (2014). Polycomb and Trithorax group protein-mediated control of stress responses in plants. *Biol. Chem.* 395, 1291–300. doi:10.1515/hsz-2014-0197.

Kuwabara, A., and Gruissem, W. (2014). Arabidopsis RETINOBLASTOMA-RELATED and Polycomb group proteins: cooperation during plant cell differentiation and development. *J. Exp. Bot.* 65, 2667–76. doi:10.1093/jxb/eru069.

Lafos, M., Kroll, P., Hohenstatt, M. L., Thorpe, F. L., Clarenz, O., and Schubert, D. (2011). Dynamic regulation of H3K27 trimethylation during arabidopsis differentiation. *PLoS Genet.* 7. doi:10.1371/journal.pgen.1002040.

Laguri, C., Duband-Goulet, I., Friedrich, N., Axt, M., Belin, P., Callebaut, I., et al. (2008). Human Mismatch Repair Protein MSH6 Contains a PWWP Domain That Targets Double Stranded DNA †. *Biochemistry* 47, 6199–6207. doi:10.1021/bi7024639.

Lindner, M., Simonini, S., Kooiker, M., Gagliardini, V., Somssich, M., Hohenstatt, M., et al. (2013). TAF13 interacts with PRC2 members and is essential for Arabidopsis seed development. *Dev. Biol.* 379, 28–37. doi:10.1016/j.ydbio.2013.03.005.

Liu, Y., Taverna, S. D., Muratore, T. L., Shabanowitz, J., Hunt, D. F., and Allis, C. D. (2007). RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. 1530–1545. doi:10.1101/gad.1544207.structures.

Lodha, M., Marco, C. F., and Timmermans, M. C. P. (2008). Genetic and epigenetic regulation of stem cell homeostasis in plants. *Cold Spring Harb. Symp. Quant. Biol.* 73, 243–51. doi:10.1101/sqb.2008.73.044.

Lodha, M., Marco, C. F., and Timmermans, M. C. P. (2013). The ASYMMETRIC LEAVES complex maintains repression of KNOX homeobox genes via direct recruitment of Polycomb-repressive complex2. *Genes Dev.* 27, 596–601. doi:10.1101/gad.211425.112.

De Lucia, F., Crevillen, P., Jones, A. M. E., Greb, T., and Dean, C. (2008). A PHD-Polycomb Repressive Complex 2 triggers the epigenetic silencing of FLC during vernalization. *Proc. Natl. Acad. Sci.* 105, 16831–16836. doi:10.1073/pnas.0808687105.

Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M., and Richmond, T. J. (1997). Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. Mol. Biol.* 272, 301–11. doi:10.1006/jmbi.1997.1235.

Margueron, R., Justin, N., Ohno, K., Sharpe, M. L., Son, J., Drury III, W. J., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* 461, 762–767. doi:10.1038/nature08398.

Martínez-Balbás, M. A., Tsukiyama, T., Gdula, D., and Wu, C. (1998). Drosophila NURF-55, a WD repeat protein involved in histone metabolism. *Proc. Natl. Acad. Sci. U. S. A.* 95, 132–7. doi:10.1073/pnas.95.1.132.

Matsuzaki, M., Misumi, O., Shin-I, T., Maruyama, S., Takahara, M., Miyagishima, S.-Y., et al. (2004). Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. *Nature* 428, 653–657. doi:10.1038/nature02398.

Mizuguchi, T., Fudenberg, G., Mehta, S., Belton, J.-M., Taneja, N., Folco, H. D., et al. (2014). Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. *Nature* 516, 432–

435. doi:10.1038/nature13833.

Moazed, D., and O'Farrell, P. H. (1992). Maintenance of the engrailed expression pattern by Polycomb group genes in Drosophila. *Development* 116, 805–810.

Molitor, A., and Shen, W. H. (2013). The Polycomb Complex PRC1: Composition and Function in Plants. *J. Genet. Genomics* 40, 231–238. doi:10.1016/j.jgg.2012.12.005.

Mosquna, A., Katz, A., Shochat, S., Grafi, G., and Ohad, N. (2004). Interaction of FIE, a Polycomb protein, with pRb: a possible mechanism regulating endosperm development. *Mol. Genet. Genomics* 271, 651–657. doi:10.1007/s00438-004-1024-6.

Mozgova, I., and Hennig, L. (2015). The Polycomb Group Protein Regulatory Network. *Annu. Rev. Plant Biol.* 66, 269–296. doi:10.1146/annurev-arplant-043014-115627.

