Int. J. Dev. Biol. 64: 151-158 (2020)
https://doi.org/10.1387/ijdb.190167rm

THE INTERNATIONAL JOURNAL OF DEVELOPMENTAL BIOLOGY
www.intjdevbiol.com

Lessons on gene regulation learnt from the *Drosophila melanogaster* bithorax complex

ARUMUGAM SRINIVASAN and RAKESH K. MISHRA*

Centre for Cellular and Molecular Biology, Council for Scientific and Industrial Research, Hyderabad, India

ABSTRACT Homeotic or *Hox* genes determine the anterior-posterior body axis in all bilaterians. As expected, *Hox* genes are highly conserved across bilaterians. Interestingly, however, the peculiar organization of *Hox* genes in the form of clusters where the order of occurrence of genes in the genome corresponds to the order in which they regulate segmental identity of anterior-posterior body axis is also conserved. The relation between collinearity of arrangement of genes on the chromosomes and spatial function along the body axis has attracted attention to exploring its relevance in the precise regulation of *Hox* genes. Conservation of genes and their arrangement suggests a linkage between co-regulation and the higher order chromatin organization of the entire complex. To this end, we and others have used *Drosophila* as the model system to understand the cis-and trans-regulatory components of *Hox* genes. A number of chromatin-level regulatory elements, like chromatin domain boundaries, and Polycomb Response Elements (PREs) have been discovered in this process. Interestingly, much of what has emerged from the study of homeotic genes, the cis-elements and protein factors, have relevance across the genome in a large number of regulatory events beyond the *Hox* genes. Here, we review our findings and discuss their genome-wide implications in complex regulatory processes.

KEY WORDS: *bithorax complex, chromatin domain boundary, polycomb response element, GAF*

Introduction

The Homeotic genes determine Anterior-Posterior body axis in *Drosophila melanogaster* and, as it turns out, in all other bilaterians (Ferrier and Holland, 2001). Earlier work based on exhaustive genetic analysis showed that homeotic genes exist in a cluster and, more intriguingly, the order of their occurrence in the genome corresponds to their expression pattern along the anterior-posterior body axis. Thus, the order in which they determine the corresponding segments of the anterior-posterior (A-P) body axis, a phenomenon known as collinearity of organization and functions (Lewis, 1978, Lewis, 1998, Maeda and Karch, 2006, Mishra, 2004). Subsequent molecular analyses showed that the collinearity discovered in *Drosophila* is conserved, as are the homeotic genes themselves, all the way up to mammals (Duboule, 1998, Gaunt, 2015). The precise basis of this peculiar collinearity of gene organization and functions remains to be understood. One aspect of the puzzle that is better understood is the general view that the complex organization of regulatory elements is linked to the interdependent interplay of the regulatory network (Duboule, 1998). Once arrived at, during the evolutionary process, such an arrangement was retained. Therefore, it can be said that there is only one way of forming the anterior-posterior body axis in animals. A lot needs to be done, however, to understand the precise sequence and the dynamic nature of events that involve long-range interactions leading to the precise pattern and level of expression of each Hox gene along the A-P body axis. New technologies including high throughput platforms of genomics and proteomics along with genome editing approaches offer exciting new opportunities to address these knowledge gaps (Martin *et al.*, 2016, Rezsohazy *et al.*, 2015). The large scale availability of genomic sequences also opens up the possibility of exploring molecular mechanisms of this complex regulation process beyond the few model organisms and, thereby, offers a more comprehensive understanding of the evolutionary developmental, evo-devo, biology (i, 2013, Lewin *et al.*, 2018).

Abbreviations used in this paper: ANT-C, antennapedia complex; A-P, anterior-posterior; BX-C bithorax complex; GAF, GAGA associated factor; PRE, polycomb response element.

***Address correspondence to:** Rakesh K. Mishra. Centre for Cellular and Molecular Biology, Council for Scientific and Industrial Research, Uppal Road, Hyderabad, 500007, India. E-mail: mishra@ccmb.res.in - https://orcid.org/0000-0001-6636-7380

Submitted: 5 July, 2019; *Accepted:* 19 August, 2019.

ISSN: Online 1696-3547, Print 0214-6282

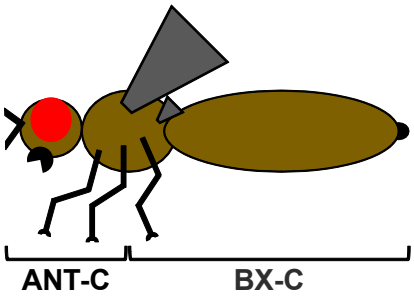

Fig. 1. Homeotic genes and regulation of anterior-posterior body axis. *In* Drosophila melanogaster, *the Hox genes are clustered in two complexes, antennapedia complex (ANT-C) and bithorax complex (BX-C), which determine the identity to the body segments in anterior and posterior parts, respectively.*

Drosophila bithorax complex, a model system of choice to study chromatin level regulatory elements

Homeotic genes in *Drosophila* are organized in two complexes, Antennapedia Complex (ANT-C) and Bithorax Complex (BX-C). The genes in ANT-C dictate the anterior part of the body segments starting from the head to the second thoracic segment, and the rest of the segments, comprising the posterior part of the body from third thoracic segment to the entire abdomen, are determined by the BX-C genes (Fig. 1) (Kaufman *et al.*, 1990, Lewis, 1978, Sanchez-Herrero *et al.*, 1985). The *Drosophila* body has large and distinct segments with unique features which helped geneticists to uncover the regulatory mutations involved in developmental gene regulation. In fact, it was *Drosophila melanogaster* and mutations in BX-C that revealed a number of regulatory elements for the first time (Mihaly *et al.*, 1998a, Simon *et al.*, 1993). Subsequent studies showed that aspects discovered from studies on the BX-C of *Drosophila* are not only operational at the genome scale in the fly but are also conserved across evolution.

Detailed analyses on BX-C elucidated the presence of multiple regulatory domains within the BX-C and the involvement of several cis-regulatory elements like initiator elements, enhancers, chromatin domain boundaries, Polycomb Response Elements (PREs), Promoter Targeting Sequences (PTS) and Promoter Tethering Element (PTE) (Akbari *et al.*, 2008, Karch *et al.*, 1994, Mihaly *et al.*, 2006, Simon *et al.*, 1993, Zhou and Levine, 1999). Further, our sequence comparison and molecular genetic studies on regulatory elements of BX-C highlighted the role of conserved DNA motifs, that provide recruitment sites for trans-acting protein factors (Mihaly *et al.*, 1998b, Schweinsberg *et al.*, 2004)).

Collinear organization of the cis-regulatory domains of the bithorax complex

The collinearity discovered in *Drosophila*, turns out to be more peculiar than that seen in vertebrates. While in vertebrates it is the order of homeotic genes that are collinear to their order of expression and function along the body axis, in *Drosophila*, in addition to the homeotic genes- *Ubx*, *abd-A* and *Abd-B,* the order of *cis*-regulatory domains is also collinear to their function, which is to regulate the homeotic genes along the A-P body axis (Duboule, 1998, Mihaly *et al.*, 1998a). Several genetically well-characterized regulatory domains, viz., *bx, bxd, iab-2 to iab-7*, (Maeda and Karch, 2006) regulate the expression of the homeotic genes in a way that determines the identity of abdominal segments A1 to A7 in the adult abdomen (Fig. 2).

