

Chromosomal organization at the level of gene complexes

Vivek S. Chopra

Received: 15 April 2010/Revised: 17 October 2010/Accepted: 26 October 2010/Published online: 16 November 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Metazoan genomes primarily consist of non-coding DNA in comparison to coding regions. Non-coding fraction of the genome contains *cis*-regulatory elements, which ensure that the genetic code is read properly at the right time and space during development. Regulatory elements and their target genes define functional landscapes within the genome, and some developmentally important genes evolve by keeping the genes involved in specification of common organs/tissues in clusters and are termed gene complex. The clustering of genes involved in a common function may help in robust spatio-temporal gene expression. Gene complexes are often found to be evolutionarily conserved, and the classic example is the hox complex. The evolutionary constraints seen among gene complexes provide an ideal model system to understand *cis* and *trans*-regulation of gene function. This review will discuss the various characteristics of gene regulatory modules found within gene complexes and how they can be characterized.

Keywords Gene regulation · Chromatin · Enhancer · Insulator · Genomics

Introduction

The large genomes found in metazoans consist of non-coding DNA that serves to define regulatory genomic

landscapes. These regulatory DNA elements are found interspersed in the genome and regulate the function of protein coding genes. The abundant nature of such regulatory elements raises the possibility that spurious interactions could occur among neighboring regulatory elements. Thus, to overcome non-specific interactions nature has devised methods to clearly demarcate functional genomic landscapes for each gene (Fig. 1). The functional genetic intervals found within genomes are comprised of genes and their regulatory elements. These functional intervals are clearly demarcated from the neighboring genes to form a functional domain of gene function (Fig. 1). Sometimes genes that help define a specific developmental event or tissue-type occur near each other, are co-regulated during development and are termed gene complexes. The linkage among the co-expressed genes of these clusters is significantly conserved, and the expression patterns of genes within clusters generally co-evolve as suggested by cross-species analyses [1] (Fig. 2). Such evolutionary selection could be mediated by chromatin interactions with the nuclear matrix and long-range remodeling of chromatin structure. Due to this assumption such gene complexes are widely used as model systems to understand spatio-temporal gene regulatory mechanisms. Well-studied complexes include the Hox complex [2], heart complex [3] and human globin locus [4, 5]. Studies of these complexes have shown a tight genetic linkage as well as an epigenetic basis of gene regulation that includes the regulation at the higher order chromatin organization. The gene complexes can vary in size starting from a pair of genes (e.g., *en/inv* complex) [6, 7] and up to five genes or more (e.g., hox, heart and globin complex). There are also examples of single genes that are regulated by more than one tissue-specific enhancer; a classic example is that of *even-skipped* [8, 9]. This review will discuss the use of

V. S. Chopra (✉)
Department of Molecular and Cell Biology,
585, Life Sciences Addition, University of California,
Berkeley, CA 94720-3200, USA
e-mail: vschopra@berkeley.edu; v.s.chopra@gmail.com

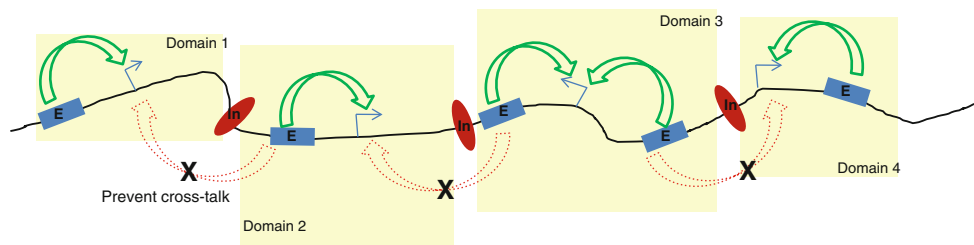


Fig. 1 Functional domains within genomic landscapes. Genomes are gene dense in nature and have genes adjacent to each other along with their regulatory elements like enhancers (blue box, *E*) and insulators (red oval, *In*). These are responsible for activating genes as well (green arrows) as insulating them from neighboring regulatory elements, respectively. The insulator elements help demarcate

domains of genetic function (yellow box) and separate them from neighboring domains. They also prevent crosstalk of adjacent enhancers (red dashed arrows) to activate genes within a given domain. These demarcated functional units of genes and their regulatory elements make up the genomic landscapes within the genome

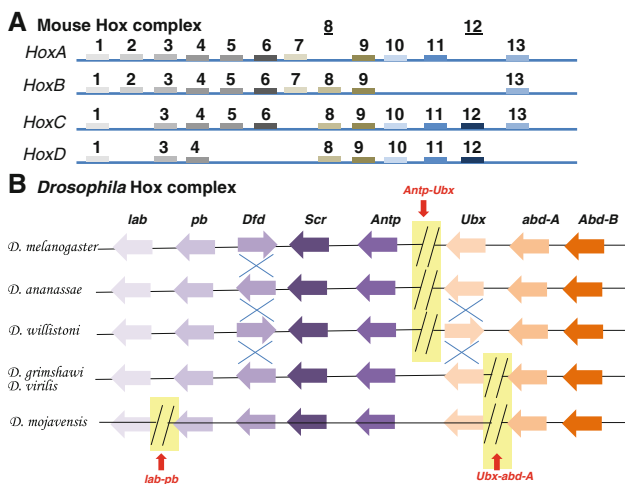


Fig. 2 Hox complex of Mouse and *Drosophila*. **a** The mouse Hox complex is separated into four clusters, namely HoxA–D. They help in formation of anterior to posterior structures of the mouse body axis, respectively. The four clusters contain from 9 to 11 paralogue genes. These genes are arranged in the same order as they are expressed along the body axis. For example, the mouse HoxA1 is expressed in the head regions and HoxD13 in the distal regions of the limbs. **b** The fly hox cluster on the other hand is split into two complexes in *D. melanogaster* to form the Antennapedia complex (determining head to T2 segments) and the bithorax complex (determining T3 to A8 segments). This complex is conserved among the other fly species with some gene inversions (shown by blue cross) as seen for *Dfd* and *Ubx* genes and shifting in breakpoints of the cluster (red arrow). There are three splits known in Drosophilids as shown in the figure between *lab-pb*, *Antp-Ubx* and *Ubx-abd-A*. These splits do not impair the collinear expression of the Hox genes along the A–P body axis

such gene complexes to dissect the different layers in gene regulation during development.

To understand the regulation of gene complexes we need to know about the different kinds of *cis*-regulatory modules (CRMs) that make up and regulate these complex genetic loci. Examples of CRMs include enhancers [9–12], insulators (boundary elements) [13 14] and maintenance elements like Polycomb response elements (PREs) [15], which coordinate proper gene regulation and are the key

regulators that mediate spatio-temporal gene activity within complexes. Enhancers bind activator proteins and help bring robust transcription of genes. Insulator elements prevent interference from foreign enhancers during transcription by demarcating domains of gene activity/inactivity. Initiators help initiate/establish tissue-specific gene activation, and maintenance elements help maintain the established gene expression/repression pattern in a given cell type. These regulatory elements are capable of bringing about novel gene expression/repression patterns and thus hold the key to dictate developmental novelties during evolution [16]. Recently, the presence of secondary enhancers has been shown to drive the expression of the target gene in the same tissue, and they have been termed shadow enhancers [17, 18]. These enhancers might help in robust gene expression and also contribute toward the evolution of the gene expression pattern [19]. There are other kinds of specialized regulatory elements that help specific enhancers find their respective target promoters and are termed promoter targeting sequences (PTS) [20]. They have been identified for Hox genes (*Scr*, *Abd-B*) and early patterning genes (*eve*) [7, 10, 20]. Another very unique feature called “homing” has been observed for a few loci and conveys the ability of a piece of non-coding DNA to insert in its native position when cloned into a transposon-based transgene [21, 22]. The mechanisms of insulator function, PTS action and homing are not clearly known, but will be important for understanding how CRMs can precisely find their respective target genes within gene complexes. The Hox and heart complex will be used as examples to illustrate the different layers of gene regulation events within gene complexes.

