Diversity of Polycomb group complexes in plants: same rules, different players?

Lars Hennig and Maria Derkacheva

Department of Biology and Zurich-Basel Plant Science Center, ETH Zurich, CH-8092 Zurich, Switzerland

Polycomb Group (PcG) proteins form an epigenetic memory system that is conserved in plants and animals and controls gene expression during development. Loss of plant PcG proteins leads to loss of organ identity and to cell overproliferation. Our understanding of plant PcG protein function has recently been advanced by the identification of additional proteins required for transcriptional repression by PcG and by the purification of an Arabidopsis PcG protein complex. These data indicate that Polycomb Repressive Complex 2 (PRC2)-like complexes in animals and plants have to associate with Plant Homeo Domain (PHD)-finger proteins for efficient deposition of histone H3 trimethylated at lysine 27 (H3K27me3) and transcriptional repression. Subsequently, H3K27me3 at target genes assist to recruit additional PcG protein complexes – PRC1 in animals and potentially LIKE HETEROCHROMATIN PROTEIN-1 (LHP1) and the RING finger gene product AtRING1 in plants. A picture is emerging in which the general mechanisms of PcG protein function are well conserved between animals and plants, but in which individual players have been exchanged during evolution.

PcG proteins comprise an evolutionary conserved transcriptional memory system

Polycomb group (PcG) proteins were first identified in Drosophila melanogaster as regulators of homeotic genes during development, but now it is clear that PcG proteins are well conserved among metazoans and plants (for review see Refs. [1,2]). PcG proteins dictate the transcriptional status of target genes and therefore control the choice between alternative development programs. PcG proteins function by forming large protein complexes. In Drosophila, for example, three different PcG protein complexes have been characterized in detail: the Polycomb Repressive Complex 1 (PRC1), the Polycomb Repressive Complex 2 (PRC2), and Pcl-PRC2 (for review see Ref. [3]). PRC2 is comprised of four proteins including the histone methyltransferase Enhancer of Zeste (E(z)). It is likely that PRC2 functions in vivo primarily as a mono- and di-methyl transferase acting on lysine 27 of histone H3 (H3K27), whereas Pcl-PRC2, which contains additional subunits such as Polycomb-like (Pcl), functions as a H3K27 trimethyltransferase. The PRC1 complex also modifies histones – its RING subunit catalyzes histone H2A lysine 119 ubiquitylation (H2AK119u). Therefore, the chromatin of PcG target genes in flies usually carries both H3K27me3 and H2AK119u. Although several plant PRC2-like complexes that control development have been known for many years, plant Pcl-PRC2- and PRC1-like complexes were identified only recently.

Plant PcG proteins are essential for the maintenance of cell fates and normal development. Here we will review the functions of plant PcG proteins in development with a strong focus on the regulation of flowering time and seed development, two traits that are not only of academic but also of great economic interest. We will discuss the current knowledge about PcG protein complexes in plants, and how the identification of a PHD-PRC2 complex, and of candidates for PRC1-like complexes, will advance our understanding of the mechanisms of PcG protein function in plants.

And then there were many – genes for PcG proteins in plants

PcG proteins were first identified in plants in 1997, and it now seems that all extant multicellular plants have a functional PcG system (Figure 1) [4,5]. A particularly impressive example of functional conservation of the PcG system over more than 450 million years of evolution is illustrated by the finding that the Arabidopsis and moss

Glossary

Apomixis: non-sexual seed formation. Apomixis is a very interesting agronomic trait because it offers a way to stabilize high levels of heterozygosity, a prerequisite for hybrid vigor. Whereas Mendelian inheritance of a heterozygous locus (Aa) leads to offspring with all possible allele combinations (AA, Aa and aa), apomictic inheritance leads to offspring with the mother’s genotype (Aa) only.

Gametophyte: the haploid stage of a plant’s life cycle that produces gametes by mitosis. In mosses, the gametophytic phase dominates, whereas it is greatly reduced (usually to 3 to 7 cells) in seed plants.

Maternal effect: the phenotype of the progeny is determined by the genotype of the mother. A gametophytic maternal effect is an effect where the genotype of the female (maternal) gametophyte determines the phenotype of the progeny. Examples of molecular mechanisms that lead to gametophytic maternal effects are (i) the gene is active only after fertilization, but only the maternal allele is expressed (genomic imprinting); (ii) the gene is active before fertilization in the female gametophyte, but the gene product functions only after fertilization (maternal carry over); (iii) the gene is active and the gene product functions before fertilization in the female gametophyte, but the effect on phenotype becomes visible only after fertilization (epigenetic gametophytic effect).

Sporophyte: the diploid stage of a plant’s life cycle that produces spores by meiosis. In mosses, the sporophytic phase is greatly reduced, whereas it dominates in seed plants.
FERTILIZATION INDEPENDENT ENDOSPERM (FIE)

PcG proteins can partially rescue each other’s mutant phenotypes [4]. It is possible that genes for PcG proteins were present in the last common ancestor of plants and animals and were subsequently lost in unicellular lineages. Such gene loss could explain the presence of only a FIE but no CURLY LEAF (CLF) or EMBRYONIC FLOWER 2 (EMF2) homologs in the green algae Chlamydomonas reinhardtii. In addition to the proteins, which are conserved between animals and plants, the plant-specific VEL proteins and LHP1 (Box 1) are also required for the function of the plant PcG system. There is a tendency for the complexity of genes encoding PcG proteins to increase during evolution. For example, the moss and fern genomes mostly have single copies of the genes encoding PcG proteins, whereas seed plants usually have small gene families with up to five members.

The model plant Arabidopsis thaliana has 12 homologs of Drosophila PRC2 subunits: the three E(z) homologs CLF, MEDEA (MEA) and SWINGER (SWN) [6–10]; the three Suppressor of zeste (Su(z)12) homologs EMF2, FERTILISATION INDEPENDENT SEED2 (FIS2) and VERNALIZATION2 (VRN2) [10–12]; the single Extra sex combs (Esc) homolog FERTILIZATION INDEPENDENT ENDOSPERM (FIE) [13]; and the five p55 homologs MSI1–5 [14–18] (Table 1). Genetic and molecular evidence strongly suggests that CLF, MEA and probably SWN catalyze trimethylation of H3K27. First, plant PcG protein target genes carry H3K27me3 that requires the presence of PRC2-like complexes [33,37,55,56,91]. Second, CLF as well as MEA both require an intact SET domain to function [27,91].

LHP1 belongs to the large family of chromodomain (CD) proteins, which also includes HP1 and Pc [79]. In vitro, LHP1 and the LHP1 CD bind with similar affinity to H3K27me3, H3K9me2/3 and H3K9me3K27me3 peptides [47,48]. Because lhp1 mutant plants with a mutation in the CD that abolishes binding to methylated histones lose silencing of PcG target genes, and have the same phenotype as lhp1 null alleles, the binding of LHP1 to methylated H3 tails appears to be essential for its function [87]. Whereas the LHP1 chromodomains mediate binding to methylated histone tails, the carboxy-terminal LHP1 chromo-shadow domain is required for binding to the MADS domain transcription factor SVP that appears to repress target genes such as FT, at least in part, by recruiting LHP1 [51]. Similarly, LHP1 interacts with the transcription factor SCARECROW (SCR) and is recruited to SCR target genes [50]. These data indicate that LHP1 can be recruited to target genes independently of H3K27me3 by direct interaction with transcription factors.
PRC2-like complexes overlap. Do these complexes generally regulate common genes, or is this the exception? Genome-wide binding studies will give a first answer, however, a thorough understanding will require the elucidation of the mechanisms of PcG protein targeting in plants.