While these regulatory domains contain multiple elements, like, initiator elements, enhancers, etc., for the appropriate combination and expression levels of the homeotic genes unique to each segment, the two elements that have emerged to play a key role in this process, are the boundary elements and PREs. Boundary elements define the extent of the *cis*-regulatory domains and prevent cross-talk across such regulatory elements. PREs, on the other hand, serve as maintenance elements that maintain the expression pattern, including the level of expression of homeotic

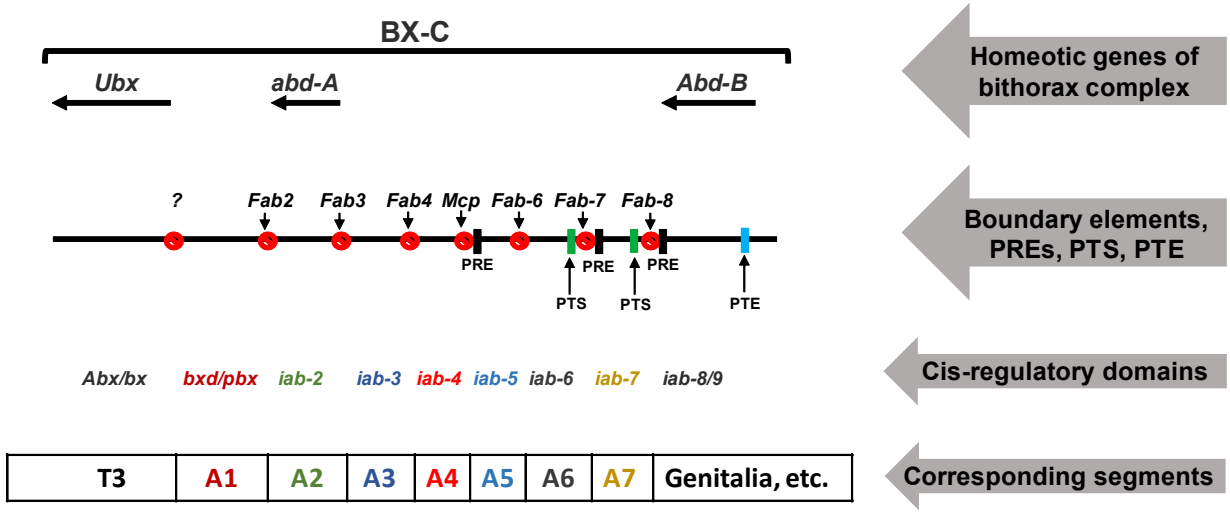

Fig. 2.The bithorax complex (BX-C) segmental identity from the third thoracic to the posterior end of the body. *Three homeotic genes,* Ubx, abd-A *and* Abd-B, *exist in an order collinear to their expression pattern along the anterior-posterior body axis and so are the nine cis-regulatory domains that determine the identity of corresponding body segments. Cis-regulatory domains are defined by the boundary elements. The representative boundaries, PREs, PTS, PTE are shown as red circle-backslash symbols, black, green and blue vertical lines respectively (not to scale).*

genes. Accordingly, mutations in boundaries and PREs lead to homeotic phenotypes as they lead to misregulation of homeotic genes. An interesting problem that remains unresolved in this context is - how a *cis*-regulatory domain that is restricted by the boundary elements from interacting with its neighbours is able to jump across several boundaries to interact with the promoter of homeotic genes in a segment specific manner (Fig. 2). One possible explanation is via coordinated long-range interactions facilitated by the boundary, PRE or associated regulatory elements like Promoter Targeting Sequences (PTS) and Promoter Tethering Element (PTE) that mediate this process (Akbari *et al.*, 2008, Ho *et al.*, 2011, Postika *et al.*, 2018). PTS helps enhancers to bypass the boundaries adjacent domains and access the target promoter (Abd-B, in BX-C). There are two PTS mapped in BX-C, one in *iab-6* domain and the other in *iab-7* domain, both are placed in close proximity to their respective boundaries *Fab-7* and *Fab-8* (Chen *et al.*, 2005, Zhou and Levine, 1999). A Promoter Targeting Element that is located in the upstream region of Abd-B promoter, tethers the *iab-5* domain to the Abd-B promoter region and facilitates the enhancer promoter interaction by avoiding intervening boundary elements (Akbari *et al.*, 2008, Ho *et al.*, 2011).

Another interesting but unexplained factor of their regulatory elements is the proximity or overlapping arrangements of boundaries and PREs in the BX-C (Fig. 2). It is likely that this arrangement holds some critical aspects of the homeotic genes that remain to be discovered. Although not quite directly, this observation points to a possible higher order mechanism on the chromatin level in this process. The PREs recruit Polycomb proteins and form protein complexes that can modify the local chromatin status by leaving specific post-translational modifications on histone H3, for example, H3K27me3. The marked histones carry their chromatin status to subsequent generations and maintain the transcriptional memory. Here the recruited PcG proteins or silent histone marks (like H3K27me3) can spread beyond PREs and affect the nearby regions (genes). To prevent this spreading of silenced marks and also to define limits of the autonomous domain of gene expression, chromatin domain boundaries are required to be placed nearby and this arrangement is seen in various *iab* domains of the bithorax complex (Brown *et al.*, 2003). Considering the crucial role the boundaries and PREs play in the bithorax complex and their utility elsewhere in the genome, their molecular nature and identification at genome level have attracted attention.

Boundaries and Polycomb Response Elements are clusters of short DNA sequence motifs

Functional significance of chromatin-level regulatory elements like boundaries and PREs indicated that such elements must exist