Musselman, C. A., Lalonde, M.-E., Côté, J., and Kutateladze, T. G. (2012). Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.* 19, 1218–1227. doi:10.1038/nsmb.2436.

Nozaki, H., Takano, H., Misumi, O., Terasawa, K., Matsuzaki, M., Maruyama, S., et al. (2007). A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga Cyanidioschyzon merolae. *BMC Biol.* 5, 28. doi:10.1186/1741-7007-5-28.

Oh, S., Park, S., and van Nocker, S. (2008). Genic and global functions for Paf1C in chromatin modification and gene expression in Arabidopsis. *PLoS Genet.* 4, e1000077. doi:10.1371/journal.pgen.1000077.

Park, S., Oh, S., and van Nocker, S. (2012). Genomic and gene-level distribution of histone H3 dimethyl lysine-27 (H3K27me2) in Arabidopsis. *PLoS One* 7, e52855. doi:10.1371/journal.pone.0052855.

Pazhouhandeh, M., Molinier, J., Berr, A., and Genschik, P. (2011). MSI4/FVE interacts with CUL4-DDB1 and a PRC2-like complex to control epigenetic regulation of flowering time in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 108, 3430–5. doi:10.1073/pnas.1018242108.

Del Prete, S., Mikulski, P., Schubert, D., and Gaudin, V. (2015). One, two, three: Polycomb proteins hit all dimensions of gene regulation. *Genes (Basel).* 6, 520–542. doi:10.3390/genes6030520.

Probst, A. V, Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. 192–206. doi:10.1038/nrm2640.

Ramakrishnan, V. (1997). Histone H1 and chromatin higher-order structure. *Crit. Rev. Eukaryot. Gene Expr.* 7, 215–30.

Raven, P. H. (2011). *Biology*. McGraw-Hill.

Ringrose, L. (2007). Polycomb comes of age: genome-wide profiling of target sites. *Curr. Opin. Cell Biol.* 19, 290–297. doi:10.1016/j.ceb.2007.04.010.

Saurin, A. J., Shao, Z., Erdjument-Bromage, H., Tempst, P., and Kingston, R. E. (2001). A Drosophila

Polycomb group complex includes Zeste and dTAFII proteins. *Nature* 412, 655–660. doi:10.1038/35088096.

Schatlowski, N., Stahl, Y., Hohenstatt, M. L., Goodrich, J., and Schubert, D. (2010). The CURLY LEAF interacting protein BLISTER controls expression of polycomb-group target genes and cellular differentiation of Arabidopsis thaliana. *Plant Cell* 22, 2291–2305. doi:10.1105/tpc.109.073403.

Schwartz, Y. B., and Pirrotta, V. (2008). Polycomb complexes and epigenetic states. *Curr. Opin. Cell Biol.* 20, 266–273. doi:10.1016/j.ceb.2008.03.002.

Schwartz, Y. B., and Pirrotta, V. (2013). A new world of Polycombs: unexpected partnerships and emerging functions. *Nat. Rev. Genet.* 14, 853–64. doi:10.1038/nrg3603.

Shaver, S., Casas-Mollano, J. A., Cerny, R. L., and Cerutti, H. (2010). Origin of the polycomb repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga Chlamydomonas. *Epigenetics* 5, 301–312. doi:10.4161/epi.5.4.11608.

Sowpati, D. T., Ramamoorthy, S., and Mishra, R. K. (2015). Expansion of the polycomb system and evolution of complexity. *Mech. Dev.* 138, 97–112. doi:10.1016/j.mod.2015.07.013.

Spivakov, M., and Fisher, A. G. (2007). Epigenetic signatures of stem-cell identity. *Nat. Rev. Genet.* 8, 263–271. doi:10.1038/nrg2046.

Steglich, B., Sazer, S., and Ekwall, K. (2013). Transcriptional regulation at the yeast nuclear envelope. *Nucleus* 4, 379–89. doi:10.4161/nucl.26394.

Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.-L., Guillaume, E., Buisine, N., et al. (2007). Arabidopsis TFL2/LHP1 Specifically Associates with Genes Marked by Trimethylation of Histone H3 Lysine 27. *PLoS Genet.* 3, e86. doi:10.1371/journal.pgen.0030086.