In Drosophila, PRC1 and Pcl-PRC2 are recruited to target genes by cis-acting Polycomb response elements (PRE) (for review see Refs. [1,29]). Binding of PRC1 to H3K27me3 via its Polycomb (Pc) subunit may help to recruit PRC1 to targets, but H3K27me3 is not absolutely required for PRC1 binding to target genes (for review see Ref. [3]). In contrast to Pcl-PRC2, which localizes mainly to PREs, the core PRC2 might bind to chromosomes with much less specificity (for review see Ref.[3]). To date, functional PREs have been identified and studied only in Drosophila, and the identification of PREs in plants will be a major contribution. It seems likely that plant PREs will soon become available because the promoter and/or intragenic regions of the three plant PcG target genes FLC, AGAMOUS (AG) and PHERES1 (PHE1) were able to direct appropriately regulated reporter gene expression [30–33].

Genomic data have revealed that the PcG system is ancient and that plants contain many PcG proteins which can form diverse complexes. Even moss has functional PcG proteins, and in seed plants a considerable diversification of genes encoding PcG proteins has occurred. The question then arises: what are the functions of the PcG complexes in plants?

A matter of identity
One central function of PcG proteins in plants and animals is to maintain cell identity, and mutations in these genes often cause homeotic transformations of organs. For example in plants, clf mutants mis-express homeotic genes leading to alterations in flower organ identity; sepals, for example, will adopt a carpel-like fate and can even bear ovules [6]. However, plant PcG proteins not only repress the genetic programs establishing specific cellular identity but also maintain cells in a differentiated state of specific identity. In FIE knock-down plants and clf; swn double mutants, de-differentiated cells accumulate in disorganized, callus-like structures [7,34]. Similarly, loss of moss FIE causes meristems to overproliferate and prevents differentiation of leafy gametophytes [4].

Altogether, it seems that plant PcG proteins are required to maintain organ identity and to keep cells in an appropriately differentiated state. This is in contrast to mammals, where over-expression and not lack of PcG proteins leads to loss of cell differentiation and to over-proliferation, as has been seen in cancer (for review see Refs. [2,35]). This might imply that mechanisms controlling cell-differentiation evolved independently, making different use of the PcG system in animals and plants.

How do PcG proteins influence flowering?
The switch to reproductive development (i.e. flowering) largely determines the reproductive success of plants and is controlled by multiple signaling pathways that integrate environmental and developmental signals (for review see Ref. [36]). It has been known for several years that PcG proteins participate in some signaling pathways to control flowering time, but it is only now becoming clear that PcG proteins control most major regulators of flowering time. In Arabidopsis, at least two PRC2-like complexes are part of this signaling network. First, the EMF complex suppresses precocious flowering and promotes vegetative
development by repressing the transcription of flowering activators such as FLOWERING LOCUS T (FT), the main flowering time integrator, and AGAMOUS-LIKE 19 (AGL19) [12,28,37].

A second PRC2-like complex controlling flowering time is the VRN complex that enables Arabidopsis to flower after vernalization, which in Arabidopsis causes the epigenetic repression of the flowering repressor FLOWERING LOCUS C (FLC) (Box 2). The VRN complex is composed of VRN2, CLF/SWN, FIE and MSI1 [22,23] (Figure 2). VRN2 was found to be required for FLC repression after vernalization [11], and it seemed likely that cold exposure would

**Box 2. Vernalization – reproduction enabled by cold**

Vernalization is the promotion of the competence to flower by long periods of low temperatures such as those typically experienced during winters (for review see Ref. [92]). In temperate climate zones, spring is usually the most suitable season for the onset of reproduction, and many plants from these zones match flowering to the spring season by requiring vernalization and long photoperiods. The photoperiod and vernalization pathways converge to regulate flower integrator genes. Arabidopsis has both winter-annual accessions, which require vernalization to flower, and summer-annual accessions, which do not require vernalization, and there is much natural variation in the extent of vernalization requirement [93]. The major determinants of vernalization requirement in Arabidopsis are the MADS box gene FLOWERING LOCUS C (FLC) and the FLC activator FRIGIDA (FRI). FLC delays flowering, in part by repressing transcription of the potent flowering activator FLOWERING TIME (FT) that is induced by the photoperiod pathway and is part of the mobile flowering signal termed Florigen. Ecotypes with weak or inactive alleles of FRI or FLC flower without vernalization, but ecotypes with active FRI and FLC flower only once vernalization has repressed FLC.

Although FLC is the major gene for vernalization in Arabidopsis, other genes act in parallel to FLC in the vernalization response [92]. VERN3 is required not only for FLC-dependent but also for FLC-independent branches of the vernalization pathway [56], suggesting that PcG mediated gene silencing might be involved in all branches of the Arabidopsis vernalization response.
target the VRN complex to FLC. Surprisingly, however, it was found that association of the core VRN complex with FLC locus, and increased expression of FLC in vrn2 mutants, were both independent of cold exposure [11,23]. In addition, the EMF complex represses FLC under warm conditions [28].

A major breakthrough in the field was the discovery that the VRN complex associates with VERNALIZATION INSENSITIVE 3 (VIN3), VRN5, and VEL1 to form a PHD-PRC2 complex during prolonged cold [22,23]. VIN3, VRN5 and VEL1, together with VEL2 and VEL3, form a small Arabidopsis protein family that is characterized by a PHD finger and a fibronectin III domain [32,38]. The PHD-VRN complex increases H3K27me3 levels in FLC chromatin, leading to stable silencing. Another key finding was the observation that VRN5 and H3K27me3 are initially restricted to a small region from the transcriptional start to the beginning of the first intron, and only spread across the entire FLC locus after return to warm conditions [23,39] (Figure 3). However, neither VRN5 nor H3K27me3 enter the body of the neighbouring genes. Interestingly, this spreading (as well as continuous stable silencing of FLC) takes place only in mitotically active cells, and not in non-dividing cells [39]. Mammalian PRC2 binds to H3K27me3, and it has been suggested that H3K27me3 recruits PRC2 to maintain the mark at sites of DNA replication [40]. It remains to be established whether spreading of H3K27me3 on FLC during cell proliferation relies on the same mechanisms. It has long been known that vernalization requires mitotic cell division [41] and the new results suggest that establishing stability in PcG-mediated epigenetic silencing of target genes such as FLC is the underlying molecular mechanism for this physiological process.