Hox complex

The Hox complex genes were identified based on their striking homeotic transformations (transformation of one

body part into another, suggesting that these genes control cellular identity). The *Drosophila* hox genes are organized into two homeotic complexes (HOM-C), the bithorax complex (BX-C) and the Antennapedia complexes (ANT-C), which are equivalent to the four HOM-Cs found in humans (HoxA-D) [23–25] (Fig. 2). These genes control body patterning in animals, from nematodes to vertebrates, as they are conserved across the species [26, 27] and specify the identity of body segments along the anterior-posterior (A-P) body axis [28, 29]. These genes were found to contain a conserved DNA sequence that encodes a 60-amino-acid DNA-binding motif, the homeodomain, and were coined homeobox (Hox) genes [30, 31]. The Hox genes are known to control the transcription of downstream target genes that are responsible for morphological diversification.

In addition to the sequence conservation between homeotic genes, there are remarkable similarities in their organization and regulation. Hox genes are generally found in clusters that most likely arose from the duplication and divergence of a single ancestral gene. Nematodes, for example, have one cluster with four genes; *Drosophila* has a split cluster with eight genes, whereas vertebrates have four clusters with a total of 39 genes [32, 33] (Fig. 2). The proximal to distal order of genes within each cluster corresponds to their functional domains along the A-P body axis. This “spatial colinearity” of organization on the chromosome and respective functional domain along the A-P axis is also conserved during evolution. Furthermore, in some organisms, the anterior to posterior order of Hox gene expression is accompanied by an early to late temporal order of expression, a phenomenon called “temporal colinearity” [34–37]. This remarkable genomic organization has fascinated biologists for many years, but its functional link to regulatory mechanisms is yet to be satisfactorily explained. Several models implicating chromatin organization [38], employment of shared regulatory elements [39, 40] or global control elements shared among multiple genes [41] have been proposed, but these have yet to be firmly established.

Although colinearity was discovered in the BX-C of the fruitfly *Drosophila melanogaster* [28], flies also show many exceptions to the general rule concerning colinearity and Hox clustering. For example, the HOM-C of *Drosophila* is split into two gene clusters, the BX-C and the ANT-C. Furthermore, flies not only have a split complex, but the BX-C can be further split without apparently compromising its function [42, 43]. It is also known that the HOM-C in different *Drosophila* species is split at different positions [44, 45] (Fig. 2b). However, these splits within the Hox complex do not impair the spatial expression pattern of Hox genes. For example, the split between

Ubx and *abd-A* genes seen in *D. virilis* and *D. grimshawi* does not affect the haltere or anterior abdominal segment formation, suggesting that these splits do not cause loss or ectopic gene expression leading to morphological changes (Fig. 2b). In fact, *D. mojavensis* has two splits within the complex one between *lab* and *pb*, and another between *Ubx* and *abd-A*, but still retains its segment-specific Hox gene expression intact. Taken together, these lines of evidence suggest that although the fly Hox genes remain collinear, they have lost the evolutionary constraints to exist as a single unified complex.

In vertebrates the hox genes form several homologous clusters because of duplication during the course of evolution. Their genes play significant and complementary roles in axial patterning during development. Vertebrates have four sets of homeotic paralogous genes, each one organized in one intact cluster as compared to one set of homeotic genes organized into two clusters in *Drosophila* (Fig. 2). These are *HoxA*, *HoxB*, *HoxC* and *HoxD*, located on different chromosomes [46]. Genes are numbered from 1 to 13 in their physical order along the 3′–5′ direction on the chromosome. All four clusters have gone through different “gene loss” events and contain between 9 and 11 genes. Alignment of the sequences of hox genes based on relative position, sequence identity and domains of expression along the anterior-posterior axis shows a clear relationship among genes in the mouse and *Drosophila* complexes, suggesting that these complexes arose from a common ancestor, present before the divergence of lineages that gave rise to arthropods and vertebrates [47].

The vertebrate hox complexes are compact and smaller in size as compared to the *Drosophila* complex, and are the most repeat-free region of their genomes and show extensive conservation of non-coding DNA sequences associated with them [48]. All vertebrate hox genes are transcribed in one direction, unlike the *Drosophila* hox complex [49, 50]. The vertebrate hox genes are differentially activated by retinoic acid (RA) according to their physical location within the four chromosomal loci [51]. The genes located at the 3′ end of each one of the four hox loci are activated by RA in a sequential order collinear with their 3′–5′ arrangement in the cluster: 3′ hox genes respond to a lower concentration of RA, whereas upstream genes respond progressively to higher concentrations [52]. Subsequent studies showed that the vertebrate hox genes could respond to RA because of the presence of enhancers called retinoic acid response elements (RAREs) that are also conserved across species [53].

Gene complexes are regulated in proper time and space by a myriad of gene regulatory elements that will be illustrated using the BX-C complex as an example.

Layers of gene regulatory modules within BX-C

One of the most extensively studied gene complexes is BX-C. The *Drosophila* BX-C genes *Ubx*, *abd-A* and *Abd-B* genes control the identity of nine parasegmental units in the posterior two-thirds of the fly [2, 29, 54, 55]. The homeotic genes of the BX-C are expressed in intricate temporal and spatial patterns in an overlapping set of parasegments (PS) in embryonic development that give rise to segments in the adult fly. *Ubx* is expressed from PS5 to PS12-13, *abd-A* from PS7 to PS12 and *Abd-B* from PS10 to PS14 [56–61]. Morphogenesis of segment-specific structures requires the elaboration of the precise parasegmental expression patterns of these genes. Mutations that alter *Ubx*, *abd-A* or *Abd-B* expression can transform the parasegment identity. The complex transcription pattern of the BX-C genes is generated by a large (about 315 Kb) *cis*-regulatory region. Genetic and molecular analysis has defined nine PS-specific *cis*-regulatory subregions within this large DNA segment. These PS-specific subregions *abx/bx* (*anteriobithorax/bithorax*), *bxd/pbx* (*bithoraxoid/posteriobithorax*), *iab-2* (*infrabdominal-2*), *iab-3*, *iab-4*, *iab-5*, *iab-6*, *iab-7* and *iab-8,9*, are arranged in the same order along the chromosome as the PS they affect. The *abx/bx* and *bxd/pbx* *cis*-regulatory subregions are responsible for proper *Ubx* expression in PS5 and PS6, respectively [60]. Similarly, the *abd-A* expression in PS7, 8 and 9 is under the control of the *iab-2*, *iab-3* and *iab-4* *cis*-regulatory units, respectively [58]. Finally, the *iab-5* through *iab-8, 9* subregions direct *Abd-B* expression in PS10-14, in the same order [57, 62] (Fig. 3). Thus, colinearity applies not only to

gene order, but also to the order of *cis*-regulatory domains along the chromosome. These *cis*-regulatory elements are protected from interference from the neighboring *cis*-regulatory elements by the presence of chromatin domain boundary/insulator elements. These chromatin domain boundary elements help in demarcating the domain of action of the *iab*'s and prevent crosstalk of the *cis*-regulatory elements [63] (Fig. 3).

Mutations in *cis*-regulatory elements

Several mutations in these *cis*-regulatory regions lead to interesting homeotic phenotypes, and their molecular analysis has revealed the chromatin level regulatory mechanisms involved in the regulation of BX-C. Loss-of-function (LOF) mutations in any of these nine *cis*-regulatory subregions typically transform the corresponding PS into a copy of the PS immediately anterior (Fig. 4). Consistent with the observed phenotypic effects on segmental identity, the normal spatial and temporal expression pattern in the affected PS is replaced by an expression pattern that mimics the one immediately anterior, for example, *iab-7^{SZ}* in which almost the entire *iab-7* region is deleted and transforms PS12 to PS11 (A7 to A6) [64, 65] (Fig. 4). Gain-of-function (GOF) mutations have the opposite phenotype in which the affected PS is transformed into a more posterior PS. For example, the *Fab7* (Frontoabdominal-7) mutation in which the boundary between *iab-6* and *iab-7* is deleted leads to ectopic activation of *iab-7* in PS11, transforming it to PS12 (A6 to A7) [66, 67] (Fig. 4).