Tyler, J. K., Bulger, M., Kamakaka, R. T., Kobayashi, R., and Kadonaga, J. T. (1996). The p55 subunit of Drosophila chromatin assembly factor 1 is homologous to a histone deacetylase-associated protein. *Mol. Cell. Biol.* 16, 6149–6159.

Veluchamy, A., Jégu, T., Ariel, F., Latrasse, D., Mariappan, K. G., Kim, S.-K., et al. (2016). LHP1 Regulates H3K27me3 Spreading and Shapes the Three-Dimensional Conformation of the Arabidopsis Genome. *PLoS One* 11, e0158936. doi:10.1371/journal.pone.0158936.

Veluchamy, A., Rastogi, A., Lin, X., Lombard, B., Murik, O., Thomas, Y., et al. (2015). An integrative analysis of post-translational histone modifications in the marine diatom Phaeodactylum tricornutum. *Genome Biol.* 16, 102. doi:10.1186/s13059-015-0671-8.

Vermeulen, M., Eberl, H. C., Matarese, F., Marks, H., Denissov, S., Butter, F., et al. (2010). Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers. *Cell* 142, 967–980. doi:10.1016/j.cell.2010.08.020.

Vieux-Rochas, M., Fabre, P. J., Leleu, M., Duboule, D., and Noordermeer, D. (2015). Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. *Proc. Natl. Acad. Sci. U. S. A.* 112, 4672–7. doi:10.1073/pnas.1504783112.

Whitcomb, S. J., Basu, A., Allis, C. D., and Bernstein, E. (2007). Polycomb Group proteins: an evolutionary perspective. *Trends Genet.* 23, 494–502. doi:10.1016/j.tig.2007.08.006.

Wood, C. C., Robertson, M., Tanner, G., Peacock, W. J., Dennis, E. S., and Helliwell, C. A. (2006). The Arabidopsis thaliana vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. *Proc. Natl. Acad. Sci. U. S. A.* 103, 14631–6. doi:10.1073/pnas.0606385103.

Wu, H., Zeng, H., Lam, R., Tempel, W., Amaya, M. F., Xu, C., et al. (2011). Structural and histone binding ability characterizations of human PWWP domains. *PLoS One* 6. doi:10.1371/journal.pone.0018919.

Yin, H., Sweeney, S., Raha, D., Snyder, M., and Lin, H. (2011). A high-resolution whole-genome map of key chromatin modifications in the adult Drosophila melanogaster. *PLoS Genet.* 7, e1002380. doi:10.1371/journal.pgen.1002380.

Yong, D. C., Ying, H., and Shen, R. W. (2016). The evolutionary landscape of PRC1 core components in green lineage. *Planta* 243, 825–846. doi:10.1007/s00425-015-2451-9.

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y. V, Pellegrini, M., Goodrich, J., et al. (2007). Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. *PLoS Biol.* 5, e129. doi:10.1371/journal.pbio.0050129.

Zhao, D., Zhang, X., Guan, H., Xiong, X., Shi, X., Deng, H., et al. (2016). The BAH domain of BAHD1 is a histone H3K27me3 reader. *Protein Cell* 7, 222–6. doi:10.1007/s13238-016-0243-z.

## 7. List of publications

**Publications resulting from this work:**

del Prete S*, Mikulski P*, Schubert D, Gaudin V. One, Two, Three: Polycomb Proteins Hit All Dimensions of Gene Regulation. Genes 2015, 6, 520-542. (*equal contribution)

Bey T, Jamge S, Klemme S, Komar DN, Le Gall S, Mikulski P, Schmidt M, Zicola J, Berr A. Chromatin and epigenetics in all their states: Meeting report of the first conference on Epigenetic and Chromatin Regulation of Plant Traits - January 14 – 15, 2016 - Strasbourg, France. Epigenetics 2016, 11:8, 625-634.