The increase in H3K27me3 at the FLC locus brought about by the PHD-VRN complex is reminiscent of the situation in animals where Pcl-PRC2-like complexes, that also contain PHD-finger proteins, introduce high levels of H3K27me3 into target gene chromatin (for review see Ref. [3]). In vitro studies have shown that the mammalian PHD finger protein PHF1 specifically promotes the ability of Polycomb Group protein EZH2 to catalyze tri-methylation (and not mono- or di-methylation) of H3K27. It will be important to see whether VEL proteins associate also with the FIS and EMF complexes. An association of VEL proteins with the EMF complex is possible because vrn5 mutants have FLC-independent phenotypes that resemble weak alleles of EMF complex mutants. These phenotypes include mild leaf curling, an increase in mean petal number, and early flowering in short-day conditions [32]. Together, the new results of de Lucia and colleagues [23] suggest, in both animals and plants, that association of
PRC2-like complexes with PHD-finger proteins creates the active protein machinery for efficient H3K27me3 deposition and transcriptional repression. Notably, no sequence homology exists outside the PHD domain between the plant and animal PHD-finger proteins that associate with PRC2-like complexes, suggesting that stimulation of PRC2 HMTase activity by PHD-finger proteins independently evolved at least twice.

Silence – just for now or forever?

Initially, it was thought that once PcG protein-mediated silencing is established that it was generally irreversible. However, FT and AGL19 are both temporally repressed by PcG proteins in young Arabidopsis plants, but escape from silencing and lose H3K27me3 in response to day-length and vernalization signals, respectively [28,37]. Therefore, we now know that PcG protein-mediated silencing is often only transient and can be reversed. At least three mechanisms could lead to H3K27me3 clearance from target gene chromatin. Modified H3 could be passively diluted during mitosis or it could be actively exchanged by unmodified H3. There are several replacement histone genes in Arabidopsis [42], and replication-independent removal of H3 has been observed in Arabidopsis [43]. Alternatively, modified H3 could be actively demethylated. Active demethylation of H3K27me3 is found in animals (for review see Ref. [44]), and plant genomes encode homologs of the metazoan JmjC-domain histone demethylases [45]. However, there are no obvious Arabidopsis homologs of the KDM6/JMJJD3 H3K27me3 demethylases, and it will be important to identify plant H3K27me3 demethylases. The extent to which the potential mechanisms contribute to reactivation of PcG protein target genes in plants is currently unclear. Although the mechanisms of H3K27me3 clearance in plants remain elusive, recent results from many laboratories have demonstrated that PcG-mediated silencing is not a one-way road but instead affords a dynamic means of modulating gene expression during development [12,28,37,46–52].

In contrast to most other studied plant PcG protein target genes, which are reactivated at certain stages of the plant life cycle, FLC remains inactive once it is silenced by vernalization. Stable resetting of FLC expression occurs only in heart to torpedo stage embryos [53,54]. Interestingly, in addition to the typical PcG protein chromatin mark H3K27me3, H3K9me2 was also found at FLC after vernalization; both marks required VRN2 and the PHD-VRN complex [55,56]. In Arabidopsis, H3K9me2 is mostly associated with heterochromatic regions [57], and PcG protein target genes generally do not carry this mark [47,58,59]. It is therefore possible that FLC acquired stable heterochromatin-like long-term silencing as a consequence of potentially reversible PcG protein-mediated repression. However, loss of the H3K9 dimethyltransferase KRYPTONITE does not cause increased expression of FLC or late flowering [60]. Clearly, more work is required to clarify the relationship between heterochromatin H3K9me2 gene silencing and PcG protein-mediated repression, but the results on FLC demonstrate that PcG protein-mediated silencing and heterochromatic H3K9me2 can sometimes function together.

PcG proteins and sex

The third PRC2-like complex in plants – the FIS complex – prevents the initiation of endosperm and seed development in the absence of fertilization (for review see Ref. [61]) (Figure 2). The FIS complex consists of four main subunits: MEA/SWN, FIE, FIS2 and MSI1 [19,24,25,62]. The FIS complex functions in the female gametophyte before fertilization, and because this function is needed to control gametophyte and seed development after fertilization, fis mutants are true epigenetic female gametophytic maternal-effect mutants [63]. Initiation of fertilization-independent seed development, such as that observed in fis mutants, is a hallmark of apomictic development. However, PcG proteins might not generally be involved in apomixes, because knockdown of FIE in sexually reproducing Hieracium piloselloides does not lead to autonomous seed development; on the contrary, knock down of FIE in apomictic Hieracium leads to inhibition of autonomous embryo and endosperm formation [64]. Similarly, loss of genes for three PRC2 subunits did not lead to autonomous seed development in rice [65]. Thus, PRC2-like complexes are required for seed development in diverse species and do not simply act to prevent apomixis as initially thought.

The function of the FIS complex after fertilization in the endosperm is less well understood. One direct FIS complex target gene in the endosperm is PHERES1 (PHE1) that encodes a MADS domain transcription factor [33]. PHE1 is paternally imprinted (Box 3); the maternal allele is repressed by the FIS complex [27,66]. Interestingly, FIS2 and MEA are themselves imprinted with only the maternal alleles being expressed in the endosperm [20,67–69]. MEA imprinting is controlled by an autoregulation process through H3K27 trimethylation at the promoter and at the 3′ coding region of the gene [70–72]. By contrast, FIS2 imprinting is independent of PcG proteins but depends on differential DNA methylation [69]. The Arabidopsis FIS-complex subunit genes FIE and MSI1 are not imprinted, but one of the three genes encoding E(z) homologs in maize, and one of the two FIE homologs in maize and rice, are expressed only from the maternal alleles in the endosperm [65,73,74]. It is therefore possible that imprinting of genes for PRC2 subunits is widespread among seed plants.

**Box 3. Genomic imprinting**

Genomic imprinting is an epigenetic phenomenon leading to allele-specific gene expression depending on the parent of origin, and has evolved independently in mammals and flowering plants (for review see Ref. [95]). Parent-of-origin specific expression is generally established through epigenetic modifications, usually involving DNA and/or histone methylation. Several theories have attempted to rationalize the emergence of imprinting. The most prevailing theory proposes that the evolution of imprinting was driven by the conflict of interest between two parents about how many resources mothers should invest into the offspring [96]. According to this theory, genes with expression from the paternal allele promote embryo growth while genes expressed from the maternal allele restrict embryo growth. Consistent with this theory, many imprinted genes act in specialized tissue for the transfer of nutrients from the mother to the embryo, the mammalian placenta and the plant endosperm. Because not all known imprinted genes control off-spring growth, it is likely that other evolutionary forces have contributed to the origin of genomic imprinting.
The hunt for PRC1-like functions in plants
In contrast to PRC2-like complexes, that have been known in plants for several years, the existence of PRC1-like complexes in plants has remained enigmatic, primarily because plant genomes do not encode bona fide Pc homologs. Recent evidence suggests that the plant chromodomain protein LHP1, also known as TERMINAL FLOWER 2, is a major candidate for a Pc-like function in Arabidopsis. It was initially thought that LHP1 functioned in plants, as does HP1 in animals, to silence heterochromatic loci. However, LHP1 (unlike HP1) was usually found in euchromatin and was found to be needed for the silencing of euchromatic genes, including many PcG protein targets, but not for the silencing of genes in heterochromatin [75,76]. LHP1 has a chromodomain that binds H3K27me3 in vitro, and its genome-wide localization overlaps strongly with regions rich in H3K27me3 [47,48]. Thus, LHP1 might fulfill in plants the role that Pc fulfills in animals, namely to bind to the PRC2-depend H3K27me3 mark. However, LHP1 seems not to be required for establishment of chromosome-wide H3K27me3 profiles in Arabidopsis [47], in contrast to animal Pc which may be involved in spreading of H3K27me3 across target loci [1].