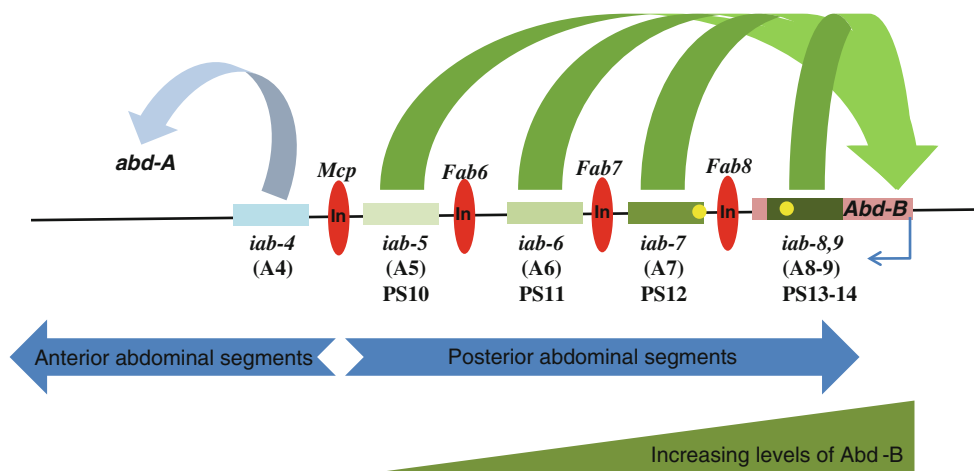


Fig. 3 CRMs within the *Abd-B* locus of the bithorax complex leading to collinear gene expression. The *Abd-B* gene is activated from PS10-14 in the developing embryo. The levels of *Abd-B* are the lowest in PS10 and highest in PS14, and this is believed to be either due to the strength of the *iab*'s (shown in shades of green rectangles) or the distance of *iab*s from the *Abd-B* promoter. The *iab-5* enhancer drives

the expression of *Abd-B* in PS10, *iab-6* in PS11, *iab-7* in PS12 and *iab-8,9* in PS13-14. Each of these enhancer domains is demarcated by insulator elements (red oval) like *Mcp*, *Fab6*, *Fab7* and *Fab8*. The *iab*s are also known to contain PRE elements (yellow circle) as experimentally identified for the *iab-7* and *iab-8* PREs. The earlier *iab* elements such as the *iab-4* act upon the *abd-A* gene in anterior PS

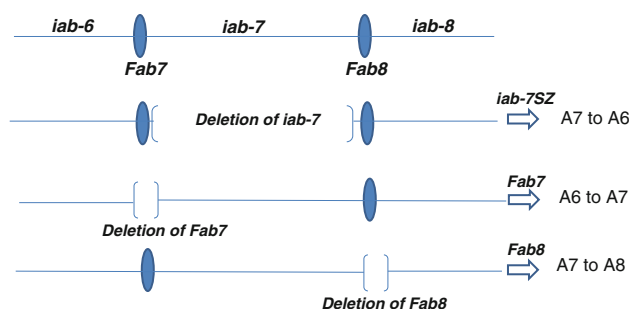


Fig. 4 Genetic and molecular nature of deletions within regulatory elements of the BX-C. The *iab-6*, *iab-7* and *iab-8* cis-regulatory elements drive the expression of *Abd-B* gene in increasing levels in PS11, PS12 and PS13, respectively, which give rise to the abdominal segments A6 to A8. The *iab-6*, *iab-7* and *iab-8* cis-regulatory elements are separated by the *Fab7* and *Fab8* insulator elements. The function of these insulators and enhancers has been deciphered by genetic ablations of these regions resulting in homeotic phenotypes. The deletions that remove the *iab-7* element (*iab-7SZ*) lead to the transformation of A7 to A6 as *iab-6* takes over the function of the deleted *iab-7* in the A7 region, giving rise to the loss of the function anteriorization phenotype. This happens because in PS12/A7 all the regulatory elements from *iab-2* to *iab-7* are in an open conformation state, but since *iab-7* is proximal to the *Abd-B* gene it prevails in driving it in the PS12/A7 region. In case of the *iab-7* deletion, the last regulatory module to be in an open chromatin state is *iab-6* in PS12/A7; hence it takes over the *iab-7* function and leads to A7 to A6 transformation. In case of the deletion of the boundary element *Fab7*, the *iab-6* and *iab-7* domains fuse to form a hybrid element where *iab-7* prevails and is ectopically activated within the *iab-6* domain (PS11/A6), thus leading to A6 to A7 dominant gain of function transformation. A similar posteriorization phenotype is observed in the case of *Fab8* deletion where the *iab-7* and *iab-8* domains are fused, leading to ectopic activation of *iab-8* in the *iab-7* domain (A7), thus leading to A7 to A8 transformation

Transvection

An unusual feature of the Diptera is that homologous chromosomes are synapsed in somatic cells during interphase. At a number of loci in *Drosophila*, this pairing can significantly influence gene expression. E.B. Lewis detected this phenomenon in *Drosophila* and coined the term transvection. Transvection can be described as the phenomenon in which the expression of a gene on one chromosome depends on the pairing with its homologous region [68] (Fig. 5). For example, the deletion analysis of the *Abd-B* gene strongly suggests the existence of transvection that tethers cis-regulatory regions to the promoter-upstream region [69, 70]. It has been found that while *Abd-B* point mutations do not complement the phenotype of an *iab-7* deletion in A7, *Abd-B* alleles deleted for the promoter region do complement *iab-7* deletions in *trans*-heterozygotes. The complementation is a result of the action of the wild-type *iab-7* on the wild-type *Abd-B* in *trans*. As this *trans*-regulation is not detected when the somatic pairing of homologous chromosomes is disturbed by chromosomal

rearrangements, it represents a case of 'transvection.' The degree of complementation in A7 depends on the size of the promoter deletion: the larger the deletion is, the stronger the *trans*-regulation, suggesting that the promoter upstream region of the *Abd-B* gene consists of numerous discrete elements that cooperate in locking individual cis-regulators to the *Abd-B* gene [69, 70].

Transvection can also occur by the action of silencers in *trans* or by the spreading of position effect variegation from rearrangements having heterochromatic breakpoints to paired non-rearranged chromosomes [71]. Several cases of transvection require ZESTE, a DNA-binding protein that is thought to facilitate homolog interactions by self-aggregation [72]. Recently, condensins have been shown to negatively regulate transvection at the yellow locus [73]. Genes showing transvection can differ greatly in their response to pairing disruption. In several cases, transvection appears to require intimate synapsis of homologs [74, 75]. However, in at least one case (transvection of the *iab-5,6,7* region of the BX-C), transvection is independent of synapsis within and surrounding the interacting gene [69, 76]. The latter example suggests that transvection could well occur in organisms that lack somatic pairing. In support of this, transvection-like phenomena have been described in a number of different organisms, including plants, fungi and mammals [77, 78].

The *Abd-B* gene controls the morphogenesis of posterior abdominal segments in *Drosophila*, and its expression is regulated by a series of 3' enhancers that are themselves transcribed. Studies utilizing RNA Fluorescence In situ Hybridization (FISH) to visualize nascent transcripts associated with coding and non-coding regions of *Abd-B* in developing embryos and confocal imaging suggest that distal enhancers often loop to the *Abd-B* promoter region. Surprisingly, enhancers located on one chromosome frequently associated with the *Abd-B* transcription unit located on the other homolog. These *trans*-homolog interactions could be interpreted as the direct visualization of the genetic phenomenon, transvection, whereby certain mutations in *Abd-B* can be rescued in *trans* by the other copy of the gene [79]. It has also been shown that a 10-kb sequence in the 3' flanking region mediates pairing of *Abd-B* alleles, thereby facilitating *trans* looping of distal enhancers [79]. Such *trans*-homolog interactions might be a common mechanism of gene regulation in higher metazoans.