## 8. CV

**For reasons of data protection, the curriculum vitae is not published in the electronic version.**

**For reasons of data protection, the curriculum vitae is not published in the electronic version.**

**For reasons of data protection, the curriculum vitae is not published in the electronic version.**

# 9. Appendix

## 9.1. Abbreviations

| | | | |
|---|---|---|---|
| 3D | three-dimensional | mCh | mCherry fluorophore |
| °C | degree Celsius | mg | milligram |
| aa | amino acid | min | minute |
| bp | base pair | ml | milliliter |
| Col | Ecotype Columbia-0 | mM | millimolar |
| ChIP | chromatin immunoprecipitation | MS | Murashige & Skoog Medium |
| ChIP-seq | ChIP-sequencing | µg | microgram |
| *C. merolae* | *Cyanidioschyzon merolae* | µl | microliter |
| dH2O | distilled water | µm | micrometer |
| DNA | deoxyribonucleic acid | ng | nanogram |
| EtOH | ethanol | NL | nuclear lamina |
| Gal4-AD | Gal4-activating domain | NLS | nuclear localization signal |
| Gal4-BD | Gal4-binding domain | nm | nanometer |
| GFP | green fluorescent protein | PCR | polymerase chain reaction |
| GO | gene ontology | PcG | Polycomb group |
| H1-4 | histone 1-4 | qPCR | quantitative PCR |
| INM | inner nuclear membrane | RNA-seq | RNA-sequencing |
| IP | immunoprecipitation | RT-qPCR | Reverse Transcription-quantitive PCR |
| L | liter | RNA | ribonucleic acid |
| M | molar | | |

| | | | |
|---|---|---|---|
| T-DNA | transfer DNA | YEB | yeast extract broth |
| w/v | weight per volume | YPD | yeast extract peptone dextrose |
| wt | wild type | | |

## 9.2. Epigenetics in simple terms

### What epigenetics is

**Pawel Mikulski1*, BlaiseWeber2*, Daniel Schubert1a, Maike Stam2a**

1 Institute for Biology, Free University Berlin, Germany

2 SILS Institute, University of Amsterdam, Netherlands

* equal contribution

a corresponding authors: dan.schubert@fu-berlin.de, M.E.Stam@uva.nl

Keywords: epigenetics, histone modifications, dna methylation, outreach, music partition

Author contributions:

PM and BW wrote the manuscript and created figure 3. BW created figures 1 and 2. DS and MS revised the manuscript.

## What epigenetics is:

What is epigenetics? Let us explain this term using some analogies.

Take a look at your body. You can easily distinguish your eyes from your skin, and your nose from your mouth. Similarly, plants are also made of different organs: roots, stems, leaves, flowers etc. The reason behind such large diversity in living organisms comes from the existence of different tissues – groups of similar cells with a specialized function. Without this specialization we would be just a mass of simple cells that cannot, for example, walk or talk. Consequently, a plant would not be able to produce flowers or absorb sunlight. Even though the difference between cell types in your body is pretty evident, almost all the cells contain the same DNA. The DNA code is identical but is interpreted differently by different cell types. Epigenetics can be described as the tool by which DNA can be interpreted in different ways without changing the DNA sequence.