The lhp1 mutant has a pleiotropic phenotype, including changes in flowering time, plant architecture, leaf morphology, inflorescence determinacy and hormone levels [46,77–80]. Important LHP1 target genes responsible for the mutant phenotype, are that are also targets for PRC2-like complexes, include FT, FLC and AGAMOUS (AG) [60,80,81]. Nevertheless, lhp1 null mutants have a much milder phenotype than strong alleles of the genes encoding PRC2 subunits, indicating that other proteins act redundantly with LHP1 in plants to fulfil the role of Pc.

Two proteins are known to act downstream of PRC2-like complexes in plants, VRN1 and EMF1, and these have been proposed to be involved in PRC1-like functions [52,60]. VRN1 is a plant-specific DNA-binding protein that contains two B3 domains [82] and is needed for stable repression of FLC; however, VRN1 is not needed for H3K27me3 accumulation at FLC chromatin, suggesting that it acts downstream of the VRN PRC2-like complex [60]. The plant-specific DNA-binding protein EMF1, which is needed for stable repression of AG, binds to silent AG, and this binding depends on the EMF PRC2-like complex [52,83]. In contrast to VRN1, EMF1 is required for normal H3K27me3 levels at target genes and might thus affect
PRC2 function. Although EMF1 is needed for silencing of AG, EMF1 is not needed for silencing of other EMF complex targets including FT [52,84]. The available data suggest that VRN1 and EMF1 are required to lock specific PeG target genes in a silent state; however, neither protein seems essential for the silencing of all PeG target genes. Thus, VRN1 and EMF1 are plant-specific proteins that contribute to the stabilization of PeG protein-mediated silencing.

RING1 is another core component of animal PRC1. RING1 functions as an E3 ubiquitin ligase that mono-ubiquitylates lysine 119 of histone H2A via its RING domain. Although the precise role of this ubiquitylation is unknown, it is required for normal PeG-dependent silencing [3]. Two Arabidopsis RING1 homologs were recently identified [49,85]. AtRING1a and AtRING1b contain both the RING domain and the Ubiquitin-like RAWUL domain, the key characteristics of RING1 in animal PRC1 [85]. These two RING1 homologs, as well as the EMF PRC2-like complex, are required to silence the class I KNOTTED-like targets including FT AG PRC2 function. Although EMF1 is needed for silencing of the key characteristics of RING1 in animal PRC1 [47,58].

Although these results suggest the existence of a PRC1-like complex in plants, this complex might function differently from animal PRC1 because H2AK119 ubiquitination has not so far been detected in Arabidopsis [86]. In mammals and insects, H3K27 trimethylation by Pcl-PRC2, and H2AK119 mono-ubiquitylation by PRC1, are both essential for stable repression. PRC1 PeG proteins can induce nucleosome compaction in vitro, but in vivo transcription initiation and/or elongation are blocked rather than access and binding of the basal transcriptional machinery (for review see Ref. [3]). It is not known whether LHP1 or AtRING1a/b can repress transcription in plants by a similar mechanism. For EMF1, however, interference with transcription has been observed in vitro [52].

The view is now emerging that, in addition to PRC2-like complexes, several other proteins contribute to PeG-mediated silencing in plants. Recent work on LHP1, AtRING1a and AtRING1b indicates that these proteins might form a complex similar in domain composition and function to animal PRC1 [47–49,60,75,76,81,87]. One next step in the field will be to establish how LHP1, AtRING1a and AtRING1b repress transcription and whether they function together with the plant-specific EMF1 and VRN1 proteins.

**To spread or not to spread?**

Genome-wide profiling techniques have revealed that 3–4% of mammalian genes (i.e. 500–700 genes), and 1.3% of the fly genes (i.e. 200 genes), are directly regulated by PeG proteins (for review see Ref. [29]). In flies, PeG binding was usually found in small peaks of ~1 kb that often overlapped with known or presumptive PREs, whereas H3K27me3 covered chromosomal domains of tens or even hundreds of kilobases with depletion of H3K27me3 at the PREs. It was initially suggested that H3K27me3 in mammals does not form the extended domains found in flies, but instead forms small focal peaks at silent gene promoters (for review see Ref. [29]). More recently, however, it was reported that in mouse embryonic fibroblasts H3K27me3 covered about 11% of the analyzed chromosome 17, forming large domains of average size 43 kb but with many domains larger than 100 kb [88]. Differences in the domains detected can be caused by differences in profiling technology and algorithms used for data analysis (for review see Ref. [29]).

Several differences are known in the genome-wide distributions of H3K27me3 in Arabidopsis [47,58], mammals and flies. First, H3K27me3 decorated ~4400 genes (~15% of all genes) in Arabidopsis, considerably more than in animals. It should be noted that the animal experiments were usually performed with single cell types, while the plant experiments were performed with tissues composed of diverse cell types. Because most Arabidopsis genes marked by H3K27me3 have tissue-specific patterns of expression [58], H3K27me3 profiles are often cell-type specific. Thus, the higher number of target genes found in Arabidopsis might not necessarily reflect a more widespread role of PeG protein-based regulation of gene expression in plants. Second, plant H3K27me3 domains were largely restricted to the transcribed regions of single genes and did not extend over the large genomic regions found in flies or recently in mammals. Third, plant H3K27me3 domains were not significantly associated with low-nucleosome density regions. The establishment, spreading, and functional roles of H3K27me3 may thus differ between plants and animals. This idea is supported by the existence of plant-specific members of PeG protein complexes such LHP1 and VEL proteins. Despite these differences, profiling studies in plants, insects and mammals have revealed several similarities. First, H3K27me3 was usually enriched in gene-dense genome regions. Second, H3K27me3 usually did not overlap significantly with heterochromatic H3K9me2 marks in Arabidopsis or H3K9me3 marks in mouse or fly. Third, there was usually a strong negative correlation between H3K27me3 and transcription. These findings are consistent with the idea that PeG protein-dependent H3K27me3 constitutes a mark for transcriptional silencing predominantly of euchromatic gene-dense regions distinct from the heterochromatic gene-poor regions that are rich in H3K9me2 or H3K9me3.

