Mechanisms of Development xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Mechanisms of Development

journal homepage: www.elsevier.com/locate/mod

Expansion of the polycomb system and evolution of complexity

Divya Tej Sowpati [1], Senthilkumar Ramamoorthy [1,2], Rakesh K. Mishra *

Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad 500007, India

ARTICLE INFO

Article history:
Received 13 June 2015
Received in revised form 27 July 2015
Accepted 29 July 2015
Available online xxxx

Keywords:
Polycomb
Epigenetic regulation
PRC1
PRC2

ABSTRACT

Polycomb group (PcG) proteins regulate and maintain expression pattern of genes set early during development. Although originally isolated as regulators of homeotic genes, PcG members play a key role in epigenetic mechanisms that maintain the expression state of a large number of genes. All members of the two polycomb repressive complexes (PRC1 and PRC2) are conserved during evolution and while invertebrates generally have one gene for each of these, vertebrates have multiple homologues of them. It remains unclear, however, if different vertebrate PcG homologues have distinct or overlapping functions. We have identified and compared the sequence of PcG homologues in various organisms to analyze similarities and differences that shaped the evolutionary history of these proteins. Comparative analysis of the sequences led to the identification of several novel and signature motifs in the vertebrate homologues of these proteins, which can be directly used to pick respective homologues. Our analysis shows that PcG is an ancient gene group dating back to pre-bilaterian origin that has not only been conserved but also expanded during the evolution of complexity. The presence of unique motifs in each paralogue and its conservation for more than 500 Ma indicates their functional relevance and probable unique role. Although this does not rule out completely any overlapping function, our finding that these homologues only minimally overlap in their nuclear localization suggests that each PcG homologue has distinct function. We further propose distinct complex formation by the PcG members. Taken together, our studies suggest non-redundant and specific role of multiple homologues of PcG proteins in vertebrates and indicate major expansion event preceded by emergence of vertebrates that contributed as enhanced epigenetic resource to the evolution of complexity.

1. Introduction

A complex network of signaling, maternally deposited RNA and proteins, and a cascade of regulatory events set the expression pattern of genes early during development to affect the process of embryogenesis. Expression pattern of hox genes, which determine the anterior–posterior body axis of the embryo, is also set by a number of early expressing genes. Once the pattern of expression is set, however, a completely different set of factors kicks in to maintain this pattern. For instance, the combination and level of expression of hox genes determine the identity of a cell, and this identity is preserved for normal development. Polycomb group (PcG) and trithorax group (trxG) of genes contribute to the repressive and active states, respectively.

In *Drosophila*, PcG proteins are associated with H3K27Me3 and trxG mediated activation is associated with H3K4Me3 mark (Simon and Tamkun, 2002). PcG proteins function as multi-protein complexes, polycomb repressive complexes 1 and 2 (PRC1 and PRC2). PRC2 mainly consists of ESC, E(Z), CAF1 and SUZ12 (Kuzmichev et al., 2002; Tie et al., 2001), in which, E(Z) contains the SET domain which has enzymatic activity of tri-methylation of H3K27 (Cao and Zhang, 2004). The methylation mark set by PRC2 is subsequently recognized by PRC1 to establish and maintain the repressed state. PRC1 consists of PC, PSC, PH and SCE (Shao et al., 1999). The chromodomain of PC binds to H3K27me3 marks (Fischle et al., 2003). Furthermore, in fly, glycosylation of PH is shown to be essential for the PcG mediated repression (Gambetta et al., 2009) and, likewise, SCE is involved in ubiquitylation of H2A (Wang et al., 2004), which brings additional functionality to the polycomb system.