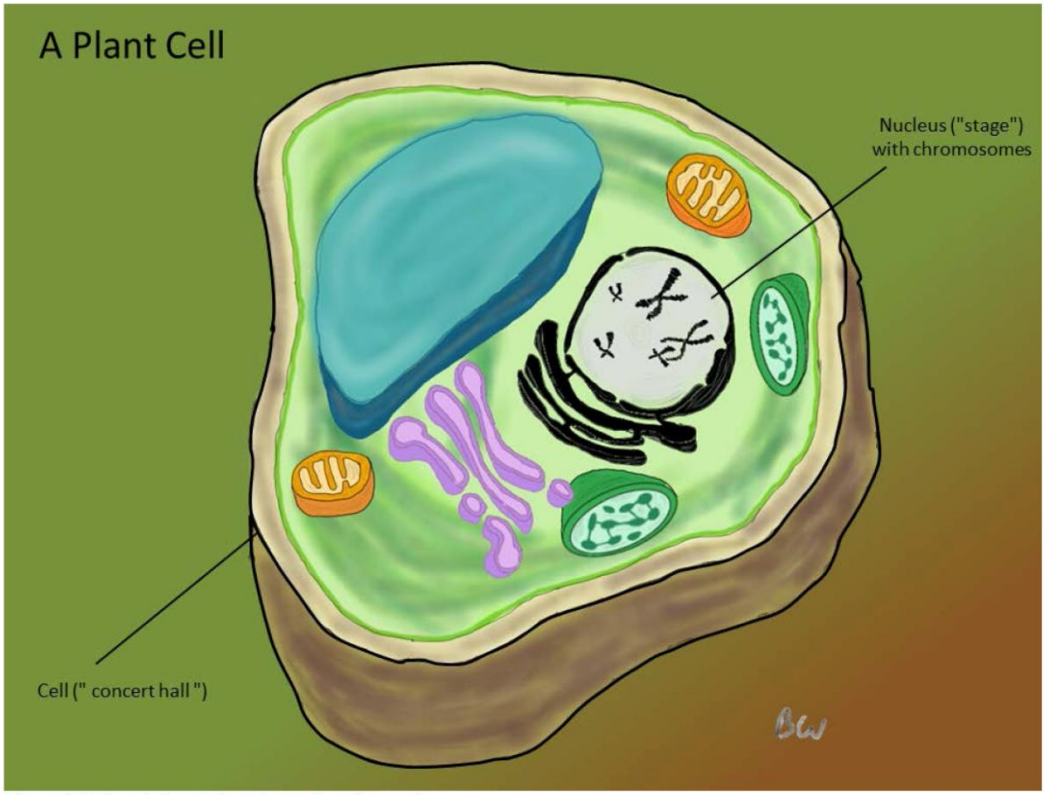



Figure 1 Cell and the nucleus in analogy to music.

We will explain epigenetics using an analogy to playing music. First, imagine each cell as a concert hall. At the center of it, stands the stage. This stage is the core of the cell: the nucleus. It is the place where chromosomes containing all the DNA sequences including genes are gathered. Genes

encode the genetic information that makes an organism: "the book of life". The genetic code itself is based on the use of four letters A, T, G and C called nucleotides that assemble together to form the DNA double helix. Translation of one encoded sequence (one gene) leads to the formation of one factor (protein) and life itself relies on the simultaneous translation of many of these sequences (and therefore proteins). To transpose this in our musical allegory, DNA sequences (genes) are like music partitions where nucleotides are the musical notes. Translating the code of the different genes into proteins then resemble a situation where musicians come together to play different partitions and produce all the melodies that compose the symphony of life. To guide musicians on how and when to play a partition, music sheets are usually annotated with symbols that indicate the dynamics of the music, for example: "*ff*" for fortissimo (very loud) and "*pp*" for pianissimo (very soft). Importantly, these symbols do not change the notes themselves, just the way they are played. In similar manner, cells also contain "guidelines" on how a gene should be interpreted: Epigenetic marks. According to the given epigenetic indications, a gene can be (strongly) activated or silenced, depending on the cell you are looking at. Last but not least, a crucial aspect of epigenetic marks is their heritability. Marks will be transmitted through cell divisions or even from one generation to the other. This allows the establishment of a cellular memory. Interpretative annotations will be transmitted from the mother cell to both daughter cells therefore allowing maintenance of tissue identity. To explain this in music terms, it would be like parents giving a set of partitions to their children that encompass all the manuscript indications they have been adding to the partitions. Thus children would play partitions just like their parents used to do.

## When compaction shapes gene expression:

Thus, everything revolves around the nucleus where all the genes are being stored and regulated. But this nucleus is a very small structure: The tenth of the width of a human hair. Given the number of DNA contained in the nucleus, compaction of DNA is required to fit in such small space. Nature indeed found a way to tackle this problem. A compaction is accomplished via wrapping of the DNA double helix around complexes of particular proteins (histones) therefore forming a structure called nucleosomes. To get a better picture, one could imagine a string (DNA) wrapped around beads (the histones). This process is in a way comparable to the use of compacted MP3s instead of a raw file that takes much more storage space.