Regulatory non-coding RNAs

A number of studies in parallel suggest that the cis-regulatory elements in the BX-C may in fact be operating in synchrony with a system of intergenic, non-coding RNA transcripts. While it has been known for decades that such

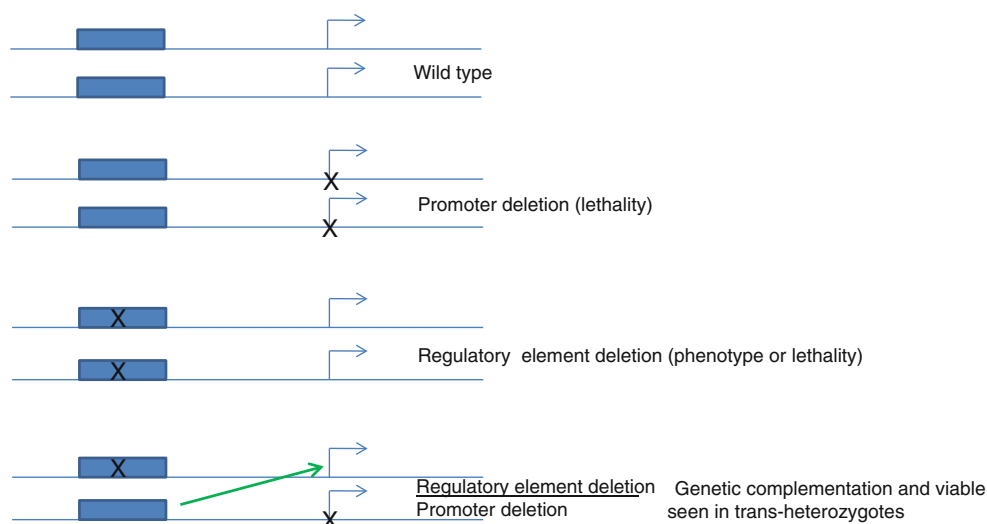


Fig. 5 Transvection was first observed within the bithorax complex as genetic lesions removing promoter and *cis*-regulatory elements were easily available. Mutations that delete the promoter region are usually homozygous lethal, whereas the *cis*-regulatory deletions are homozygous viable, displaying hypomorphic phenotypes or lethality. When these two mutations were combined to form a *trans*-

heterozygous genetic complementation was observed. This can be explained only by the ability of the *cis*-regulatory element to activate the gene in *trans* (green arrow) present in the homologous chromosome due to pairing. Hence, the genetic complementation of a promoter deletion with a *cis*-regulatory deletion is termed as transvection

transcripts are produced in abundance at the BX-C, their functional role has not been clear. Lipshitz et al. [80] worked with such transcripts as early as 1987, focusing on those produced in the *bithoraxoid* (*bxd*) region of the *Ubx* gene of the BX-C. Similarly, substantial transcription through the intergenic region between *abd-A* and *Abd-B* was identified [81] and believed to play a role in maintaining the epigenetic state of the regulatory elements within the BX-C. Although the generation of these transcripts was observed and their molecular characteristics investigated, their function remained unclear. In fact, repressive nature exerted by PREs can be reverted to an active heritable state upon forced transcription through PREs [82, 83]. Studies from other genetic loci have indicated that non-genic transcription may be a common feature at tightly regulated gene complexes. Studies at the human β -globin locus revealed a transcription program in which various regulatory domains are subject to chromatin remodeling via intergenic transcription [84]. More recently, work on the immunoglobulin heavy chain locus in mice revealed a similar active role for non-coding transcription in the alteration of gene function, in which antisense transcription through the VH region correlates with a switch from DJH to VDJH recombination [85]. The presence of non-genic transcription programs at these distinct gene complexes suggests that the BX-C intergenic transcripts may have a functional activity.

The generation of a comprehensive profile of the BX-C non-genic transcripts has provided valuable information about their role upon interaction with the *cis*-regulatory

elements at the BX-C. This task was accomplished by designing a series of in situ hybridization probes spanning from *iab-2* to *iab-8* in the *abd-A-Abd-B* intergenic region [86]. These studies provided a temporal and spatial map of transcription from this region in the developing embryo and offered some insight into the function of the non-genic BX-C transcripts. Spatially, the transcripts in the embryo are expressed in the same co-linear pattern as their chromosomal organization on the BX-C. The transcripts from *iab-2* are found more anterior to those from *iab-3*, those from *iab-3* are found anterior to those from *iab-4*, and so on. Furthermore, transcripts are also retained within individual *iab* chromosomal regions. In this way, a single type of RNA is produced per *iab* region, and the transcription does not appear to traverse the characterized insulator elements. While the expression patterns of all the transcripts have a defined anterior margin in the embryo, the posterior limits can spread into the regions of the other *iab* transcripts. Almost all of the transcripts are generated from the sense strand relative to the direction of transcription for *abd-A* and *Abd-B*. If this non-genic transcription was spurious, then the transcripts should be generated randomly from both strands, showing no preference. The predominant transcription of the sense strand and the specific expression patterns in the embryo argue for a functional role for the intergenic transcripts.

Reports have shown that ectopic transcription through the boundary element could lead to the abolition of the insulator activity [87]. It was observed that this ectopic transcription through the endogenous boundary elements at

the *bxd/iab-2* junction in the BX-C subsequently activated more posterior regulatory domains leading to abdominal segment A1 to A2 transformation similar to *Ultrabdominal* (*Uab*) mutation [87]. The phenotypic effect was therefore not caused by a deletion of a boundary element but rather by transcription through the insulator. Similar work, using different experimental procedures, also indicated a function for ectopic transcription at the BX-C. The functionality of a trimmed-down version of the *scs* insulator was examined by replacing endogenous *Fab-7* with *scs* using gene conversion. When the new *scs* fragment, which also contained a promoter, was inserted in an orientation such that the promoter could drive transcription through the PRE adjacent to the *Fab-7* region, the proper segmental identity was disrupted in a manner similar to that of a *Fab-7* deletion [88]. The ectopic transcription resulted in a transformation of abdominal segment 6 into segment 7. This result signifies that, despite the insulating activity of the *scs* fragment, the transcription through the adjacent PRE serves to remove its silencing effects. This causes the *cis*-regulatory information in the *iab-7* domain to become active anterior to its normal position in the embryo. It was shown that a deletion at the *Mcp* region [28] results in a loss of the non-genic transcription in the adjacent *iab-4* domain, presumably due to inactivation of the promoter for this transcript. The absence of the *iab-4* transcript is correlated with a transformation of abdominal segment 4 into 5 [89]. Taken together, these studies suggest that controlled non-genic transcription in the *iab* regions is critical to the proper function of the BX-C and appears to play a role in activating *cis*-regulatory domains during development.

Trans-acting regulators of the BX-C

Initiation factors

In addition to the *cis*-acting DNA elements like *iab*'s, boundary elements and PREs, there are crucial *trans*-acting factors that regulate the expression of the BX-C by binding to these *cis*-regulatory elements. These *trans*-acting factors can be grouped into two classes: those required for establishing the initial domains of BX-C expression and those required to maintain the initial patterns throughout development. The specific pattern of BX-C gene expression is initially established by the combinatorial action of transcription factors encoded by the segmentation genes [90–95]. For example, the *iab-2* regulatory region activates the *abd-A* gene in PS7. Amazingly, two independently isolated mutations that transform PS5 to PS7 both alter the exact same base pair within the *iab-2* domain, destroying a binding site for the gap gene *Krüppel* [94]. This provides good evidence that *Krüppel* is one of the repressive factors

that prevents the activation of the *iab-2* domain in PS anterior to PS7. *Krüppel* has also been shown to act as a repressor in the *Superabdominal* (*Sab*) mutation, which causes A3 to A5 transformation [96]. A point mutation removing the KRUPPEL binding site in the *iab-5* enhancer was shown to cause ectopic activation of *Abd-B* in A3 leading to A5 transformation [96]. Another gap gene product, *Hunchback*, represses the *Ubx* gene in regions of the embryo anterior to PS5 [92, 95], while the pair-rule gene *fushi tarazu* (*ftz*) is required to activate at least a subset of the *Ubx* enhancers [91, 92]. The general picture that has emerged from these analyses is that gap proteins are direct repressors of homeotic genes, whereas pair-rule proteins are direct activators. These two classes of proteins may compete for binding sites within the control region, such that the balance between their interactions at each *cis*-regulatory domain sets up the initial homeotic expression patterns [97].

Maintenance factors

Since the products of the gap and pair-rule genes are present only transiently in the early embryo, the activity state established during the initiation phase must be preserved by a maintenance system in each *cis*-regulatory domain. In simple terms, cells must remember which regulatory domain has to be kept in an active state and those that must be kept in a silent state. The expression of the BX-C in restricted patterns is required throughout life, so maintenance proteins are essential for normal [98] development. Also, the maintenance system must be stable through many cell divisions that occur between the time when homeotic expression patterns are initiated and the time when segmental differentiation occurs. The maintenance system involves two antagonistic sets of genes: the Polycomb group (PcG) and trithorax group (trxG) genes [99, 100]. The products of the PcG genes function as negative regulators, maintaining the inactive state of the homeotic genes, while the trxG gene products function as positive regulators, maintaining the active state. Mutations in PcG genes do not alter the initial selection of homeotic expression patterns, but instead cause their inappropriate de-repression later in embryogenesis [101–103]. trxG genes appear to be required for homeotic gene transcription during both the initiation and maintenance phase [98, 104, 105]. Mutations in trxG genes thus resemble LOF mutations in the ANT-C and BX-C [99]. Despite the fact that many PcG and trxG mutants have homeotic phenotypes themselves, they are not solely dedicated to homeotic gene control; members of both classes are transcriptional regulators of many other genes as well [106–108]. The PcG and trxG protein complexes contain histone-modifying activity that helps in condensing the chromatin for repression or

opening of chromatin for activation, respectively [109]. The DNA sequences to which PcG or trxG proteins bind are termed PRE/TREs and are known to require non-coding RNAs to bring about the silencing or activating functions [110, 111]. All these *cis* and *trans* regulatory components ensure that the Hox genes are expressed within respective tissues at the proper time during development and comprise an integral part of gene regulation events occurring within such gene complexes.