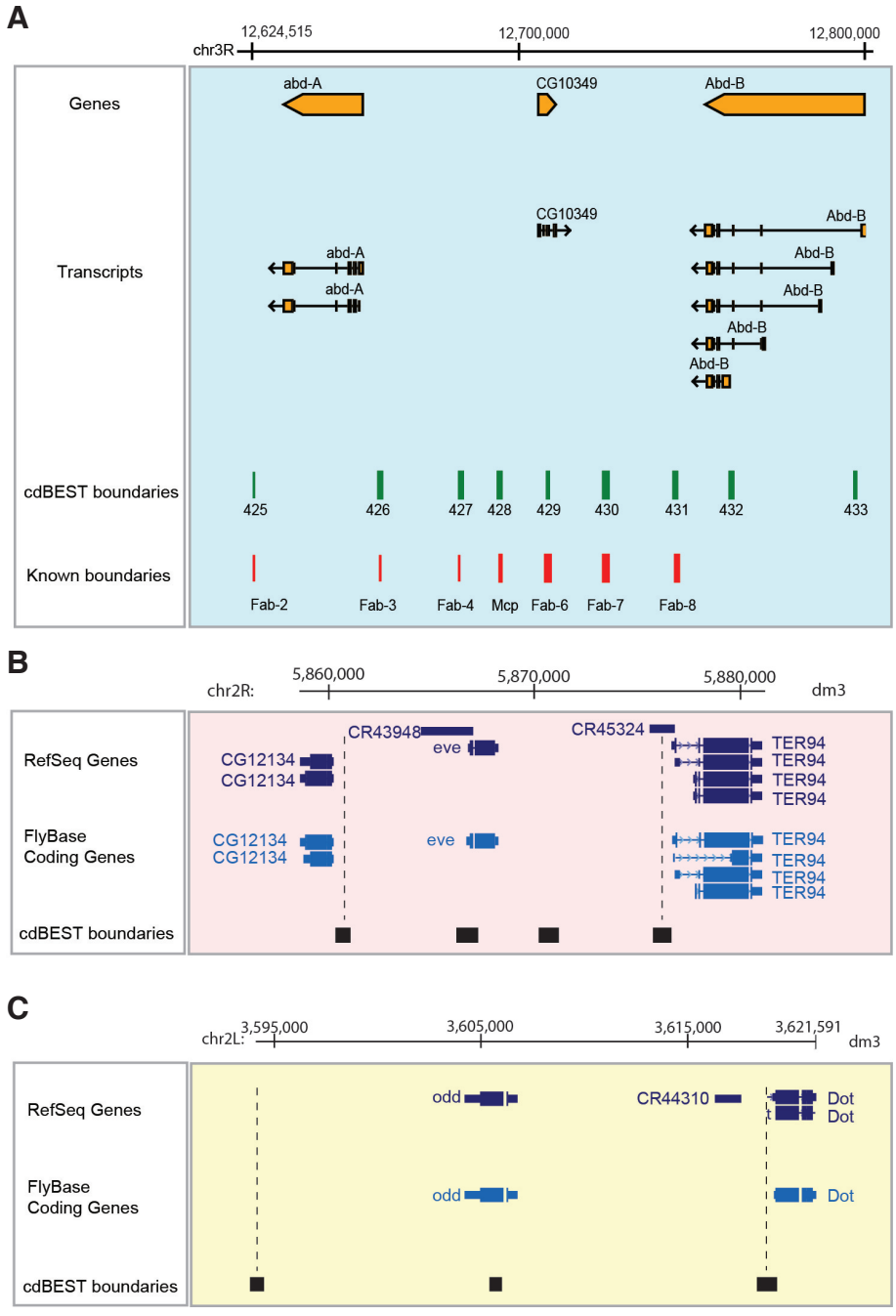

Fig. 3. The cdBEST tool identified boundaries in BX-C region and pair-rule gene loci. (A) *The* Hox *genes, abd-A and Abd-B and their annotated transcripts are drawn according to scale. The cdBEST boundaries are shown as green bars, with their boundaries IDs (Srinivasan and Mishra, 2012) corresponding to chromosome 3R (Release 5). The red bars represent known boundaries.* **(B,C)** *The pair-rule genes,* even-skipped *(*eve*) and* odd-skipped *(*odd*) have insulted chromatin domains. The UCSC genome browser (dm3) views show cdBEST boundaries (black boxes) and genes. The dashed intersecting vertical lines on the image show the borders of insulated chromatin domain.*

TABLE 1

SHORT DNA MOTIFS DEFINE BOUNDARIES AND POLYCOMB RESPONSE ELEMENTS (PREs)

A. DNA MOTIFS PRESENT IN BOUNDARIES

S. No	Motif Sequence (IUPAC code)	Binding Protein	Associated BX-C Boundary*	Reference
1	GAGAG	GAF	*Fab-2, Fab-3, Mcp, Fab-7, Fab-8*	(Schweinsberg *et al.*, 2004)
2	MCAATAAG	Elba	*Fab-3, Fab-4, Fab-7, Fab-8*	(Aoki *et al.*, 2008)
3	MHRGRKGKCGCY, YAGRKGKCGC, RRCGCCMYCYRKY	CTCF	*Fab-2, Fab-3, Fab-4, Mcp, Fab-6, Fab-8*	(Bartkuhn *et al.*, 2009, Holohan *et al.*, 2007, Smith *et al.,* 2009)
4	CGATA	BEAF	*Fab-3, Fab-4, Mcp, Fab-8*	(Cuvier *et al.*, 1998)
5	GCTGMG	Zw5	*Fab-3, Fab-4, Fab-6, Fab-7*	(Gaszner *et al.*, 1999)
6	YRYTGCATAYYY, YWGCMTACTTHY	Su(Hw)	*Fab-2*	(Parnell *et al.,* 2003, Ramos *et al.,* 2006)

*based on the physical motif presence on considered DNA sequence

B. DNA MOTIFS PRESENT IN PREs

S. No.	Motif Sequence (IUPAC DNA code)	Binding Protein	Associated BX-C PREs*	Reference
1	GAGAG	GAF	*bxd,lab-2, Mcp, iab-6, iab-7, iab-8*	(Strutt *et al.,* 1997)
2	CNGCCATNDNND	Pho	*bxd,lab-2, Mcp, iab-6, iab-7, iab-8*	(Mihaly *et al.*, 1998b)
3	GAAAA	DSP1	*bxd,lab-2, Mcp, iab-6, iab-7, iab-8*	(Dejardin *et al.,* 2005)
4	YGAGYG	Zeste	*bxd,lab-2, Mcp, iab-6, iab-7, iab-8*	(Ringrose *et al.*, 2003)
5	RRGGYG	Sp1/KLF	*bxd,lab-2, Mcp, iab-6, iab-7, iab-8*	(Brown *et al.,* 2005)
6	TGTTTTTT, WCHGGTT	Grh	*bxd , lab-2, Mcp*	(Blastyak *et al.*, 2006, Kassis and Brown, 2013)
7	GTGTGT	Cg	*lab-2, Mcp, iab-8*	(Okulski *et al.*, 2011, Ray *et al.,* 2016)
8	KCRRCRGCRRCR	ADF1	*Mcp*	(Orsi *et al.*, 2014)

*based on the physical motif presence on considered DNA sequence

across the genome, although simple searches based on sequence comparisons did not lead to any success. It, therefore, remained a challenge to locate boundaries and PREs in genomes. Biochemical studies from our laboratory (Mishra *et al.*, 2001, Schweinsberg *et al.*, 2004) and several others (Aoki *et al.*, 2008, Bartkuhn *et al.*, 2009, Blastyak *et al.*, 2006, Brown *et al.*, 2005, Cuvier *et al.*, 1998, Dejardin *et al.*, 2005, Gaszner *et al.*, 1999, Holohan *et al.*, 2007, Kassis and Brown, 2013, Mihaly *et al.*, 1998b, Okulski *et al.*, 2011, Orsi *et al.*, 2014, Parnell *et al.*, 2003, Ramos *et al.*, 2006, Ray *et al.*, 2016, Ringrose *et al.*, 2003, Smith *et al.*, 2009, Strutt *et al.*, 1997), revealed clusters short DNA sequence motifs in boundaries and PREs (Table 1). These studies indicated that a cluster of short sequence motifs that provide a site of interaction for sequence-specific DNA binding proteins might be providing the functional input for such elements.

Chromatin domain Boundary Element Search Tool, cdBEST

Using short sequence motifs and their occurrence patterns on boundary element sequences, we developed a boundary search tool, cdBEST, to look for boundary elements in *Drosophila* (Srinivasan and Mishra, 2012). We used the *Drosophila* bithorax complex as a standard region to look for boundaries. The tool identified 12 boundaries in the whole bithorax complex and included all the known boundaries, like Fab-7, Fab-8, etc (Fig. 3). As would be expected, the predicted boundary elements separate genes into distinct domains for the autonomy of their regulation and prevent them from being influenced by nearby regulatory elements outside the domain, a large number of boundary elements are expected to be present in the genome. With the cdBEST, we could identify 4576 boundary elements across the *Drosophila melanogaster* genome (Srinivasan and Mishra, 2012). Apart from bithorax complex, such boundaries are also present in several other loci including the pair-rule genes, *eve* and *odd* where the presence of boundary is necessary for their highly restricted pattern of expression and to maintain autonomous domain of gene expression, Fig. 3. Other than *Drosophila* species, we could identify boundaries, using cdBEST in the mosquito Hox complex and validate them using trans-gene based approaches in *Drosophila* (Ahanger *et al.,* 2013). Applicability of such tools for vertebrate genomes, however, remains elusive as necessary biochemical and genetic data are lacking.