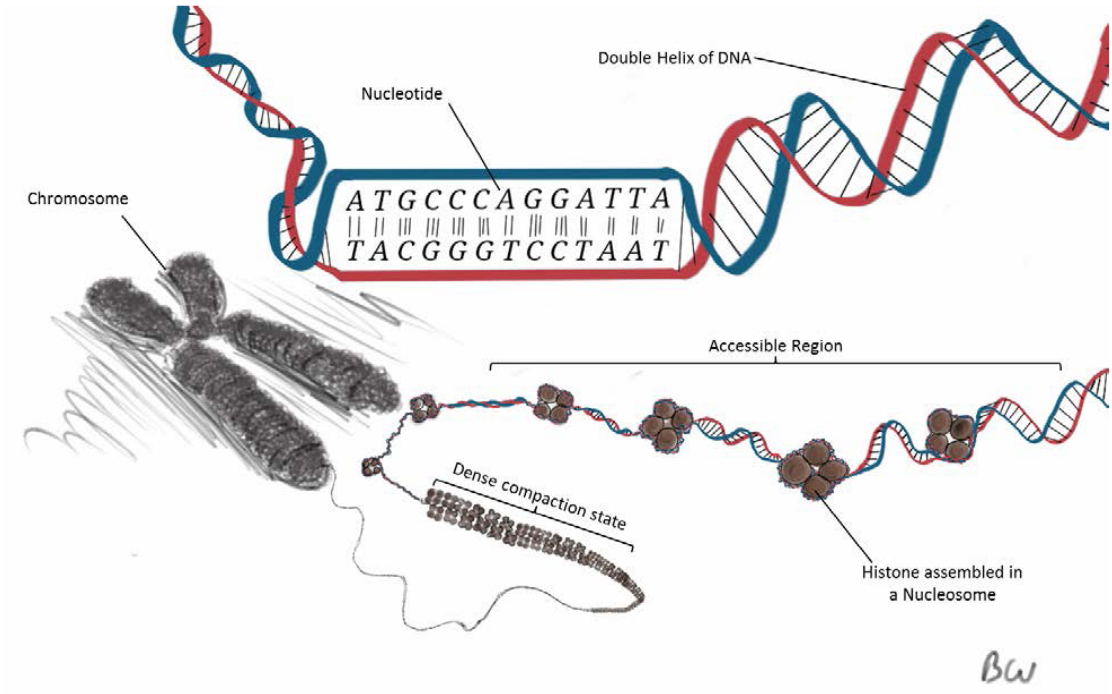



Figure 2 Compaction of DNA

Unfortunately, beauty often lies in complexity and the process of compaction is not that straightforward. High compaction of all the DNA sequences will almost certainly erase all the necessary information about a cell's behavior. No accessible partitions for the musicians, no music, no life. Therefore compaction needs to be applied in a selective way where dispensable genes would be compressed while crucial ones remain active. To annotate particular regions of DNA, the cell follows special instructions, called in biology: epigenetic modifications.

## Epigenetic modifications:

We will distinguish the two major modifications associated with epigenetic gene regulation: DNA methylation and histone modifications. But before entering into a detailed description, it is important to remember that epigenetic factors and modifications do not change the DNA sequence. They leave the sequence of nucleotides as it is. Instead, they influence the interpretation of DNA – in a symphony, the way the partition will be played by the musicians e.g silent, pianissimo (*pp*) or loud, fortissimo (*ff*).

DNA methylation:

This layer of epigenetic regulation acts on the sequence of DNA, more precisely on the C letters of the genetic code. Remember, no modification is made to the original sequence. So the

DNA sequences stay the same. Only, DNA methylation adds a “red mark” called a Methyl group on top of the C. Overall, DNA methylation mostly impairs gene expression, this by reducing accessibility either directly or indirectly. The context of this mark is also an important aspect of it. In mammals, only C preceding G letters can be methylated. This point is actually crucial for the heritability. But in plants, our main interest among the EpiTRAITS consortium, the context of DNA methylation is broader with C preceding G letters, C preceding H and G or H and H letters (H being either a A, a T or a G) and brings extra interrogation on how heritability is established in the latter context.

Histone Modifications:

This second layer of epigenetic regulation acts on histones associated with DNA, the ‘beads’ on which DNA is wrapped around. Histones, similarly to the C letters in DNA, can be modified. Histone modifications are ruled by a specific code referred to as the “histone code”. According to the category of genes they are targeting, the modifications will not be the same. For instance, active genes are often associated with acetylation marks which act as “green positive flags” while methylation marks often act as “red negative flags”. “Activating” histone marks will ensure a relaxed state of compaction and shapes accessible DNA regions while “repressive” Histone marks will lead to a higher degree of compaction, meaning less accessibility and inactive genes.