Heart complex

The heart complex is another example of a gene complex that has been extensively studied in model organisms like *Drosophila* and *Tribolium castaneum*, and is also called the tinman complex (Tin-C). The Tin-C contains seven genes: *tinman* (*tin*), *bagpipe* (*bap*), *ladybird early* (*lbe*), *ladybird late* (*lbl*), *C15*, *slouched* (*slou*) and *Drop* (*Dr/Msh*) [16]. The use of comparative genomics has revealed that rapid chromosomal arrangements are known to occur in insects, leading to their ever increasing diversity in morphology and behavior. Unlike the human genome, insect genomes rarely retain similar linkage arrangements [112–114]. For example, the heart genes in *Drosophila*, flower beetle and honey bee are conserved, but have undergone multiple inversions and translocations within the gene complex (Fig. 6). The Tin-C genes are evolutionarily ancient and pre-date the Hox complex. They contain a series of NK homeobox genes. The Tin-C gene linkage has been conserved in protostomes such as flies, but is lost in deuterostomes [115]. All members of the Tin-C are involved in muscle cell differentiation in flies, and many of the mesodermal patterning functions of Tin-C genes are conserved among flies, annelids and vertebrates.

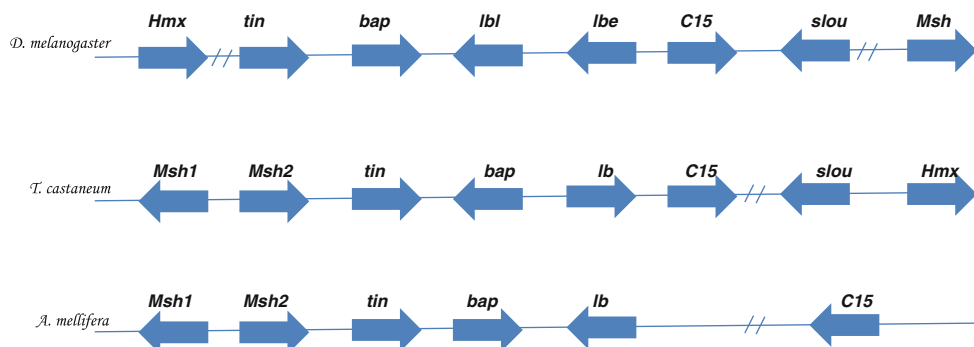


Fig. 6 Heart complex genes. The *Drosophila* heart complex contains eight NK-homoeobox genes and has two breakpoint between *Hmx-tin* and *slou-Msh* genes. The *Tribolium* heart complex also contains eight genes with duplication of the *Msh* genes and inversion of *bap* and breakpoint between *C15-slou* genes. The Honey bee heart complex

Comparative studies in *Drosophila* and *Tribolium* have recently shown that the ladybird gene is differentially expressed within the heart field. Ladybird is expressed in cardiac mesoderm of *Drosophila* and *Apis mellifera* (honeybee), and subdivides it into distinct pericardial and cardinal lineages [116, 117]. A striking dissimilarity is that *ladybird* expression is lost in the *Tribolium* heart field. Instead the *C15* gene is expressed and helps to pattern the heart in this insect. This intriguing replacement of ladybird expression with *C15* in *Tribolium* was mapped back to the altered enhancer-promoter interaction due to an inversion within the gene complex that bypasses an insulator located in the *ladybird-C15* region [16] (Fig. 7). The *lbe* gene promoter displays the paused/stalled polymerase not seen at the *lbl* and *C15* promoters, as shown by RNA Pol II ChIP-Seq experiments. Stalled Hox promoters have been shown to contain insulator activity [12]. Likewise it was shown that the *lbe* promoter could itself behave as an insulator, thus demarcating the cardiac enhancer in *Drosophila* [16]. In *Tribolium* the gene inversion event removes the insulator activity of the *lbe* promoter between the cardiac enhancer and *C15* promoter [16] (Fig. 7). These kinds of inversions may lead to redirection of conserved enhancers and might be an important mechanism of regulatory evolution.

How to identify and characterize CRMs

It is important to identify and characterize CRMs to understand how they function as switches of gene regulation. Traditionally the CRMs, such as enhancers, insulators and PREs, were isolated in genetic screens as they display dominant homeotic transformations, e.g., *Fab7* [66], *Fab8* [118] and *Mcp* [67] insulators. Few were recessive

displays only six genes with the absence of *slou* and *Hmx* and inversions of *bap*, *lb* and *C15*, as well as a breakpoint between *lb-C15* genes. This clearly demonstrates that gene complexes can evolve to give new distinct patterns of gene expression to accommodate the evolution of different species (adapted from [16])

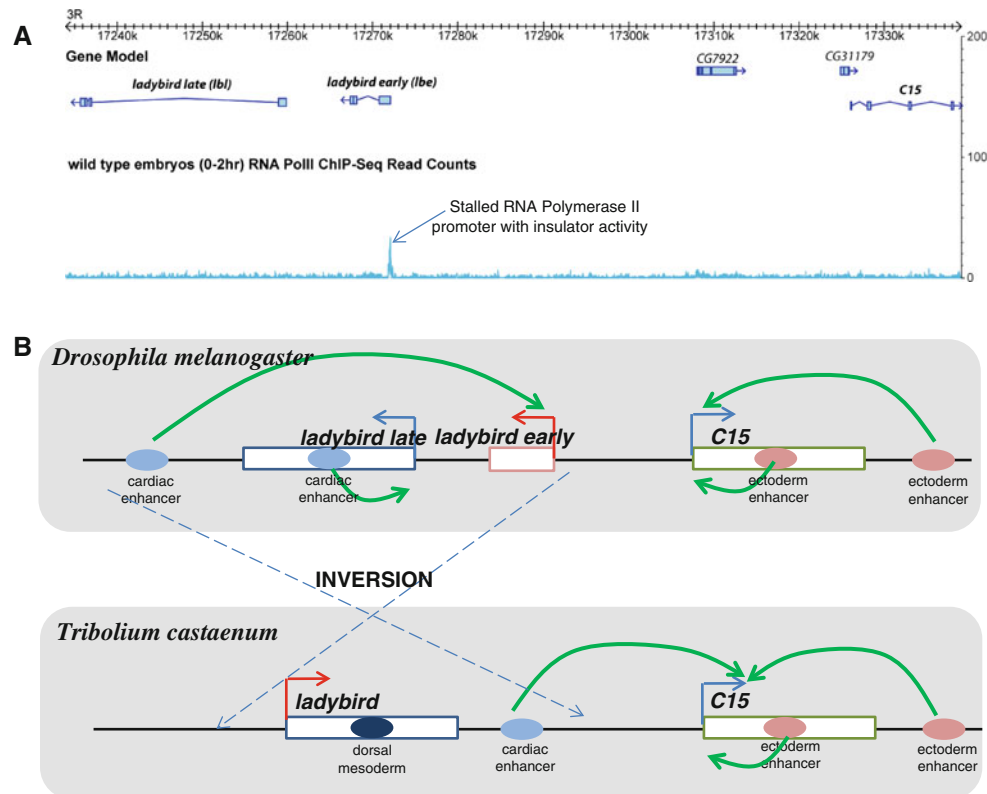


Fig. 7 Gene inversion within heart complex leads to novel patterning and insect evolution. **a** The RNA Pol II ChIP-Seq tracks within the *ladybird* locus from early *Drosophila* embryos (Chopra VS, Hendrix DA and Levine M. unpublished). The *lbe* promoter displays a strong stalled Pol II signal, whereas the *lbl* and *C15* promoters are not stalled. This stalled *lbe* promoter contains insulator activity in a transgenic assay [16]. **b** In *Drosophila*, the cardiac enhancer is located 3' of the *ladybird* early and *ladybird* late genes, and is unable to activate *C15* expression owing to insulator activity at the *ladybird* early promoter. The chromosomal inversion in *Tribolium* relocates