The PcG and trxG proteins have been best characterized in *Drosophila*. In vertebrates, multiple homologues of these proteins have been identified, which reflects the enhanced complexity of the PcG system (Fig. 1). Though the vertebrate homologues of PRC1 members were identified based on sequence homology, the functional roles of the multiple homologues are poorly understood. The *Drosophila Psc* has a paralogue, *Su(z)2*, while vertebrates have six paralogues, viz., *Pcgf1*, *Pcgf2*, *Pcgf3*, *Pcgf4*, *Pcgf5* and *Pcgf6*. Among them, *Pcgf2* (*Mel-18*) and *Pcgf4* (*Bmi1*), homologues of fly *Psc* and *Su(z)2*, respectively, are better studied. Though the Mel18 and Bmi1 deficient mice exhibit almost similar posterior transformation and axial skeletal defects (Akasaka et al., 1996; van der Lugt et al., 1994), cerebellum abnormality in *Bmi1* mutant and intestinal obstruction in *Mel18*

* Corresponding author.
E-mail address: mishra@ccmb.res.in (R.K. Mishra).
[1] These authors contributed equally.
[2] Present address: Max Planck Institute of Immunobiology and Epigenetics, Stübeweg 51, 79108 Freiburg, Germany.

http://dx.doi.org/10.1016/j.mod.2015.07.013
0925-4773/

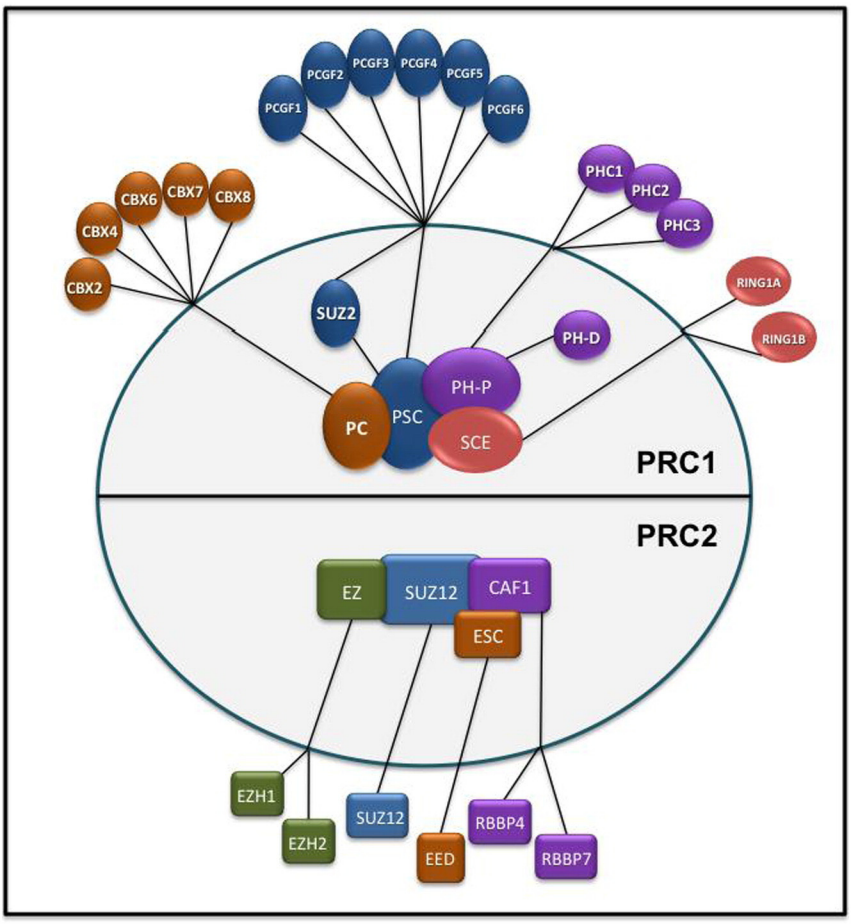

Fig. 1. Expansion of polycomb repressive complexes in vertebrates. Members of PRC1 are represented as ellipses, and PRC2 members are represented as rounded rectangles. The *Drosophila* PRC1 and PRC2 members are shown inside the large circle, and the corresponding vertebrate homologues are shown as branches outside the circle.