**Concluding remarks**

The PeG system is an ancient and conserved molecular machinery that represses expression of developmental regulators in plants and animals. Seed plants contain many genes encoding PeG proteins and have diverse PeG complexes. Plant PeG proteins, as with their animal counterparts, maintain organ identity. By contrast, cell overproliferation is caused by loss of PeG proteins in plants, whereas it is caused by over-expression of PeG proteins in mammals. New data suggest that PRC2-like complexes must associate with PHD-finger proteins for efficient H3K27me3 deposition and transcriptional repres-
sion. In animals, H3K27me3 at target genes helps to recruit PRC1, whereas in plants the same mark might recruit a PRC1-like LHP1-AtRING1 complex. Thus, the general mechanisms of PcG protein function are well conserved, but individual players have been substituted during evolution. We can expect to see plant PcG protein complex composition firmly established in the near future. A challenging task will be to identify the mechanism(s) of transcriptional repression by PcG proteins in order to fully appreciate why evolution has placed this protein machinery at such prominent control positions in both plant and animal development. Undoubtedly there are exciting times to come!

Acknowledgments
We thank Claudia Köhler and Cristina Madeira Alexandre for helpful comments on the manuscript. Research in the authors’ laboratory is supported by SNF grant 3100AO-116060 and by ETH project ETH-08 02.

References
4 Mosquera, A. et al. (2009) Regulation of stem cell maintenance by the Polycomb protein FIE has been conserved during land plant evolution. Development 136, 2435–2444
19 Köhler, C. et al. (2003) Arabidopsis MS1 is a component of the MEA/FIE Polycomb group complex and required for seed development. EMBO J. 22, 4804–4814
46 Takada, S. and Goto, K. (2003) TERMINAL FLOWER 2, an Arabidopsis homolog of HETEROCHROMATIN PROTEIN 1,
counteracts the activation of FLOWERING LOCUS T by CONSTANS in the vascular tissues of leaves to regulate flowering time. Plant Cell 15, 2856–2865


Cui, H. and Benfey, P.N. (2009) Interplay between SCARECROW, GA and LIKE HETEROCROMATIN PROTEIN 1 in ground tissue patterning in the Arabidopsis root. Plant J. 58, 1016–1027


Bernatavichute, Y.V. et al. (2008) Genome-wide association of histone H3 Lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS ONE 3 (9), e3156


Leroy, O. et al. (2007) Polyclomb group proteins function in the female gametophyte to determine seed development in plants. Development 134, 3638–3648


Hermon, P. et al. (2007) Activation of the imprinted Polyclomb Group FIE1 gene in maize endosperm requires demethylation of the maternal allele. Plant Mol. Biol. 64, 387–395


Nakagai, K. et al. (2005) The Arabidopsis HETEROCHROMATIN PROTEIN 1 homolog (TERMINAL FLOWER2) silences genes within the euchromatic region but not genes positioned in heterochromatin. Plant Cell Physiol. 46, 1747–1756


Kotake, T. et al. (2003) Arabidopsis TERMINAL FLOWER 2 gene encodes a LIKE HETEROCROMATIN PROTEIN 1 homolog and represses both FLOWERING LOCUS T to regulate flowering time and several floral homeotic genes. Plant Cell Physiol. 44, 555–564


Sánchez-Pulido, L. et al. (2008) RAWUL: A new Ubiquitin-like domain in PRC1 Ring finger proteins that unveils putative plant and worm PRC1 orthologs. BMC Genomics 9, 308


Exner, V. et al. (2009) The chromodomain of LIKE HETEROCROMATIN PROTEIN 1 is essential for H3K27me3 binding and function during Arabidopsis development. PLoS ONE 4

Pauler, F.M. et al. (2007) H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone binding pattern on a mouse autosomal chromosome. Genome Res. 19, 221–233


Epigenetics

24. B. Brower-Toland et al., Genes Dev. 21, 2300 (2007).
34. H. Kano et al., Genes Dev. 23, 1303 (2009).
37. T. A. Bell et al., Genetics 172, 411 (2006).
40. V. Grandjean et al., Development 136, 3647 (2009).
43. P. Dunoyer et al., EMBO J. 29, 1699 (2010).
45. We are indebted to J. Brennecke, S. Duharcourt, and A. Péllissot for valuable discussions and advice. D.B. is supported by a European Young Investigator award and the Fondation Schlumberger pour l’Enseignement et la Recherche. D.V. is supported by a European Research Council starting grant (210890) “Frontiers of RNA” and an award from the Bettencourt Foundation.

REVIEW

Epigenetic Reprogramming in Plant and Animal Development

Suhua Feng,1 Steven E. Jacobsen,1,2 Wolf Reik2*

Epigenetic modifications of the genome are generally stable in somatic cells of multicellular organisms. In germ cells and early embryos, however, epigenetic reprogramming occurs on a genome-wide scale, which includes demethylation of DNA and remodeling of histones and their modifications. The mechanisms of genome-wide erasure of DNA methylation, which involve modifications to 5-methylcytosine and DNA repair, are being unraveled. Epigenetic reprogramming has important roles in imprinting, the natural as well as experimental aspects of chromatin. For example, methylation of histone octamers wrapped with DNA—influence the transcriptional state and other functional aspects of chromatin. For example, methylation of DNA and certain residues on the histone H3 N-terminal tail [e.g., H3 lysine 9 (H3K9)] are important for transcriptional gene silencing and the formation of heterochromatin. Such marks are essential for the silencing of nongenic sequences—including transposons, pseudogenes, repetitive sequences, and integrated viruses—that could become deleterious to cells if expressed and hence activated. Epigenetic gene silencing is also important in developmental phenomena such as imprinting in both plants and mammals, as well as in cell differentiation and reprogramming.

DNA methylation occurs in three different nucleotide sequence contexts: CG, CHG, and CHH (where H = C, T, or A). In both mammals and plants, CG methylation is maintained by the maintenance DNA methyltransferase, termed DNMT1 [DNA (cytosine-5)-methyltransferase 1] in mammals and MET1 (DNA METHYLTRANSFERASE 1) in Arabidopsis, and by a cofactor that recognizes hemimethylated DNA at replication foci, called UHRF1 (ubiquitin-like containing PHD and RING finger domains 1) in mammals and VIM (VARIATION IN METHYLATION) family proteins in Arabidopsis (1). In addition, the mammalian de novo DNA methyltransferases DNMT3A and Dnmt3b are required for the maintenance of CG methylation at some loci (2). CHG methylation is common in Arabidopsis and other plant genomes and is maintained by a feedback loop that is formed by a plant-specific DNA methyltransferase, CMT3 (CHROMOMETHYLAISE 3), and a histone methyltransferase, KYP (KRYPTONITE) (3, 4). CHH methylation is also abundant in plants and is maintained by the RNA-directed DNA methylation (RdDM) pathway, which actively targets the DNA methyltransferase DRM2 (DOMAINS

1Howard Hughes Medical Institute and Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA 90095, USA. 2Laboratory of Developmental Genetics and Imprinting, Babraham Institute, Cambridge CB2 2ST, UK, and Centre for Trophoblast Research, University of Cambridge, Cambridge CB2 2EG, UK.

*To whom correspondence should be addressed. E-mail: jacobsen@ucla.edu (S.E.J.); wolfr@bbrc.ac.uk (W.R.)
by the combined action of both histone methyltransferases and histone demethylases, as well as by the proteins that read these histone marks (16).

Through the developmental regulation of these epigenetic mechanisms, both plants and animals undergo epigenetic reprogramming in various cell types and developmental stages, which serve either to transmit epigenetic information between cells or between sexual generations, or to reset epigenetic marks to reduce the risk of perpetuating dangerous epigenetic alleles.