this enhancer so that the *ladybird* promoter is no longer positioned between the enhancer and *C15* gene. As a result, the cardiac enhancer is able to activate *C15* expression in *Tribolium*, but not *Drosophila*. Thus, a novel pattern of gene expression, *C15* expression in *Tribolium* pericardial and cardiac cells, is not due to the modification of gene regulatory networks or the de novo evolution of enhancer sequences, but rather results from the interaction of a conserved enhancer with different target genes: *ladybird* in *Drosophila* and the neighboring *C15* gene in *Tribolium* [16]

mutations and required sensitized mutant backgrounds as in the case of enhancers like *bxd*, *iab-7* and *iab-7PRE* [23, 119–122]. With the advent of biochemical and molecular biological techniques, these CRMs could be precisely mapped by DNaseI assays and in transgenic contexts [119]. The transgenic techniques also helped dissect the functions of *cis*-regulatory elements like enhancers [9], insulators [123–125], PREs [119, 126, 127], homing sequences [7, 21, 22] and PTS [10, 20] (Appendix 1). Gene conversion and transposon-mediated deletion techniques have immensely helped in dissecting functions of insulators and PREs in vivo [65, 118, 128, 129]. The recent advancement in BAC recombineering techniques could also aid the in vivo manipulation and characterization of CRMs [18, 19, 130].

Whole genome methods like ChIP-chip and ChIP-Seq have helped in generating profiles of nucleosomal proteins [131], histone marks [132–134], DNaseI [135], restriction enzyme accessibility [136], insulator proteins [137, 138],

PcG proteins [139] and tissue-specific transcription factors [140, 141]. This has remarkably helped identify regulatory codes underlying the function of gene complexes. ChIP-chip studies have shown that there are histone marks that preferentially mark regulatory elements (H3K4me1 that marks enhancers) [132], and cofactors like CBP/p300 also bind enhancers [142], leading to the identification of novel tissue-specific enhancers.

The discovery that stalled Hox and Heart promoters can function as insulator elements and demarcate functional boundaries of gene complexes would have been impossible without the availability of whole genome ChIP-chip and ChIP-Seq RNA Pol II profiles [12, 16, 143]. The Pol II ChIP-Seq data clearly showed that the first and last genes of the BX-C and Antp-C were stalled, whereas the genes within the complexes were not stalled. Thus, the stalled versus the non-stalled promoters were tested in transgenic enhancer blocking assays, and the stalled promoters displayed insulator activity. These findings are supported by the interaction

observed between endogenous insulators and their target promoters (example: *Abd-B* and *Fab7*) in vivo using the DamID technique [144] and in vitro using 3C assays [145]. ChIP-chip studies have also shown that a negative elongation factor (NELF) co-localizes with BEAF binding sites near promoters, again suggesting the link between boundary elements with stalled promoters [146, 147].

The current use of 3C techniques [148] at individual genes, as well as at the whole genome level, has taken higher order gene regulation understanding to the next level. It is now possible to visualize gene regulation events at the level of chromosomal topology using techniques like 5C [149], Hi-C [150], ChIA-PET [151] and FAIRE-Seq [152]. These techniques have helped to understand the regulation of mating type locus in yeast, human disease susceptibility loci, hormonal response regulation and stem cell fate maintenance. They generate complete genomic landscapes and represent long-range interactions as snapshots of a developmental window. These techniques have the capability to capture long-range chromosomal interactions like those of enhancer-promoter and insulator-promoter, and can be readily used to address mechanistic questions related to enhancer and insulator function. Perhaps with further advancement it may be possible to study transvection using these long-range genome-wide techniques. The future of such studies will be to try and resolve these genomic landscapes using high-resolution imaging techniques and visualize higher order gene regulation in real time and space within a cell.

Acknowledgments Thanks to Mike Levine for help and support. I would also like to acknowledge Eileen Wagner, Preeti Paliwal, Dave Hendrix, Kevin Tsui and anonymous reviewers for helpful insights during the writing process of this review. This review is dedicated to the memory of late Prof. Veronica Rodrigues, for her ever inspiring comments and feedback.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix 1. Transgenic techniques for identification of CRMs

Enhancer: Enhancers are *cis*-regulatory sequences that maybe present upstream, downstream or within (intronic) a gene, and help bind tissue-specific transcription factors that bring about enhanced levels of target gene transcription. Traditionally enhancers have been tested using transgenic reporter genes. A non-coding DNA fragment suspected to be an enhancer (oval) is cloned 5' to a lacZ or GFP reporter gene (rectangle) with a minimal promoter (bent arrow), and the plasmid is used to generate a transgenic model

organism. Within the transgenic organism the fragment drives the expression pattern of the reporter in a desired pattern in vivo confirming that it is an enhancer. Nowadays, enhancers are first predicted computationally using TF-binding site information from biochemical assays and whole genome assays (like ChIP-chip and ChIP-Seq) and then tested in transgenic contexts. Examples: *eve* stripe enhancers and *rho* NEE [9, 153].

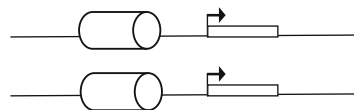


Polycomb response elements: PREs are cellular memory modules that bind PcG and trxG proteins, and help maintain the expression pattern of genes. The classic way to test a PRE is by Pairing Dependent-Silencing Assay. This assay uses the *white* (*w*) reporter gene (rectangle with arrow), which encodes for the red eye color. The PRE is placed 5' to a *w* reporter gene (cylinder), and it generally leads to a variegation phenotype in the eye. The *w* reporter gene is expressed in some ommatidia in the eye and repressed in some by PcG binding to the PRE under test, thus leading to this patchy expression. However, when these transgenic lines are allowed to homozygous (having two copies of the transgene), the *w* promoter shows less expression (lighter eye color) or is silenced completely (white eye color) compared to heterozygous (one copy of transgene). This is caused by PcG proteins forming higher order complexes on the paired PREs. This pairing-dependent silencing can be relieved in PcG mutant backgrounds [154–158]. Examples: *iab-7PRE*, *en-PRE* and *bx-d-PRE*.

Heterozygous (variegated white expression)

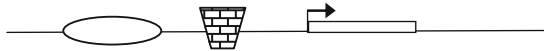


Homozygous (no white expression)



Insulator or boundary elements: These elements when placed between an enhancer (oval) and a promoter (bent arrow) disrupt the enhancer-promoter interaction and lead to loss of reporter gene expression. DNA sequences suspected to contain boundary activity (brick wall) are cloned in between a tissue-specific enhancer and reporter gene, and its ability to prevent the activation of the reporter is assayed in transgenic context. Insulators help demarcate domains of gene expression and prevent ectopic regulatory interactions from outside the domain. These elements also

sometimes prevent the spread of heterochromatin (barrier activity) and help maintain functional domains [12, 16, 123, 124]. Examples: *Fab7*, *Mcp* and *gypsy*.



Promoter targeting sequences: PTS are sequences found near promoter regions that facilitate the targeting of distal enhancers to the promoter containing the PTS. These sequences are found in genes that contain distal enhancers usually found downstream of neighboring genes. These sequences (diamond) when cloned between the insulator element and enhancer can lead to the activation of downstream reporter in spite of the insulator being present between enhancer and reporter. Hence, these elements are also known as anti-insulators as they can overcome even insulator elements and target the PTS containing promoter to the target gene. [10, 20].