Genome-wide identification of Polycomb Response Elements in Drosophila

Taking clues from the PREs of bithorax complex, and other regions of *Drosophila* genome, the efforts were on to map the PREs across the *Drosophila* genome using bioinformatics approaches (Fiedler and Rehmsmeier, 2006, Ringrose *et al.*, 2003, Zeng *et al.*, 2012). Several approaches have been used that led to identification of ~700 PREs in the *Drosophila* genome (Fiedler and Rehmsmeier, 2006, Schwartz *et al.*, 2006, Zeng *et al.*, 2012). However, these available methods miss out a large number of potential PREs as they could only show 16% overlap with *in vivo* PcG binding sites (Kassis and Brown, 2013, Schuettengruber *et al.*, 2009). To address this deficiency, we developed an exhaustive PRE search tool, PRE Mapper (Srinivasan and Mishra, 2020). In addition to the motifs used on previous methods, we also included additional PRE motifs such as Grh (Blastyak *et al.*, 2006, Kassis and Brown, 2013), Adf1 (Orsi *et al.*, 2014) and Cg (Okulski *et al.*, 2011, Ray *et al.*, 2016) that are recently shown to contribute for PRE function. PRE Mapper identified >20000 PREs in the *Drosophila melanogaster* genome (Srinivasan and Mishra, 2020). A sample region extracted from PRE mapper data is shown in Fig. 4, along with a repressive histone modification and PcG binding profiles (Schwartz *et al.*, 2006). As evident from the figure, the mapped PREs coincide with PcG bound regions. Additionally, these mapped PREs also encompass DNaseI hypersensitive sites which is in agreement with the suggestion that there is a global association of such hypersensitive sites with a variety of regulatory elements, including PREs (Thomas *et al.*, 2011).

These results on the genome-wide search of boundaries and PREs support the view that the cis-regulatory elements identified on the *Drosophila* bithorax complex are relevant in other loci as well

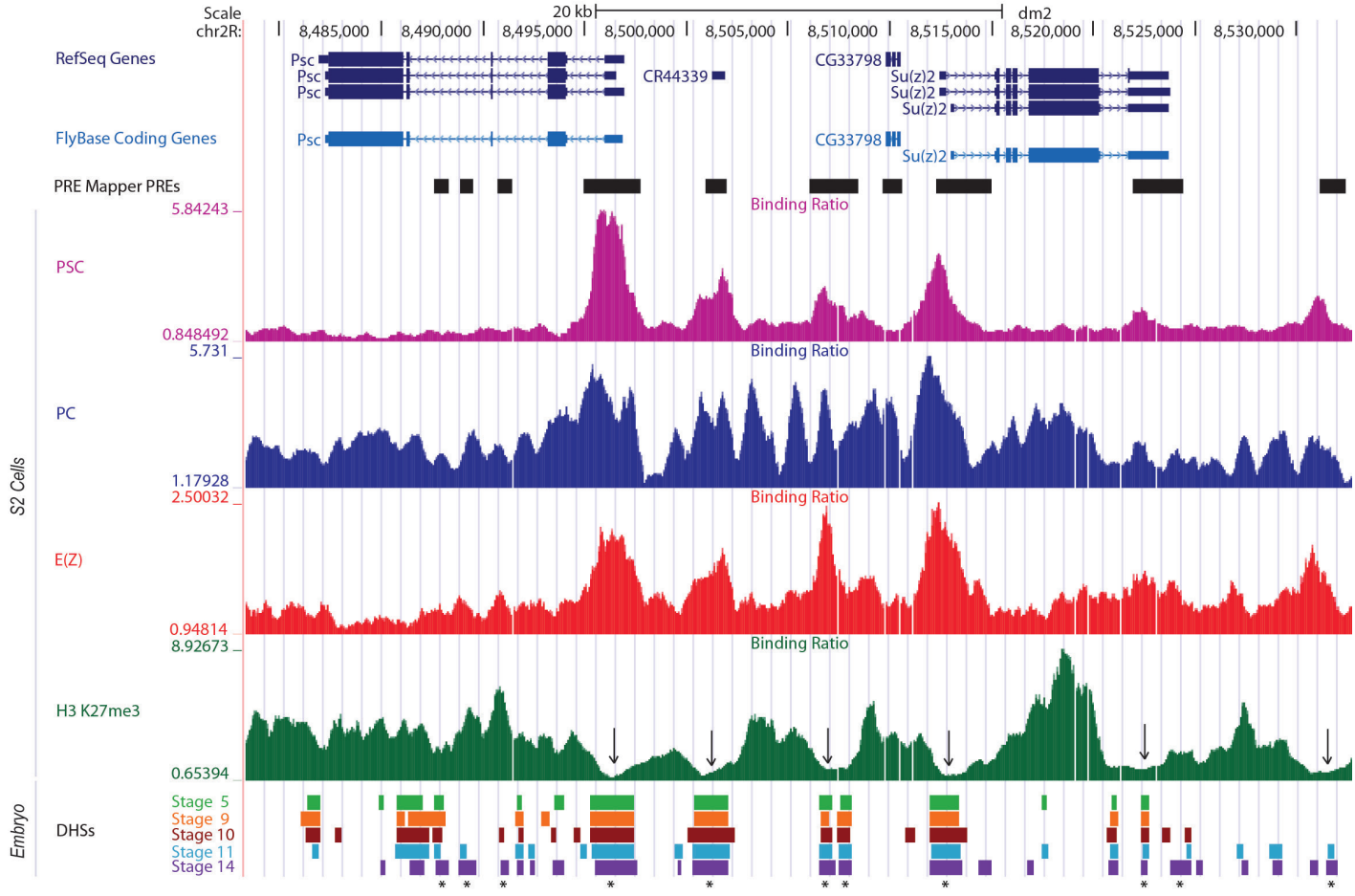

Fig. 4. Polycomb Response Element (PRE) Mapper tool derived PREs correlate with *in vivo* Polycomb group protein occupancies and DNase I Hypersensitive sites (Srinivasan and Mishra, 2020). *The PREs are plotted as a custom track on the UCSC genome browser (dm2) and shown with RefSeq and FlyBase protein-coding genes. The binding profiles of Polycomb proteins, PSC, PC, E(Z) and its associated H3K27me3 marks were obtained from Schwartz* et al.*, data (Schwartz* et al.*, 2006) and used on the UCSC browser for comparison. The black arrows over H3K27me3 profile indicate the PREs that correlate with strong Polycomb group protein occupancy. The bottom-most track shows the DNase I Hypersensitive sites (DHSs) for* Drosophila *embryonic developmental stages (Stage 5, 9, 10,11, 14). The asterisks represent the PRE-overlapping DHSs.*

and that such elements are spread across the genome to govern the chromatin-mediated gene regulatory system. The boundary, PRE search methods developed for *Drosophila* system could also be applied on other closely associated species or insects, if the boundary and PRE interacting proteins are conserved in those organisms. Identification of these elements in higher animals like zebrafish or mouse, however, will require additional experimental inputs such as well-defined DNA motifs and their interacting protein factors.