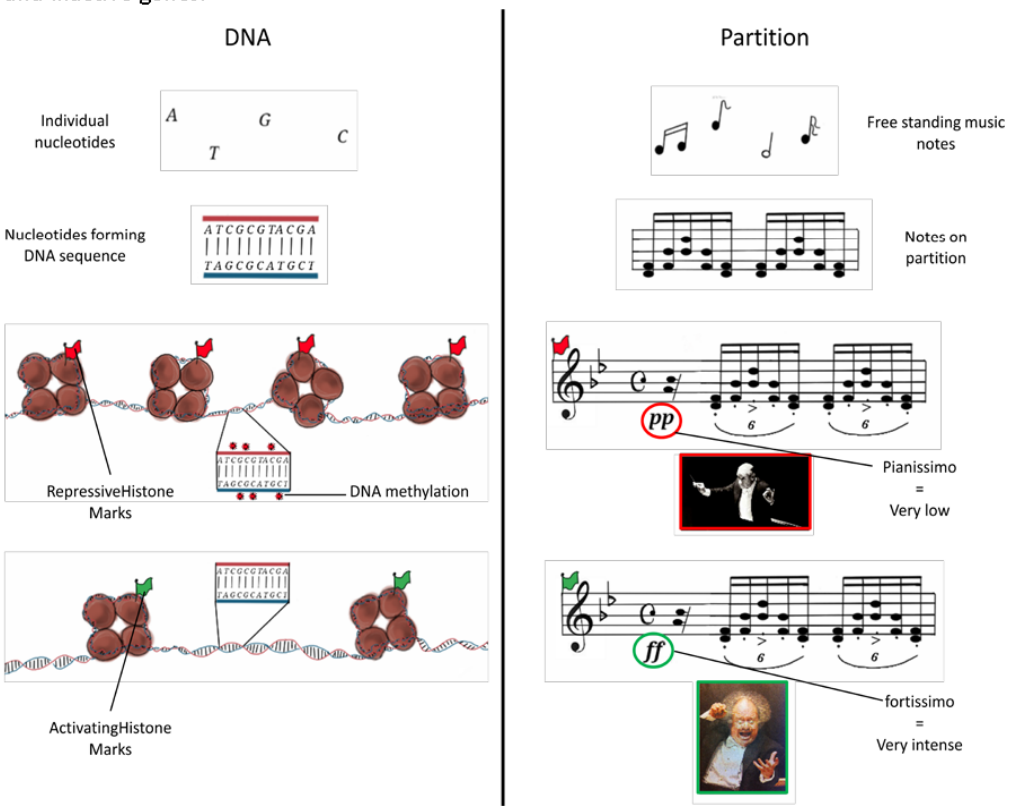



Figure 3 Dynamics of gene regulation as musical notation.

## Conclusion:

Altogether, from one "book of life" containing the same partitions in every cells, epigenetics conducts a cell specific symphony. By selectively silencing or fine tuning the way each partitions has to be played, epigenetic regulation adapts the symphony depending on the cell type. If you could hear it, it would be like having one specific composition for leaf cells, another for root cells and another one for flower composing cells. As mentioned earlier, epigenetic defines guidelines that are heritably passed from one generation to the next. But beyond serving the purpose of establishing a cellular memory, this heritable aspect, combined with the dynamic aspect of epigenetic modifications is deeply questioning our vision of classic evolution. Interestingly, epigenetic marks have been found to be under the influence of both internal and environmental signals which ultimately influence gene expression. For instance, an organism such as plant, when exposed to a specific stress (drought, cold, warm, pathogens etc.) can acquire a higher stress resistance via epigenetic mechanisms and possibly transmit this better fitness to its progeny. Thus, epigenetic inheritance tends to blur the boundary between nature and nurture, what comes from the fixed DNA code and what you gain from the environment, by editing an extra code on top of the first one. The final word should be given to Sir Conrad Waddington, who firstly coined the modern term of epigenetics and declared that an epigenetic consideration of the theory of evolution would go some way toward healing it.

## See more:

For more information about epigenetics:

http://www.epigenesys.eu/en/homepage

http://learn.genetics.utah.edu/content/epigenetics/

http://www.nature.com/scitable/spotlight/epigenetics-26097411

**Blaise Weber, Pawel Mikulski**

www.epitraits.eu


## 9.3. Broader current view on chromatin

**Chromatin and epigenetics in all their states: Meeting report of the first conference on Epigenetic and Chromatin Regulation of Plant Traits - January 14 – 15, 2016 - Strasbourg, France**

Till Bey1, Suraj Jamge2, Sonia Klemme3, Dorota Natalia Komar4, Sabine Le Gall5, Pawel Mikulski6, Schmidt5, Johan Zicola7, Alexandre Berr8

1 SILS Institute, University of Amsterdam, Netherlands

2 Laboratory of Molecular Biology, Wageningen University, Netherlands

3 Crop Science Division, Bayer CropScience SA-NV Zwijnaarde, Belgium

4 Centre for Plant Biotechnology and Genomics, UPM-INIA Madrid, Spain

5 Department of Plant Biotechnology and Bioinformatics, Ghent University, Belgium

6Institute for Biology, Free University Berlin, Germany

7Max Planck Institute for Plant Breeding Research Cologne, Germany

8 Institute for Plant Molecular Biology, CNRS Strasbourg, France

a corresponding author: alexandre.berr@ibmp-cnrs.unistra.fr

Keywords: Chromatin reprogramming, development, enhancers, epigenetics and heritability, nuclear organization, plants

Author contributions:

Each 1-2 authors wrote separate chapters in the manuscript. PM wrote chapter: “In movement: Chromatin remodelling research stays in shape”. All authors revised the manuscript.