DNA Methylation Throughout the Arabidopsis Life Cycle

To maintain genome integrity from generation to generation, transposons and repetitive DNA elements must be kept under tight regulation in reproductive cells. One of the ways that plants achieve this is through the stable inheritance of DNA methylation. Plants frequently show meiotic inheritance of gene silencing (1). Furthermore, plants are not known to undergo genome-wide waves of demethylation in germ cells, as occurs in animals. However, large-scale reprogramming occurs in non-germ line reproductive cells, and this reprogramming may function to reinforce silencing of transposable elements in germ cells (see below).

One way to actively reprogram the epigenome is to remove methylated cytosines. The Arabidopsis genome encodes four bifunctional helix-hairpin-helix DNA glycosylases and AP lyases—ROS1 (REPRESSOR OF SILENCING 1), DME (DEMETER), DML2 (DEMETER-LIKE 2), and DML3 (DEMETER-LIKE 3)—which recognize and remove methylated cytosines, resulting in a 1-nt gap in the DNA double strand. Subsequently, as yet unidentified DNA repair polymerase and DNA ligase enzymes are thought to fill in the gap with an unmethylated cytosine (1, 17). ROS1, DML2, and DML3 mainly function in vegetative tissues, and genomic studies suggest that they demethylate hundreds of specific loci across the genome with a bias toward genes (1, 18). Knocking out all three genes does not markedly affect the overall levels or patterns of methylation in the Arabidopsis genome (18, 19). Instead, these enzymes appear to be acting as a counterbalance to the RNA-directed DNA methylation system to quantitatively fine-tune methylation levels at particular genomic locations.

By contrast, DME functions to cause global hypomethylation in the endosperm (the extra-embryonic tissue of flowering plants) of Arabidopsis (20, 21), and thus contributes to large-scale epigenetic reprogramming (Fig. 1). In Arabidopsis, female gametogenesis begins when a somatically derived megasporocyte mother cell undergoes meiosis to give rise to a haploid megasporocyte, which subsequently develops into a mature female gametophyte (embryo sac) that contains one egg cell, one central cell (two nuclei), and several other accessory cells. During double fertilization (which is common in plants), the egg cell fuses with a sperm cell from the male gametophyte (pollen grain) to form an embryo, and the central cell fuses with the other sperm cell from pollen to form the triploid endosperm, which nourishes the embryo, and thus bears a function similar to that of the placenta in mammals. DME is expressed primarily in the central cell before fertilization, and thus only the maternal genome is demethylated by DME. This leads to maternal allele-specific gene expression (imprinting) in the endosperm (22). Until recently, only six imprinted Arabidopsis genes were known, but recent genomic studies of endosperm have revealed genome-wide differences in DNA methylation, including a substantial reduction of CG methylation; hence, many additional genes are likely to be imprinted in Arabidopsis, some of which have been verified by single-gene studies (20, 21) (Fig. 1). Demethylation by DME may also reactivate transposon expression, which shunts transposon transcripts into the RNAi pathway, producing additional siRNAs that can guide DNA methylation to non-CG sites whose methylation is high in wild-type endosperm but decreased in dme mutant endosperm (Fig. 1). Curiously, there are even higher levels of non-CG methylation in the wild-type embryo that could be explained by movement of siRNAs produced in the central cell into the egg cell; this attractive idea awaits experimental support (20). Because the endosperm genome does not contribute to the next generation, mild reactivation of transposons in endosperm may not be deleterious and may re-inforce the silencing of transposons in the egg cell and later embryo, contributing to the genome integrity of offspring. Indeed, there is a class of RNA polymerase IV (Pol IV)–dependent siRNAs that only accumulate in flowers and young siliques, likely originating from the endosperm (23). Notably, these siRNAs are derived from maternal alleles only, which suggests that they may be produced in part during female gametogenesis and then retained after karyogamy. However, these siRNAs are expressed more highly after fertilization, and therefore imprinted maternal expression of siRNA loci may also occur as the endosperm develops (23). It is tempting to speculate that the maternal Pol IV–dependent siRNAs are the “messenger” that mediates communication between endosperm and embryo (Fig. 1); however, these siRNAs were detected only in the endosperm, not in the embryo (23). Nonetheless, the possibility that they exist in low abundance in the embryo, or are ephemeral, cannot be ruled out.

The idea that siRNAs move from the endosperm to the embryo is consistent with the model put forth in an earlier study on paternal genome reprogramming in Arabidopsis (24). The male gametophyte of Arabidopsis (a pollen grain) contains two sperm cells, which fertilize the egg cell and central cell, respectively, and a vegetative nucleus (Fig. 1). Transposon expression is generally up-regulated in pollen, and certain transposons even become mobile in pollen, unlike the situation in most other tissues (24). Reduction of transposon methylation followed by transposon reactivation appears to occur in the vegetative nucleus; this is supported by the finding that transposon reactivation and movement are not inherited by the next generation (24). It has been shown that several key RdDM pathway proteins (RDR2 and DCL3) and CHG methylation maintenance pathway proteins (CMT3 and KYP) have reduced expression levels in pollen; in addition, DDM1 (DECREASE IN DNA METHYLATION 1), an important chromatin remodeler required for DNA and histone methylation and transposon silencing, is exclusively localized in sperm cells but not in the vegetative nucleus (24). These results suggest a model in which hypomethylation of the vegetative cell may reactivate transposons that could serve to reinforce transposon silencing in the adjacent sperm cells (Fig. 1) (24).

Small RNAs may be involved in this communication between the vegetative cell and the sperm cells. A class of siRNAs that is 21 nt in length and corresponds to Athila retrotransposons, the largest transposon family, is detected in sperm cells. Because Athila retrotransposons remain silenced in sperm cells but are activated in the vegetative nucleus, it is possible that the 21-nt siRNAs are produced in the vegetative nucleus and then travel to their site of action—sperm cells—where they mediate the silencing of transposons through an unknown mechanism (Fig. 1) (24). A common theme is that both male and female gametophytes contain nurse cells in which massive epigenetic reprogramming may serve to reinforce transposon silencing in the germ line (Fig. 1).

Another example of small RNAs silencing transposons at a distance occurs when the megasporocyte mother cell differentiates from somatic tissues (25). Mutations in AG09 (ARGONAU TE 9), a member of the Arabidopsis Argonaute family of proteins, result in the reactivation of transposons in the ovule (including the egg cell) (Fig. 1). Remarkably, AG09 is not expressed in the reproductive cells themselves (megasporocyte mother cell, megasporocyte, or developing female gametophyte), but is expressed in the companion cells surrounding the female gametophyte. The transposon targets of AG09 are similar to those reactivated in pollen, and evidence suggests that AG09-mediated transposon silencing uses components of known silencing pathways, including the 24-nt RNA-directed DNA methylation pathway (25). Whether the AG09-associated 24-nt siRNAs are the mobile signal remains to be tested.