References

- Elizondo LI, Jafar-Nejad P, Clewing JM, Boerkoel CF (2009) Gene clusters, molecular evolution and disease: a speculation. *Curr Genomics* 10:64–75
- Lewis EB (1998) The bithorax complex: the first fifty years. *Int J Dev Biol* 42:403–415
- Gruzdeva N, Kyrchanova O, Parshikov A, Kullyev A, Georgiev P (2005) The *Mcp* element from the bithorax complex contains an insulator that is capable of pairwise interactions and can facilitate enhancer-promoter communication. *Mol Cell Biol* 25:3682–3689
- Palstra RJ, de Laat W, Grosveld F (2008) Beta-globin regulation and long-range interactions. *Adv Genet* 61:107–142
- de Laat W, Klous P, Kooren J, Noordermeer D, Palstra RJ, Simonis M, Splinter E, Grosveld F (2008) Three-dimensional organization of gene expression in erythroid cells. *Curr Top Dev Biol* 82:117–139
- DeVido SK, Kwon D, Brown JL, Kassis JA (2008) The role of Polycomb-group response elements in regulation of engrailed transcription in *Drosophila*. *Development* 135:669–676
- Kwon D, Mucci D, Langlais KK, Americo JL, DeVido SK, Cheng Y, Kassis JA (2009) Enhancer-promoter communication at the *Drosophila* engrailed locus. *Development* 136:3067–3075
- Harding K, Rushlow C, Doyle HJ, Hoey T, Levine M (1986) Cross-regulatory interactions among pair-rule genes in *Drosophila*. *Science* 233:953–959
- Small S, Blair A, Levine M (1992) Regulation of even-skipped stripe 2 in the *Drosophila* embryo. *EMBO J* 11:4047–4057
- Akbari OS, Schiller BJ, Goetz SE, Ho MC, Bae E, Drewell RA (2007) The abdominal-B promoter tethering element mediates promoter-enhancer specificity at the *Drosophila* bithorax complex. *Fly (Austin)* 1:337–339
- Levine M (2010) Transcriptional enhancers in animal development and evolution. *Curr Biol* 20:R754–R763
- Chopra VS, Cande J, Hong JW, Levine M (2009) Stalled Hox promoters as chromosomal boundaries. *Genes Dev* 23:1505–1509
- Bushey AM, Dorman ER, Corces VG (2008) Chromatin insulators: regulatory mechanisms and epigenetic inheritance. *Mol Cell* 32:1–9
- Smith ST et al (2009) Genome wide ChIP-chip analyses reveal important roles for CTCF in *Drosophila* genome organization. *Dev Biol* 328:518–528
- Ringrose L, Paro R (2007) Polycomb/trithorax response elements and epigenetic memory of cell identity. *Development* 134:223–232
- Cande JD, Chopra VS, Levine M (2009) Evolving enhancer-promoter interactions within the tinman complex of the flour beetle, *Tribolium castaneum*. *Development* 136:3153–3160
- Hong JW, Hendrix DA, Levine MS (2008) Shadow enhancers as a source of evolutionary novelty. *Science* 321:1314
- Frankel N, Davis GK, Vargas D, Wang S, Payre F and Stern DL (2010) Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature*
- Perry MW, Boettiger AN, Bothma JP, Levine M (2010) Shadow enhancers foster robustness of *Drosophila* gastrulation. *Curr Biol* 20:1562–1567
- Zhou J, Levine M (1999) A novel *cis*-regulatory element, the PTS, mediates an anti-insulator activity in the *Drosophila* embryo. *Cell* 99:567–575
- Fujioka M, Wu X, Jaynes JB (2009) A chromatin insulator mediates transgene homing and very long-range enhancer-promoter communication. *Development* 136:3077–3087
- Bender W, Hudson A (2000) P element homing to the *Drosophila* bithorax complex. *Development* 127:3981–3992
- Maeda RK, Karch F (2006) The ABC of the BX-C: the bithorax complex explained. *Development* 133:1413–1422
- Kmita M, Duboule D (2003) Organizing axes in time and space; 25 years of colinear tinkering. *Science* 301:331–333
- Kaufman TC, Seeger MA, Olsen G (1990) Molecular and genetic organization of the antennapedia gene complex of *Drosophila melanogaster*. *Adv Genet* 27:309–362
- Kenyon C (1994) If birds can fly, why can't we? Homeotic genes and evolution. *Cell* 78:175–180
- Krumlauf R (1992) Transforming the Hox code. *Curr Biol* 2:641–643
- Lewis EB (1978) A gene complex controlling segmentation in *Drosophila*. *Nature* 276:565–570
- Duncan I (1987) The bithorax complex. *Annu Rev Genet* 21:285–319
- McGinnis W, Levine MS, Hafen E, Kuroiwa A, Gehring WJ (1984) A conserved DNA sequence in homeotic genes of the *Drosophila* Antennapedia and bithorax complexes. *Nature* 308:428–433
- McGinnis W, Garber RL, Wirz J, Kuroiwa A, Gehring WJ (1984) A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. *Cell* 37:403–408
- Carroll SB, Weatherbee SD, Langeland JA (1995) Homeotic genes and the regulation and evolution of insect wing number. *Nature* 375:58–61
- Carroll SB (1995) Homeotic genes and the evolution of arthropods and chordates. *Nature* 376:479–485
- Dolle P, Izpisua-Belmonte JC, Falkenstein H, Renucci A, Duboule D (1989) Coordinate expression of the murine Hox-5 complex homeobox-containing genes during limb pattern formation. *Nature* 342:767–772
- Dolle P, Duboule D (1989) Two gene members of the murine HOX-5 complex show regional and cell-type specific expression in developing limbs and gonads. *EMBO J* 8:1507–1515