Boundaries and PREs share at least one common factor

One of the protein that appeared multiple times and sites in the regulation of BX-C is the Trithorax like/GAGA Associated Factor, TRL/GAF. Earlier studies have shown the role of GAF as an activator interacting in the promoter region of Ubx (Biggin and Tjian, 1988, Farkas *et al.*, 1994). We showed that GAF plays an essential role of the function of BX-C boundary, *Fab-7* (Schweinsberg *et al.*, 2004). While a factor that operates at a large number of sites in the genome is likely to have pleiotropic effects, using specific genetic resources we could demonstrate the role of GAF in the *Fab-7* boundary (Schweinsberg *et al.*, 2004). We also showed that GAF also plays a role in a number of other boundaries of the BX-C. This suggests a role of GAF in multiple boundaries of the BX-C either by direct binding to the relevant sites or by a long-range interaction, probably by clustering of the domain boundaries.

Role of GAF in PRE function was also shown by us using multiple approaches in *Drosophila* (Mishra *et al.*, 2001). A variety of other observations indicated multiple roles GAF in the regulation of BX-C genes that include an activator role at the Ubx promoter that was shown both biochemically and genetically (Chopra *et al.*, 2008), role in boundary function (Schweinsberg *et al.*, 2004) and repressive function (Mishra *et al.*, 2003, Mishra *et al.*, 2001). One of the questions that emerged from these findings was, how GAF carries out such different roles. To address question, we performed an extensive analysis of interactors of GAF and showed that while GAF is a recruiter protein, the choice of its partner decides whether the action is going to be that of activator, repressor, or boundary (Chopra *et al.*, 2008). The combination of protein-protein interac-

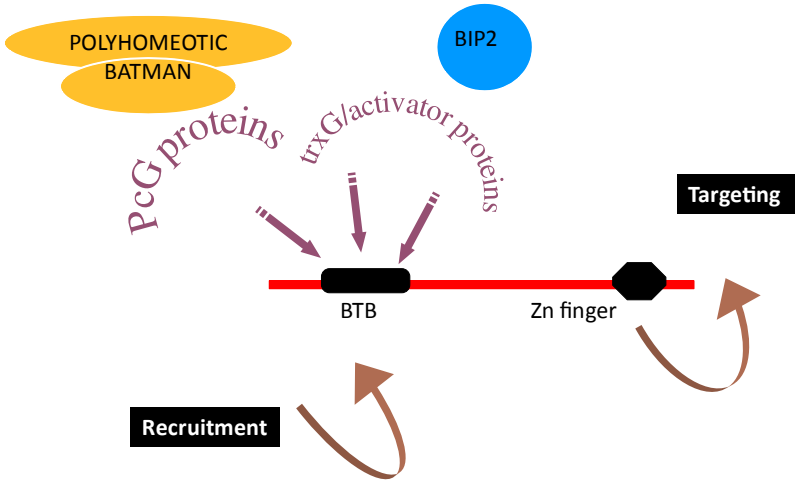

Fig. 5.The multifunctional GAGA associated factor (GAF). Protein-protein interaction domain (BTB) and DNA binding domain (Zn-finger) provided the versatility of recruiting relevant factors to respective targets in the genome. *The choice of interactors (PcG proteins, trxG proteins, etc.) determines the regulatory consequence.*

tion domain (BTB) and sequence-specific DNA binding domain (Zn-finger) allow GAF to function as a mediator that picks the desired recruitment partner through BTB domain and takes it to the target site in the genome through its Zn-finger domain (Fig. 5). We subsequently found a vertebrate homolog of GAF that carries out a similar function (Matharu *et al.*, 2010).

Lessons learned from fly are relevant to vertebrate systems

While there is remarkable conservation of homeotic genes and their collinear organization-function relationship from fly to human, there are noticeable differences too (Duboule, 1998, Duboule and Dolle, 1989, Krumlauf, 1994). For example, while *Hox* genes are organised in two clusters (ANT-C and BX-C) in fly, in vertebrates this kind of split has not been found. Vertebrate complexes are also ~10 fold more compact and, unlike fly Hox cluster, do not contain non-homeotic genes in or around the Hox complex. In this context, we asked if the chromatin level regulatory elements, viz., boundaries and PRE are conserved from fly to human.

As a first step, we analyzed DNA sequence between EVX2 and HoxD13 of mouse as these two genes are expressed in very distinct parts of the body, although they are only about 8 kilobases away from each other. Sequence comparison of this region across the vertebrate led to the identification of potential sequences that we tested in *Drosophila* using the transgene-based approach. It turned out that the mouse element functioned as boundary in the fly as well. Interestingly, the EVX2-D13 boundary element has a GAGA repeat that is conserved across the vertebrates which is also the binding site for GAF (Vasanthi *et al.*, 2010). *Drosophila* experiments showed that GAF was needed for the boundary function of mouse element in flies. Subsequently, we used the *Drosophila* system to test the functionality of an element upstream of HoxD that was shown to have a repressive role in mouse (Vasanthi *et al.,* 2013). This repressive element, when tested in fly using the transgene approach, showed a *Polycomb*-dependent repressive function. Therefore, it turns out to be the first vertebrate PRE to be reported. Interestingly, like *Drosophila* PRE, this mouse PRE also had GAF binding sites. Using genetic approaches, we showed that the repressive activity of mouse PRE responds to GAF/TRL in *Drosophila*. These observations not only uncover chromatin elements that are conserved across the animal kingdom in regulating homeotic gene complexes, they also provide a major lead to understand the evolutionary constraints in this organization (Srivastava *et al.*, 2015).

As discussed above, the higher order chromatin element involved in the regulation of homeotic genes in vertebrate showed the involvement of GAF. It is important to note here that earlier reports suggested the absence of GAF homologue in vertebrate. However, compelled by the observations that GAF sites are conserved across the vertebrate and the fact that they respond to *Drosophila* GAF in the fly system, we decided to look carefully if there is a vertebrate homologue to GAF present. Using bioinformatics, we identified vertebrate GAF homologue in zebrafish which is also conserved across the vertebrate including mouse. The vertebrate homologue of GAF also known as Th-POK consists of a protein-protein interaction domain, BTB and zinc finger domain for DNA binding (Matharu *et al.*, 2010). Subsequently, a number of experiments showed that GAF is involved in the regulation of homeotic genes and, as in *Drosophila,* has a genome-wide role at multiple loci (Srivastava *et al.*, 2013).

While the bithorax complex of *Drosophila* has been one of the most rewarding loci to understand genome organization and coordinated regulation of genes during development, it also provides useful insights to test similar aspects of regulatory mechanisms operating at other loci or in other organisms. Testing of vertebrate elements in flies and thereby allowing the use of powerful genetics of *Drosophila* provided several breakthroughs about how complex mechanisms of gene regulation operate in all organisms.