Resetting of Histone Modifications in Arabidopsis

In addition to DNA methylation, plants also reprogram histones and their associated marks;
Fig. 1. Model of epigenetic silencing dynamics during the Arabidopsis life cycle. In somatic cells, three different mechanisms are responsible for repressing transcription from transposable elements (TEs), DNA methylation (in all three sequence contexts), histone H3K9 dimethylation (H3K9me2), and histone H3K27 monomethylation (H3K27me1). Methyltransferases and proteins regulating these epigenetic marks are shown. See text for details. In the female gametophyte, the central cell is demethylated by DME, which leads to TE activation and up-regulation of RdDM. The siRNAs produced from TEs not only direct non-CG methylation in the central cell, but also might travel to the egg cell and enhance the silencing of TEs there. In addition, AGO9-associated siRNAs produced in somatic companion cells also contribute to the silencing of TEs in the egg cell. In the male gametophyte, the vegetative nucleus does not express DDM1 and has reduced RdDM, which leads to TE activation and mobilization. A new class of 21-nt siRNAs, produced from TEs in the vegetative nucleus, travels to sperm cells to reinforce TE silencing. After double fertilization, maternal TEs in the endosperm stay activated and produce Pol IV-dependent 24-nt siRNAs, which could function to silence the paternal TEs in the endosperm. The methylation levels in the embryo are elevated, possibly as a result of the siRNA signals transmitted from the endosperm. Different shadings indicate the level of DNA methylation (high, black; medium, gray; low, white).
as opposed to DNA methylation (which is typically inherited), some histone modifications are known to be reset in each generation. Because plants do not set aside a germ line early in development (germ cells are differentiated from adult somatic cells), some type of “reprogramming” process is likely needed to erase the effects of epigenetic marks caused by external stimuli (such as development or stress). For example, PGC (Polycomb group) proteins mediate the silencing of FLC (FLOWERING LOCUS C) in Arabidopsis, which controls flowering time (26). In winter-annual accessions of Arabidopsis, FLC is expressed at high levels to repress the initiation of flowering. During vernalization (prolonged exposure to cold, such as during winter), FLC becomes modified by H3K27 trimethylation, which helps to turn off FLC expression epigenetically. When winter passes and temperatures become warmer, trimethylation and silencing of FLC persists, and therefore Arabidopsis can flower in response to environmental cues such as photoperiod. When gametes are formed through meiosis, H3K27 trimethylation marks on FLC are removed by an unknown mechanism and FLC becomes reexpressed in the seeds. Thus, flowering is inhibited by FLC until the next-generation plants encounter cold weather.

Resetting of histone marks may involve, in part, global replacement of histones (27). The histone variant H3.3 can be incorporated in the absence of DNA replication, and thus is a candidate for the “replacement” histone H3 during reprogramming. HTR10 (HISTONE THREE RELATED 10) is exclusively expressed in male plant cells, for reproductive cells undergoing meiosis, and perhaps for early stages of embryo development.


Mechanisms of Epigenetic Reprogramming in Mammalian Development

Genome-wide epigenetic reprogramming occurs in mammalian development at two distinct stages: in primordial germ cells (PGCs) primarily once they have reached the embryonic gonads (embryonic day E10.5 to E13.5), and in the early embryo beginning in the zygote immediately after fertilization and extending to the morula stage of preimplantation development (Fig. 2) (28–30). This reprogramming entails erasure of DNA methylation and loss of histone modifications (as well as loss of histones and histone variants); hence, we focus on demethylation of DNA. The loss of DNA methylation by E13.5 (the developmental endpoint of reprogramming) is truly global; in mouse female PGCs, only 7% of CpGs remain methylated (versus 70 to 80% in embryonic stem (ES) cells and somatic cells), and most promoters and genic, intergenic, and transposon sequences are hypomethylated at this stage (31). The only clear exclusion to global erasure is intracisternal A particles (IAPs), an active family of retrotransposons that have only recently been acquired in the rodent lineage, which only show partial demethylation in PGCs (31). Promoters of germ cell–specific genes (such as Dazl or Vasat) are methylated in early PGCs and become demethylated and expressed during reprogramming (32). Imprinted genes have allele-specific methylation in early PGCs and the Xist promoter is methylated, and this methylation is all erased in PGCs by E13.5 (Fig. 2) (33, 34). Although most of the genome-wide demethylation appears to occur in E11.5 to E13.5 PGCs, it remains possible that some loci become demethylated at slightly earlier stages (35); hence, demethylation is not necessarily coordinated synchronously throughout the genome. Nothing is currently known about the possible occurrence or erasure of non-CG methylation in PGCs.

DNA demainases and the base excision repair pathway have recently been implicated in erasure, which suggests that active demethylation is involved at least in part (31, 36). The cytosine deaminases AID and APOBEC1 are capable in vitro of deaminating 5-methylcytosine (5mC) as well as cytosine and are expressed, albeit at a low level, in PGCs (36, 37). Notably, AID deficiency in PGCs results in a deficit in demethylation of 20% of all CpGs if it is assumed that early PGCs have methylation levels similar to those of ES cells or somatic cells (31). Because of this partial effect of AID deficiency on erasure, the potential redundancy with other DNA demainases needs to be examined. The 5mC hydroxylases TET1 and TET2 are also expressed in PGCs (36), suggesting the possibility that 5mC could be modified by different mechanisms (deamination, hydroxylation) in order to initiate active demethylation. It is also possible that a combination of passive (resulting in hemimethylated substrates in G1 phase of the cell cycle) and active demethylation could be involved. Finally, it is possible that the genome-wide nature of the demethylation process and its relatively coordinate timing require different mechanisms and different modifications of 5mC to join forces in order to achieve such large-scale reprogramming.

Initial modification of 5mC would require further modification or DNA repair in order to achieve demethylation. DNA repair pathways that might be involved in resolving mismatches or in excising 5-hydroxymethylcytosine (5hmC) are nucleotide excision repair, mismatch repair, and especially base excision repair (BER), which is also involved in demethylation during reprogramming in plants (37). BER components such as PARP1, APE1, and XRCC1 are all up-regulated at E11.5 in PGCs, together with enhancement of chromatin-bound XRCC1; thus, it is possible that BER is activated at this time point (Fig. 2) (36).

Global losses of several histone modifications (e.g., H3K27me3, H3K9ac) as well as the linker histone H1 are observed after demethylation of DNA, indicating that widespread DNA repair might be associated with global remodeling of nucleosomes in PGCs (38). It is also possible that specific histone modification or demodification enzymes (deacetylases, demethylases) are in part responsible for erasure of histone marks in PGCs, but none have been identified so far.