mutant (Akasaka et al., 1996) are among their distinct phenotypes. Similarly, *Drosophila Sce*, a.k.a. *dRing*, has two homologues, *Ring1A* and *Ring1B* in vertebrates, and mutants of these genes in mice exhibit distinct phenotypes. Mice lacking one copy of *Ring1A* exhibit homeotic transformations and skeletal defects (del Mar Lorente et al., 2000; Voncken et al., 2003), while different *Ring1B* mutant mice show gastrulation defects, posterior transformation of axio skeleton (Suzuki et al., 2002), and die at E10.5 (Voncken et al., 2003) This indicates a non-redundant function for the *Ring1A* and *Ring1B* paralogues in mice. *Ph*, a core PRC1 member, is duplicated in fly and is known as *Ph-proximal* (*Ph-p*) and *Ph-distal* (*Ph-d*). In vertebrates, *Ph* has 3 homologues, *Phc1*, *Phc2* and *Phc3*. *Phc1* and *Phc2* mutant mice show similar skeletal defects and both interact with the same set of PcG members (Isono et al., 2005; Takihara et al., 1997), although several phenotypic defects are not identical for the two paralogues (Isono et al., 2005). The best studied member of PcG, *Pc*, has at least five paralogues *Cbx2*, *Cbx4*, *Cbx6*, *Cbx7* and *Cbx8* in vertebrates. It has been shown that *Cbx2* deficient mice have defective cellular proliferation and severe homeotic transformation, viz., defective skeleton, limb and sternal malformations, and lethality within six weeks (Core et al., 1997). In a different study, a truncated *Cbx2* showed male to female sex reversals (Katoh-Fukui et al., 1998). Our earlier study indicated that each polycomb paralogue has its own signature motif and has diverged early during the course of evolution (Senthilkumar and Mishra, 2009). Taken together, these observations indicate the presence of multiple PRC1 in vertebrates.

Specialized role for different paralogues of PcG members in vertebrates is also indicated by a number of studies. CBX4 interacts with SUV39H1 histone lysine methyl transferase, and the SET domain of SUV39H1 is essential for the localization of CBX4 on H3K9 methylation mark (Sewalt et al., 2002), and mouse CBX8 interacts with a transcriptional activator AF9 while CBX2 does not interact with these proteins (Hemenway et al., 2001). C-terminal binding protein (CTBP) is a transcriptional repressor that binds specifically to CBX4 but not CBX2 (Sewalt et al., 1999). SUMOylation is essential for CBX4 mediated repression, where it functions as E3 SUMO ligase and is involved in SUMOylation of CTBP, SIP1, HIPK2, CTCF and DNMT3A (Li et al., 2007; Long et al., 2005; MacPherson et al., 2009; Roscic et al., 2006). De-SUMOylation activity of CENP2 establishes the removal of CBX4 and subsequent de-repression of the locus. *Cenp2* null mutant shows accumulation of SUMOylated CBX4 at *Gata4* and *Gata6* loci leading to their repression (Kang et al., 2010). The SUMO mediated regulation is specifically seen in CBX4. In addition, previous studies have shown that various CBX paralogues show differential association with chromatin, and their nuclear localization changes with differentiation (Ren et al., 2008; Vincenz and Kerppola, 2008). These findings point to diversification and conservation of distinct functions of expanded polycomb system.

In the present study, we analyzed the homologues of vertebrate counterparts of PRC1 and PRC2 genes and identified sequence motifs unique to different homologues that were acquired early during vertebrate evolution and have been conserved. We also show that different paralogues are functionally distinct as they tend to localize in non-overlapping manner in the nucleus. These observations suggest the crucial role of epigenetic regulators in the evolution of complexity.

2. Results

Polycomb system consists two major kinds of complexes, PRC1 and PRC2. Pc is one of the components of PRC1 (Fig. 1). In an earlier study, we identified and compared homologues of Pc gene across the species

Please cite this article as: Sowpati, D.T., et al., Expansion of the polycomb system and evolution of complexity, Mechanisms of Development (2015), http://dx.doi.org/10.1016/j.mod.2015.07.013