The non-homeotic function of *Hox* genes

While regulation of homeotic genes and their role in A-P body axis formation has been investigated extensively, their function, if any, and regulation later during development is still emerging. The first indication of the non-homeotic function of these genes was suspected by the observation that in several instances, more than one *Hox* gene is expressed in one region and even in the same cell. As the primary/homeotic function of the hox genes is to determine cell identity at specific regions of the body axis, why there should be two such proteins in the same place? Which one determines the cell identity and what is, if any, the role of the other gene? While the rule of posterior prevalence or posterior dominance explains that it is the posterior gene that decides the cell identity in such situations, it is not very obvious why sophisticated mechanisms that set the expression pattern of *Hox* genes leave undesirable genes to be expressed only to be countered by posterior ones. It is, therefore, likely that while posterior gene may dominate in determining the cell identity, the anterior gene may have a non-homeotic function in such situations. In vertebrates, several studies have been reported where homeotic genes play an important role in organogenesis in a very specific way (Rux and Wellik, 2017). There are few reports where the role of homeotic

genes has been implicated in diseases, particularly cancer (Cillo *et al*., 1999, Quinonez and Innis, 2014). In one of our studies, we showed a non-homeotic role of *abd-A* as a growth promoter in adult cuticle formation (Singh and Mishra, 2014). It turns out to be an important function of *abd-A* in the regions of the body where the identity of the segment is determined by *Abd-B* (Singh and Mishra, 2014). This finding, therefore, partially contradicts the rule of posterior prevalence. Regulating growth in the tissues like adult cuticle may also have implication in size and shape of the organisms. The variety of shape and sizes that we see in animals can be explained, at least partly, by such non-homeotic functions of homeotic genes.

In an independent study, we came across the role of the *Hox* gene, particularly *abd-A* in melanoma when ectopically expressed in the hematopoietic system (Ponrathnam and Mishra, 2018). While misregulation of *Hox* genes has been noticed in different types of cancer, a causal effect of *Hox* gene in melanoma was not shown earlier. Our findings support a broader role of *abd-A*, beyond its primary role in body axis formation, as a gene that determines the body shape and size of organs by promoting cell proliferation and growth. As may be anticipated, when misexpressed, such genes may lead to disease condition due to unregulated cell division.

Concluding remarks

BX-C of *Drosophila* has served as a remarkably efficient model not only to understand the underlying mechanisms of body axis formation, but also regulatory mechanisms operating across the genome. The regulatory elements deciphered from these studies have turned out to be relevant in vertebrates, including human (Heger and Wiehe, 2014, Kassis and Brown, 2013). These advantages in the context of new genomic approaches, viz., DNA-Protein interaction (ChIP-Seq), long-range interactions (5C, Hi-C) and advanced imaging techniques offer unprecedented opportunity to understand the complex and dynamic factors of coordinated regulation of genes (Mateo *et al*., 2019). Adding to that, genome editing technology that has been efficiently adopted by fly researchers, creates scope to ask question regarding genome organization and regulation that was not possible earlier (Gambetta and Furlong, 2018, Martin *et al*., 2016). For example, almost all the mutants in the *Drosophila* hox clusters are isolated from random mutagenesis approaches followed hy extensive screening. Now we can alter DNA sequence motifs and change the nature of cis regulatory domains in a precise and desired manner. Using such approaches, even the order of the genes or regulatory domains/elements can changed. These possibilities present opportunity to directly address the question of fine regulation and its link with the peculiar phenomenon of collinearity. Taken together, these new findings will help us solve the mystery of the mechanism that involves collinearity of *Hox* gene organization and function which turn out to be the only way for bilaterians to achieve A-P body axis formation. Understanding of such complex regulatory processes provides insight into the evolution of developmental mechanisms that contributed to the evolution of variety and complexity in living systems.

Acknowledgements

Authors acknowledge financial support from Council of Scientific and Industrial Research (CSIR), India. RKM acknowledges JC Bose fellowship from Science & Engineering Research Board (SERB), Department of Science and Technology (DST), Government of India.

References

AHANGER, S.H., SRINIVASAN, A., VASANTHI, D., SHOUCHE, Y.S. and MISHRA, R.K. (2013). Conserved boundary elements from the Hox complex of mosquito, Anopheles gambiae. *Nucleic Acids Res* 41: 804-816.

AKBARI, O.S., BAE, E., JOHNSEN, H., VILLALUZ, A., WONG, D. and DREWELL, R.A. (2008). A novel promoter-tethering element regulates enhancer-driven gene expression at the bithorax complex in the *Drosophila* embryo. *Development* 135: 123-131.

AOKI, T., SCHWEINSBERG, S., MANASSON, J. and SCHEDL, P. (2008). A stage-specific factor confers Fab-7 boundary activity during early embryogenesis in *Drosophila*. *Mol Cell Biol* 28: 1047-1060.

BARTKUHN, M., STRAUB, T., HEROLD, M., HERRMANN, M., RATHKE, C., SAUMWEBER, H., GILFILLAN, G.D., BECKER, P.B. and RENKAWITZ, R. (2009). Active promoters and insulators are marked by the centrosomal protein 190. *EMBO J* 28: 877-888.

BIGGIN, M.D. and TJIAN, R. (1988). Transcription factors that activate the Ultrabithorax promoter in developmentally staged extracts. *Cell* 53: 699-711.

BLASTYAK, A., MISHRA, R.K., KARCH, F. and GYURKOVICS, H. (2006). Efficient and specific targeting of Polycomb group proteins requires cooperative interaction between Grainyhead and Pleiohomeotic. *Mol Cell Biol* 26: 1434-1444.

BROWN, J.L., GRAU, D.J., DEVIDO, S.K. and KASSIS, J.A. (2005). An Sp1/KLF binding site is important for the activity of a Polycomb group response element from the *Drosophila* engrailed gene. *Nucleic Acids Res* 33: 5181-5189.

BROWN, J.P., SINGH, P.B. and COWELL, I.G. (2003). Composite cis-acting epigenetic switches in eukaryotes: lessons from *Drosophila* Fab-7 for the Igf2-H19 imprinted domain. *Genetica* 117: 199-207.

CHEN, Q., LIN, L., SMITH, S., LIN, Q. and ZHOU, J. (2005). Multiple Promoter Targeting Sequences exist in Abdominal-B to regulate long-range gene activation. *Dev Biol* 286: 629-636.

CHOPRA, V.S., SRINIVASAN, A., KUMAR, R.P., MISHRA, K., BASQUIN, D., DOCQUIER, M., SEUM, C., PAULI, D. and MISHRA, R.K. (2008). Transcriptional activation by GAGA factor is through its direct interaction with dmTAF3. *Dev Biol* 317: 660-670.

CILLO, C., FAIELLA, A., CANTILE, M. and BONCINELLI, E. (1999). Homeobox genes and cancer. *Exp Cell Res* 248: 1-9.

CUVIER, O., HART, C.M. and LAEMMLI, U.K. (1998). Identification of a class of chromatin boundary elements. *Mol Cell Biol* 18: 7478-7486.

DEJARDIN, J., RAPPAILLES, A., CUVIER, O., GRIMAUD, C., DECOVILLE, M., LOCKER, D. and CAVALLI, G. (2005). Recruitment of *Drosophila* Polycomb group proteins to chromatin by DSP1. *Nature* 434: 533-538.

DUBOULE, D. (1998). Vertebrate hox gene regulation: clustering and/or colinearity? *Curr Opin Genet Dev* 8: 514-518.