Base excision repair also appears to be involved in demethylation in the zygote immediately after fertilization (Fig. 2). The added complication here is that it is specifically the paternal, sperm-derived, genome that is demethylated, whereas the maternal one is not; the maternal genome may be specifically protected from demethylation (39–42). Differentially methylated regions in imprinted genes are also specifically protected from demethylation, and so again are IAPs. Nonetheless, there appear to be substantial losses of methylation in the zygote, potentially of a similar scale to those occurring in PGCs (39, 43, 44). Notably, demethylation of the paternal genome may occur in two phases, one before DNA replication and one associated with the S and G2 phases (44). The first phase might involve modification of 5mC but only partial demethylation (44). Demethylation might then continue at replication or afterward. BER components are also present at these stages with an
Epigenetics

enhancement of chromatin-bound XRCC1 in the paternal pronucleus (36). Both phases show evidence of DNA strand breaks, indicating that repair may be involved in both of them, and inhibition of BER components partly interferes with demethylation (36, 44). Whether AID or TETs are involved in zygotic demethylation is not yet known, but components of the Elongator complex (Elongator complex proteins, ELPs) have been implicated in demethylation of the paternal genome (Fig. 2) (45); Elongator is involved in diverse aspects of transcriptional regulation and can also modify tRNAs. Could Elongator catalyze an as yet unknown modification of 5mC that makes it a substrate for BER? After zygotic demethylation, the embryonic genome continues to be demethylated during the following few cleavage divisions until the blastocyst stage, with DNMT1 protein being largely excluded from the nucleus by an unknown mechanism (Fig. 2). Nonetheless, the maintenance of methylation in differentially methylated regions of imprinted genes does depend on DNMT1 (46), so it will be important to understand how DNMT1 might be targeted during this reprogramming phase to key regions in the genome, such as imprinted genes (47).

Fig. 2. The two major phases of genome-wide erasure of DNA methylation in the early embryo and in primordial germ cells (PGCs) of the mouse. Thickness of the outer arrows indicates levels of DNA demethylation. Red, maternal genome; blue, paternal genome. After fertilization, the paternal genome is more rapidly demethylated than the maternal one. During gametogenesis, de novo methylation in spermatogenesis occurs earlier than in oogenesis. The inner circle shows factors or candidate factors that are implicated in de novo methylation, the maintenance of methylation, and demethylation, respectively. Solid arrows in the inner circle show at what developmental time these epigenetic regulators are thought to act. ES cells, TS cells, and XEN cells are stem cell lines that are derived from the inner cell mass (ICM), trophoderm (TE), and primitive endoderm (PE) of the blastocyst, respectively.

Hence, the current evidence for the initiation and regulation of genome-wide erasure of DNA methylation in PGCs and the zygote points to initiating events that modify 5mC (such as deamination and hydroxylation), which would trigger a BER response. Of course, it is still possible that bifunctional DNA glycosylases of the type that excise 5mC in plants also exist in animals (although none have been found so far); conversely, homologs of APOBEC deaminases and TET and ALKBH-type hydroxylases have yet to be described in plants.

Experimental Reprogramming in Mammals

Experimental reprogramming to a pluripotent state can be achieved, albeit inefficiently, by fusion of somatic cells and pluripotent cells, by cloning, and by direct reprogramming using core transcription factors (48). With all three methods, there is evidence that epigenetic reprogramming is a central component of achieving the goal of an embryonic or ES cell–like (iPS) state. In cell fusion experiments between somatic cells and ES cells, key pluripotency genes such as Oct4 and Nanog need to be demethylated; AID also seems to be important for demethylation in this system (49). Generation of iPS cells from somatic cells by the transduction of core transcription factors (such as OCT4, SOX2, KLF4, and C-MYC) probably requires multiple epigenetic reprogramming steps while the cells that undergo reprogramming divide (50). DNA demethylation is clearly critical because incompletely reprogrammed iPS cells can become completely reprogrammed by treatment with the methylation inhibitor azacytidine (48). Inhibitors of histone deacetylases and histone methyltransferases are also beneficial, showing in general that repressive epigenetic modifications acquired during differentiation and somatic development need to be reversed to achieve the pluripotent state (48). Notably, reprogramming by cloning apparently results in better resetting of the epigenome than can be achieved by direct reprogramming with transcription factors, indicating perhaps that true totipotency requires passage through germ cells or zygotes (51). Direct applications to regenerative medicine will result from unraveling the role of AID, hydroxymethylation, and the TETs, and of the base excision repair pathway as well as the methyltransferases in this process, and from knowledge of how the reprogramming network is connected with the pluripotency network.

Comparative Biology of Epigenetic Reprogramming

Whether genome-scale epigenetic reprogramming has a unified purpose is not clear; some aspects of reprogramming are clearly conserved (or have been reinvented) in animals and plants with their contrasting, although sometimes surprisingly similar, reproductive and biological strategies. In mammals, zygotic reprogramming is broadly conserved, although there may be some differences in timing or extent; by contrast, Xenopus does not appear to show demethylation of the paternal genome (52). Hypomethylation of PGCs is also seen in human and pig fetal development but has not been studied in nonmammalian organisms. Global DNA demethylation in PGCs and paternal demethylation in the
zygote may occur primarily in mammals (and in the central cell in seed plants) that also have imprinting whose mechanism is based on DNA methylation. Clearly, demethylation in PGCs is necessary for erasure of imprints so that new imprints can later be established properly, according to the sex of the germ line (Fig. 2). Plants do not seem to erase imprints; instead, they establish them by demethylation of the maternal genome in the endosperm after fertilization (with the endosperm being comparable to the placenta) (Fig. 1). Perhaps there are as yet undiscovered imprinted genes that acquire parent-specific methylation patterns by (paternal) zygotic demethylation, in analogy to plants.

A second group of genes where demethylation in PGCs seems important are the germ line–specific genes (e.g., Dazl, Vasa) that have specialized functions, for example, in meiosis and germ cell differentiation. These genes are generally demethylated and expressed in germ cells, but in early PGCs they are methylated and silenced. Genes that are demethylated in PGCs include those with a role in transposon control; for example, silences members of the ERVK transposon family (53). Hence, global demethylation, which in principle would lead to transcriptional activation and potentially to transposition of active transposon families, at the same time activates defense mechanisms against transposons that are not needed in somatic cells where transposons are methylated. An extreme view of this scenario is the possibility that demethylated transposons produce small RNAs, which in turn lead to de novo methylation and renewed silencing of transposons (Fig. 1) (24). Although it may sound paradoxical, reprogramming may have an important role in resetting the permanent silencing program for transposons across generations. Also, the fact that AID has a role in demethylation in the early mammalian embryo, selective de novo methylation occurs in ICM cells and their descendants, which is important for the identity and stability of embryonic lineages (28), whereas the placenta remains hypomethylated at the genome-wide level (31). Similarly, genome-wide demethylation in the plant endosperm but not the embryo (20, 21) indicates that epigenetic regulation between the two primary lineages (embryonic, extraembryonic) is fundamentally different, with this difference apparently being conserved—or reinvented—in plants and animals.

References and Notes
34. J. Lee et al., Development 129, 1807 (2002).
44. M. Wossidlo et al., EMBO J. 29, 1877 (2010).
47. X. Li et al., Dev. Cell 15, 547 (2008).
60. S. Ito et al., Nature 466, 1129 (2010).
61. We thank J. A. Law for reading and commenting on the manuscript, and F. Santos for help with figures.
62. S.F. is a Special Fellow of the Leukemia & Lymphoma Society. Supported by NIH grant GM60398 (S.E.J.) and by grants from the UK Biotechnology and Biological Sciences Research Council, UK Medical Research Council, and European Union (W.R.), S.F. is an investigator of the Howard Hughes Medical Institute.