See:

Bey T, Jamge S, Klemme S, Komar DN, Le Gall S, **Mikulski P**, Schmidt M, Zicola J, Berr A. Chromatin and epigenetics in all their states: Meeting report of the first conference on Epigenetic and Chromatin Regulation of Plant Traits - January 14 – 15, 2016 - Strasbourg, France. Epigenetics 2016, 11:8, 625-634.

See:

Bey T, Jamge S, Klemme S, Komar DN, Le Gall S, **Mikulski P**, Schmidt M, Zicola J, Berr A. Chromatin and epigenetics in all their states: Meeting report of the first conference on Epigenetic and Chromatin Regulation of Plant Traits - January 14 – 15, 2016 - Strasbourg, France. Epigenetics 2016, 11:8, 625-634.

See:

Bey T, Jamge S, Klemme S, Komar DN, Le Gall S, **Mikulski P**, Schmidt M, Zicola J, Berr A. Chromatin and epigenetics in all their states: Meeting report of the first conference on Epigenetic and Chromatin Regulation of Plant Traits - January 14 – 15, 2016 - Strasbourg, France. Epigenetics 2016, 11:8, 625-634.

See:

Bey T, Jamge S, Klemme S, Komar DN, Le Gall S, **Mikulski P**, Schmidt M, Zicola J, Berr A. Chromatin and epigenetics in all their states: Meeting report of the first conference on Epigenetic and Chromatin Regulation of Plant Traits - January 14 – 15, 2016 - Strasbourg, France. Epigenetics 2016, 11:8, 625-634.

See:

Bey T, Jamge S, Klemme S, Komar DN, Le Gall S, **Mikulski P**, Schmidt M, Zicola J, Berr A. Chromatin and epigenetics in all their states: Meeting report of the first conference on Epigenetic and Chromatin Regulation of Plant Traits - January 14 – 15, 2016 - Strasbourg, France. Epigenetics 2016, 11:8, 625-634.

See:

Bey T, Jamge S, Klemme S, Komar DN, Le Gall S, **Mikulski P**, Schmidt M, Zicola J, Berr A. Chromatin and epigenetics in all their states: Meeting report of the first conference on Epigenetic and Chromatin Regulation of Plant Traits - January 14 – 15, 2016 - Strasbourg, France. Epigenetics 2016, 11:8, 625-634.

See:

Bey T, Jamge S, Klemme S, Komar DN, Le Gall S, **Mikulski P**, Schmidt M, Zicola J, Berr A. Chromatin and epigenetics in all their states: Meeting report of the first conference on Epigenetic and Chromatin Regulation of Plant Traits - January 14 – 15, 2016 - Strasbourg, France. Epigenetics 2016, 11:8, 625-634.

See:

Bey T, Jamge S, Klemme S, Komar DN, Le Gall S, **Mikulski P**, Schmidt M, Zicola J, Berr A. Chromatin and epigenetics in all their states: Meeting report of the first conference on Epigenetic and Chromatin Regulation of Plant Traits - January 14 – 15, 2016 - Strasbourg, France. Epigenetics 2016, 11:8, 625-634.

See:

Bey T, Jamge S, Klemme S, Komar DN, Le Gall S, **Mikulski P**, Schmidt M, Zicola J, Berr A. Chromatin and epigenetics in all their states: Meeting report of the first conference on Epigenetic and Chromatin Regulation of Plant Traits - January 14 – 15, 2016 - Strasbourg, France. Epigenetics 2016, 11:8, 625-634.

See:

Bey T, Jamge S, Klemme S, Komar DN, Le Gall S, **Mikulski P**, Schmidt M, Zicola J, Berr A. Chromatin and epigenetics in all their states: Meeting report of the first conference on Epigenetic and Chromatin Regulation of Plant Traits - January 14 – 15, 2016 - Strasbourg, France. Epigenetics 2016, 11:8, 625-634.