DUBOULE, D. and DOLLE, P. (1989). The structural and functional organization of the murine HOX gene family resembles that of *Drosophila* homeotic genes. *EMBO J* 8: 1497-1505.

FARKAS, G., GAUSZ, J., GALLONI, M., REUTER, G., GYURKOVICS, H. and KARCH, F. (1994). The Trithorax-like gene encodes the *Drosophila* GAGA factor. *Nature* 371: 806-808.

FERRIER, D.E. and HOLLAND, P.W. (2001). Ancient origin of the Hox gene cluster. *Nat Rev Genet* 2: 33-38.

FIEDLER, T. and REHMSMEIER, M. (2006). jPREdictor: a versatile tool for the prediction of cis-regulatory elements. *Nucleic Acids Res* 34: W546-W550.

GAMBETTA, M.C. and FURLONG, E.E.M. (2018). The Insulator Protein CTCF Is Required for Correct Hox Gene Expression, but Not for Embryonic Development in *Drosophila*. *Genetics* 210: 129-136.

GASZNER, M., VAZQUEZ, J. and SCHEDL, P. (1999). The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction. *Genes Dev* 13: 2098-2107.

GAUNT, S.J. (2015). The significance of Hox gene collinearity. *Int J Dev Biol* 59: 159-170.

HEGER, P. and WIEHE, T. (2014). New tools in the box: an evolutionary synopsis of chromatin insulators. *Trends Genet* 30: 161-171.

HO, M.C., SCHILLER, B.J., AKBARI, O.S., BAE, E. and DREWELL, R.A. (2011). Disruption of the abdominal-B promoter tethering element results in a loss of long-range enhancer-directed Hox gene expression in *Drosophila*. *PLoS One* 6: e16283.

HOLOHAN, E.E., KWONG, C., ADRYAN, B., BARTKUHN, M., HEROLD, M., RENKAWITZ, R., RUSSELL, S. and WHITE, R. (2007). CTCF genomic binding sites in *Drosophila* and the organisation of the bithorax complex. *PLoS Genet* 3: e112.

I, K.C. (2013). The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *J Hered* 104: 595-600.

KARCH, F., GALLONI, M., SIPOS, L., GAUSZ, J., GYURKOVICS, H. and SCHEDL, P. (1994). Mcp and Fab-7: molecular analysis of putative boundaries of cis-regulatory domains in the bithorax complex of *Drosophila melanogaster*. *Nucleic Acids Res* 22: 3138-3146.

KASSIS, J.A. and BROWN, J.L. (2013). Polycomb group response elements in *Drosophila* and vertebrates. *Adv Genet* 81: 83-118.

KAUFMAN, T.C., SEEGER, M.A. and OLSEN, G. (1990). Molecular and genetic organization of the antennapedia gene complex of *Drosophila melanogaster*. *Adv Genet* 27: 309-362.

KRUMLAUF, R. (1994). Hox genes in vertebrate development. *Cell* 78: 191-201.

LEWIN, H.A., ROBINSON, G.E., KRESS, W.J., BAKER, W.J., CODDINGTON, J., CRANDALL, K.A., DURBIN, R., EDWARDS, S.V., FOREST, F., GILBERT, M.T.P. *et al.,* (2018). Earth BioGenome Project: Sequencing life for the future of life. *Proc Natl Acad Sci USA* 115: 4325-4333.

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

LEWIS, E.B. (1998). The bithorax complex: the first fifty years. *Int J Dev Biol* 42: 403-415.

MAEDA, R.K. and KARCH, F. (2006). The ABC of the BX-C: the bithorax complex explained. *Development* 133: 1413-1422.

MARTIN, A., SERANO, J.M., JARVIS, E., BRUCE, H.S., WANG, J., RAY, S., BARKER, C.A., O'CONNELL, L.C. and PATEL, N.H. (2016). CRISPR/Cas9 Mutagenesis Reveals Versatile Roles of Hox Genes in Crustacean Limb Specification and Evolution. *Curr Biol* 26: 14-26.

MATEO, L.J., MURPHY, S.E., HAFNER, A., CINQUINI, I.S., WALKER, C.A. and BOETTIGER, A.N. (2019). Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* 568: 49-54.

MATHARU, N.K., HUSSAIN, T., SANKARANARAYANAN, R. and MISHRA, R.K. (2010). Vertebrate homologue of *Drosophila* GAGA factor. *J Mol Biol* 400: 434-447.

MIHALY, J., BARGES, S., SIPOS, L., MAEDA, R., CLEARD, F., HOGGA, I., BENDER, W., GYURKOVICS, H. and KARCH, F. (2006). Dissecting the regulatory landscape of the Abd-B gene of the bithorax complex. *Development* 133: 2983-2993.

MIHALY, J., HOGGA, I., BARGES, S., GALLONI, M., MISHRA, R.K., HAGSTROM, K., MULLER, M., SCHEDL, P., SIPOS, L., GAUSZ, J. *et al.,* (1998a). Chromatin domain boundaries in the Bithorax complex. *Cell Mol Life Sci* 54: 60-70.

MIHALY, J., MISHRA, R.K. and KARCH, F. (1998b). A conserved sequence motif in Polycomb-response elements. *Mol Cell* 1: 1065-1066.

MISHRA, K., CHOPRA, V.S., SRINIVASAN, A. and MISHRA, R.K. (2003). Trl-GAGA directly interacts with lola like and both are part of the repressive complex of Polycomb group of genes. *Mech Dev* 120: 681-689.

MISHRA, R.K. (2004). Edward B Lewis (1918-2004). *J Biosci* 29: 231-233.

MISHRA, R.K., MIHALY, J., BARGES, S., SPIERER, A., KARCH, F., HAGSTROM, K., SCHWEINSBERG, S.E. and SCHEDL, P. (2001). The iab-7 polycomb response element maps to a nucleosome-free region of chromatin and requires both GAGA and pleiohomeotic for silencing activity. *Mol Cell Biol* 21: 1311-1318.

OKULSKI, H., DRUCK, B., BHALERAO, S. and RINGROSE, L. (2011). Quantitative analysis of polycomb response elements (PREs) at identical genomic locations distinguishes contributions of PRE sequence and genomic environment. *Epigenetics Chromatin* 4: 4.

ORSI, G.A., KASINATHAN, S., HUGHES, K.T., SAMINADIN-PETER, S., HENIKOFF, S. and AHMAD, K. (2014). High-resolution mapping defines the cooperative architecture of Polycomb response elements. *Genome Res* 24: 809-820.

PARNELL, T.J., VIERING, M.M., SKJESOL, A., HELOU, C., KUHN, E.J. and GEYER, P.K. (2003). An endogenous suppressor of hairy-wing insulator separates regulatory domains in *Drosophila*. *Proc Natl Acad Sci USA* 100: 13436-13441.

PONRATHNAM, T. and MISHRA, R. (2018). *Drosophila* Hox genes induce melanised pseudo-tumours when misexpressed in hemocytes. *bioRxiv* doi: 10.1101/363978.

POSTIKA, N., METZLER, M., AFFOLTER, M., MULLER, M., SCHEDL, P., GEORGIEV, P. and KYRCHANOVA, O. (2018). Boundaries mediate long-distance interactions between enhancers and promoters in the *Drosophila* Bithorax complex. *PLoS Genet* 14: e1007702.

QUINONEZ, S.C. and INNIS, J.W. (2014). Human HOX gene disorders. *Mol Genet Metab* 111: 4-15.

RAMOS, E., GHOSH, D., BAXTER, E. and CORCES, V.G. (2006). Genomic organization of gypsy chromatin insulators in *Drosophila melanogaster*. *Genetics* 172: 2337-2349.

RAY, P., DE, S., MITRA, A., BEZSTAROSTI, K., DEMMERS, J.A., PFEIFER, K. and KASSIS, J.A. (2016). Combgap contributes to recruitment of Polycomb group proteins in *Drosophila*. *Proc Natl Acad Sci USA* 113: 3826-3831.

REZSOHAZY, R., SAURIN, A.J., MAUREL-ZAFFRAN, C. and GRABA, Y. (2015). Cellular and molecular insights into Hox protein action. *Development* 142: 1212-1227.

RINGROSE, L., REHMSMEIER, M., DURA, J.M. and PARO, R. (2003). Genome-wide prediction of Polycomb/Trithorax response elements in *Drosophila melanogaster*. *Dev Cell* 5: 759-771.

RUX, D.R. and WELLIK, D.M. (2017). Hox genes in the adult skeleton: Novel functions beyond embryonic development. *Dev Dyn* 246: 310-317.

SANCHEZ-HERRERO, E., VERNOS, I., MARCO, R. and MORATA, G. (1985). Genetic organization of *Drosophila* bithorax complex. *Nature* 313: 108-113.

SCHUETTENGRUBER, B., GANAPATHI, M., LEBLANC, B., PORTOSO, M., JASCHEK, R., TOLHUIS, B., VAN LOHUIZEN, M., TANAY, A. and CAVALLI, G. (2009). Functional anatomy of polycomb and trithorax chromatin landscapes in *Drosophila* embryos. *PLoS Biol* 7: e13.

SCHWARTZ, Y.B., KAHN, T.G., NIX, D.A., LI, X.Y., BOURGON, R., BIGGIN, M. and PIRROTTA, V. (2006). Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nat Genet* 38: 700-705.

SCHWEINSBERG, S., HAGSTROM, K., GOHL, D., SCHEDL, P., KUMAR, R.P., MISHRA, R. and KARCH, F. (2004). The enhancer-blocking activity of the Fab-7 boundary from the *Drosophila* bithorax complex requires GAGA-factor-binding sites. *Genetics* 168: 1371-1384.

SIMON, J., CHIANG, A., BENDER, W., SHIMELL, M.J. and O'CONNOR, M. (1993). Elements of the *Drosophila* bithorax complex that mediate repression by Polycomb group products. *Dev Biol* 158: 131-144.

SINGH, N.P. and MISHRA, R.K. (2014). Role of abd-A and Abd-B in development of abdominal epithelia breaks posterior prevalence rule. *PLoS Genet* 10: e1004717.

SMITH, S.T., WICKRAMASINGHE, P., OLSON, A., LOUKINOV, D., LIN, L., DENG, J., XIONG, Y., RUX, J., SACHIDANANDAM, R., SUN, H. *et al.,* (2009). Genome wide ChIP-chip analyses reveal important roles for CTCF in *Drosophila* genome organization. *Dev Biol* 328: 518-528.

SRINIVASAN, A. and MISHRA, R.K. (2012). Chromatin domain boundary element search tool for *Drosophila*. *Nucleic Acids Res* 40: 4385-4395.

SRINIVASAN, A. and MISHRA, R.K. (2020). Genomic organization of Polycomb Response Elements and its functional implication in Drosophila and other insects. *J Biosci* 45: 12.

SRIVASTAVA, S., DHAWAN, J. and MISHRA, R.K. (2015). Epigenetic mechanisms and boundaries in the regulation of mammalian Hox clusters. *Mech Dev* 138 Pt 2: 160-169.

SRIVASTAVA, S., PURI, D., GARAPATI, H.S., DHAWAN, J. and MISHRA, R.K. (2013). Vertebrate GAGA factor associated insulator elements demarcate homeotic genes in the HOX clusters. *Epigenetics Chromatin* 6: 8.

STRUTT, H., CAVALLI, G. and PARO, R. (1997). Co-localization of Polycomb protein and GAGA factor on regulatory elements responsible for the maintenance of homeotic gene expression. *EMBO J* 16: 3621-32.

THOMAS, S., LI, X.Y., SABO, P.J., SANDSTROM, R., THURMAN, R.E., CANFIELD, T.K., GISTE, E., FISHER, W., HAMMONDS, A., CELNIKER, S.E. *et al.,* (2011). Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. *Genome Biol* 12: R43.

VASANTHI, D., ANANT, M., SRIVASTAVA, S. and MISHRA, R.K. (2010). A functionally conserved boundary element from the mouse HoxD locus requires GAGA factor in *Drosophila*. *Development* 137: 4239-4247.

VASANTHI, D., NAGABHUSHAN, A., MATHARU, N.K. and MISHRA, R.K. (2013). A functionally conserved Polycomb response element from mouse HoxD complex responds to heterochromatin factors. *Sci Rep* 3: 3011.

ZENG, J., KIRK, B.D., GOU, Y., WANG, Q. and MA, J. (2012). Genome-wide polycomb target gene prediction in *Drosophila melanogaster*. *Nucleic Acids Res* 40: 5848-5863.

ZHOU, J. and LEVINE, M. (1999). A novel cis-regulatory element, the PTS, mediates an anti-insulator activity in the *Drosophila* embryo. *Cell* 99: 567-575.

Further Related Reading, published previously in the *Int. J. Dev. Biol.*

Hox genes, clusters and collinearity
Robb Krumlauf
Int. J. Dev. Biol. (2018) 62: 659-663
https://doi.org/10.1387/ijdb.180330rr

Drosophila castor is regulated negatively by the Ubx and abdA genes, but positively by the AbdB gene
Ju-Hyun Ahn, Siuk Yoo, Hyun-Joo Park, Keuk-Il Jung, Sang-Hee Kim and Sang-Hak Jeon
Int. J. Dev. Biol. (2010) 54: 1251-1258
https://doi.org/10.1387/ijdb.093037ja

Non-genic transcription at the Drosophila bithorax complex – functional activity of the dark matter of the genome
Margaret C.W. Ho, Benjamin J. Schiller, Sara E. Goetz and Robert A. Drewell
Int. J. Dev. Biol. (2009) 53: 459-468
https://doi.org/10.1387/ijdb.082647mh

Homeotic genes influence the axonal pathway of a Drosophila embryonic sensory neuron
David J Merritt and Paul M Whitington
Int. J. Dev. Biol. (2002) 46: 633-638
http://www.intjdevbiol.com/web/paper/12141451

Trithorax and the regulation of homeotic gene expression in Drosophila: a historical perspective
P W Ingham
Int. J. Dev. Biol. (1998) 42: 423-429
http://www.intjdevbiol.com/web/paper/9654027

The bithorax complex: the first fifty years
E B Lewis
Int. J. Dev. Biol. (1998) 42: 403-415
http://www.intjdevbiol.com/web/paper/9654025

Segmental determination in Drosophila central nervous system: analysis of the abdominal-A region of the bithorax complex
H Jijakli and A Ghysen
Int. J. Dev. Biol. (1992) 36: 93-99
http://www.intjdevbiol.com/web/paper/1352698