36. Duboule D, Morata G (1994) Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet* 10:358–364
37. Izpisua-Belmonte JC, Tickle C, Dolle P, Wolpert L, Duboule D (1991) Expression of the homeobox Hox-4 genes and the specification of position in chick wing development. *Nature* 350:585–589
38. Gaunt SJ, Singh PB (1990) Homeogene expression patterns and chromosomal imprinting. *Trends Genet* 6:208–212
39. Gould A, Morrison A, Sproat G, White RA, Krumlauf R (1997) Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping Hox expression patterns. *Genes Dev* 11:900–913
40. Gerard M, Chen JY, Gronemeyer H, Chambon P, Duboule D, Zakany J (1996) In vivo targeted mutagenesis of a regulatory element required for positioning the Hoxd-11 and Hoxd-10 expression boundaries. *Genes Dev* 10:2326–2334
41. van der Hoeven F, Sordino P, Fraudeau N, Izpisua-Belmonte JC, Duboule D (1996) Teleost HoxD and HoxA genes: comparison with tetrapods and functional evolution of the HOXD complex. *Mech Dev* 54:9–21
42. Struhl G (1984) A universal genetic key to body plan? *Nature* 310:10–11
43. Tiong S, Bone LM, Whittle JR (1985) Recessive lethal mutations within the bithorax-complex in *Drosophila*. *Mol Gen Genet* 200:335–342
44. Negre B, Ranz JM, Casals F, Caceres M, Ruiz A (2003) A new split of the Hox gene complex in *Drosophila*: relocation and evolution of the gene labial. *Mol Biol Evol* 20:2042–2054
45. Lewis EB, Pfeiffer BD, Mathog DR, Celniker SE (2003) Evolution of the homeobox complex in the Diptera. *Curr Biol* 13:R587–R588
46. Krumlauf R (1994) Hox genes in vertebrate development. *Cell* 78:191–201
47. Graham A, Papalopulu N, Krumlauf R (1989) The murine and *Drosophila* homeobox gene complexes have common features of organization and expression. *Cell* 57:367–378
48. Sabarinadh C, Subramanian S, Tripathi A, Mishra RK (2004) Extreme conservation of noncoding DNA near HoxD complex of vertebrates. *BMC Genomics* 5:75
49. Krumlauf R, Holland PW, McVey JH, Hogan BL (1987) Developmental and spatial patterns of expression of the mouse homeobox gene, Hox 2.1. *Development* 99:603–617
50. Regulski M, Harding K, Kostriken R, Karch F, Levine M, McGinnis W (1985) Homeo box genes of the Antennapedia and bithorax complexes of *Drosophila*. *Cell* 43:71–80
51. Stornaiuolo A et al (1990) Human HOX genes are differentially activated by retinoic acid in embryonal carcinoma cells according to their position within the four loci. *Cell Differ Dev* 31:119–127
52. Simeone A, Acampora D, Nigro V, Faiella A, D'Esposito M, Stornaiuolo A, Mavilio F, Boncinelli E (1991) Differential regulation by retinoic acid of the homeobox genes of the four HOX loci in human embryonal carcinoma cells. *Mech Dev* 33:215–227
53. Marshall H, Studer M, Popperl H, Aparicio S, Kuroiwa A, Brenner S, Krumlauf R (1994) A conserved retinoic acid response element required for early expression of the homeobox gene Hoxb-1. *Nature* 370:567–571
54. Duncan I, Montgomery G (2002) E. B. Lewis and the bithorax complex. Part II. From *cis-trans* test to the genetic control of development. *Genetics* 161:1–10
55. Duncan I, Montgomery G (2002) E. B. Lewis and the bithorax complex. Part I. *Genetics* 160:1265–1272
56. Beachy PA, Helfand SL, Hogness DS (1985) Segmental distribution of bithorax complex proteins during *Drosophila* development. *Nature* 313:545–551
57. Celniker SE, Keelan DJ, Lewis EB (1989) The molecular genetics of the bithorax complex of *Drosophila*: characterization of the products of the Abdominal-B domain. *Genes Dev* 3:1424–1436
58. Karch F, Bender W, Weiffenbach B (1990) abdA expression in *Drosophila* embryos. *Genes Dev* 4:1573–1587
59. White RA, Wilcox M (1984) Protein products of the bithorax complex in *Drosophila*. *Cell* 39:163–171
60. White RA, Wilcox M (1985) Distribution of Ultrabithorax proteins in *Drosophila*. *EMBO J* 4:2035–2043
61. Zavortink M, Sakonju S (1989) The morphogenetic and regulatory functions of the *Drosophila* Abdominal-B gene are encoded in overlapping RNAs transcribed from separate promoters. *Genes Dev* 3:1969–1981
62. Casanova J, Sanchez-Herrero E, Morata G (1986) Identification and characterization of a parasegment specific regulatory element of the abdominal-B gene of *Drosophila*. *Cell* 47:627–636
63. Mihaly J et al (1998) Chromatin domain boundaries in the Bithorax complex. *Cell Mol Life Sci* 54:60–70
64. Galloni M, Gyurkovics H, Schedl P, Karch F (1993) The bluetail transposon: evidence for independent *cis*-regulatory domains and domain boundaries in the bithorax complex. *EMBO J* 12:1087–1097
65. Mihaly J, Hogga I, Gausz J, Gyurkovics H, Karch F (1997) In situ dissection of the Fab-7 region of the bithorax complex into a chromatin domain boundary and a Polycomb-response element. *Development* 124:1809–1820
66. Gyurkovics H, Gausz J, Kummer J, Karch F (1990) A new homeotic mutation in the *Drosophila* bithorax complex removes a boundary separating two domains of regulation. *EMBO J* 9:2579–2585
67. Karch F, Galloni M, Sipos L, Gausz J, Gyurkovics H, 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
68. Lewis EB (1985) Regulation of the genes of the bithorax complex in *Drosophila*. *Cold Spring Harb Symp Quant Biol* 50:155–164
69. Sipos L, Mihaly J, Karch F, Schedl P, Gausz J, Gyurkovics H (1998) Transvection in the *Drosophila* Abd-B domain: extensive upstream sequences are involved in anchoring distant *cis*-regulatory regions to the promoter. *Genetics* 149:1031–1050
70. Sipos L, Gyurkovics H (2005) Long-distance interactions between enhancers and promoters. *FEBS J* 272:3253–3259
71. Pirrotta V (1999) Transvection and chromosomal *trans*-interaction effects. *Biochim Biophys Acta* 1424:M1–M8
72. Gelbart WM, Wu CT (1982) Interactions of zeste mutations with loci exhibiting transvection effects in *Drosophila melanogaster*. *Genetics* 102:179–189
73. Hartl TA, Smith HF, Bosco G (2008) Chromosome alignment and transvection are antagonized by condensin II. *Science* 322:1384–1387
74. Duncan IW (2002) Transvection effects in *Drosophila*. *Annu Rev Genet* 36:521–556
75. Mathog D (1990) Transvection in the Ultrabithorax domain of the bithorax complex of *Drosophila melanogaster*. *Genetics* 125:371–382
76. Hopmann R, Duncan D, Duncan I (1995) Transvection in the iab-5, 6, 7 region of the bithorax complex of *Drosophila*: homology independent interactions in *trans*. *Genetics* 139:815–833
77. Selker EU (1990) Premeiotic instability of repeated sequences in *Neurospora crassa*. *Annu Rev Genet* 24:579–613
78. Selker EU (2002) Repeat-induced gene silencing in fungi. *Adv Genet* 46:439–450
79. Ronshaugen M, Levine M (2004) Visualization of *trans*-homolog enhancer–promoter interactions at the Abd-B Hox locus in the *Drosophila* embryo. *Dev Cell* 7:925–932

80. Lipshitz HD, Peattie DA, Hogness DS (1987) Novel transcripts from the Ultrabithorax domain of the bithorax complex. *Genes Dev* 1:307–322
81. Sanchez-Herrero E, Akam M (1989) Spatially ordered transcription of regulatory DNA in the bithorax complex of *Drosophila*. *Development* 107:321–329
82. Cavalli G, Paro R (1998) The *Drosophila* Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell* 93:505–518
83. Rank G, Prestel M, Paro R (2002) Transcription through intergenic chromosomal memory elements of the *Drosophila* bithorax complex correlates with an epigenetic switch. *Mol Cell Biol* 22:8026–8034
84. Gribnau J, Diderich K, Pruzina S, Calzolari R, Fraser P (2000) Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. *Mol Cell* 5:377–386
85. Bolland DJ, Wood AL, Johnston CM, Bunting SF, Morgan G, Chakalova L, Fraser PJ, Corcoran AE (2004) Antisense intergenic transcription in V(D)J recombination. *Nat Immunol* 5:630–637
86. Bae E, Calhoun VC, Levine M, Lewis EB, Drewell RA (2002) Characterization of the intergenic RNA profile at abdominal-A and Abdominal-B in the *Drosophila* bithorax complex. *Proc Natl Acad Sci USA* 99:16847–16852
87. Bender W, Fitzgerald DP (2002) Transcription activates repressed domains in the *Drosophila* bithorax complex. *Development* 129:4923–4930
88. Hogga I, Karch F (2002) Transcription through the *iab-7 cis-regulatory domain* of the bithorax complex interferes with maintenance of polycomb-mediated silencing. *Development* 129:4915–4922
89. Drewell RA, Bae E, Burr J, Lewis EB (2002) Transcription defines the embryonic domains of *cis-regulatory activity* at the *Drosophila* bithorax complex. *Proc Natl Acad Sci USA* 99:16853–16858
90. Casares F, Sanchez-Herrero E (1995) Regulation of the infra-abdominal regions of the bithorax complex of *Drosophila* by gap genes. *Development* 121:1855–1866
91. Muller J, Bienz M (1992) Sharp anterior boundary of homeotic gene expression conferred by the fushi tarazu protein. *EMBO J* 11:3653–3661
92. Qian S, Capovilla M, Pirrotta V (1991) The bx region enhancer, a distant *cis-control* element of the *Drosophila* Ubx gene and its regulation by hunchback and other segmentation genes. *EMBO J* 10:1415–1425
93. Qian S, Capovilla M, Pirrotta V (1993) Molecular mechanisms of pattern formation by the BRE enhancer of the Ubx gene. *EMBO J* 12:3865–3877
94. Shimell MJ, Simon J, Bender W, O'Connor MB (1994) Enhancer point mutation results in a homeotic transformation in *Drosophila*. *Science* 264:968–971
95. Zhang CC, Bienz M (1992) Segmental determination in *Drosophila* conferred by hunchback (hb), a repressor of the homeotic gene Ultrabithorax (Ubx). *Proc Natl Acad Sci USA* 89:7511–7515
96. Ho MC et al (2009) Functional evolution of *cis-regulatory* modules at a homeotic gene in *Drosophila*. *PLoS Genet* 5:e1000709
97. Bienz M (1992) Molecular mechanisms of determination in *Drosophila*. *Curr Opin Cell Biol* 4:955–961
98. Breen TR, Harte PJ (1991) Molecular characterization of the trithorax gene, a positive regulator of homeotic gene expression in *Drosophila*. *Mech Dev* 35:113–127
99. Kennison JA (1995) The Polycomb and trithorax group proteins of *Drosophila*: *trans-regulators* of homeotic gene function. *Annu Rev Genet* 29:289–303
100. Ringrose L, Paro R (2004) Epigenetic regulation of cellular memory by the polycomb and trithorax group proteins. *Annu Rev Genet* 38:413–443
101. Jones RS, Gelbart WM (1990) Genetic analysis of the enhancer of zeste locus and its role in gene regulation in *Drosophila melanogaster*. *Genetics* 126:185–199
102. Simon J, Chiang A, Bender W (1992) Ten different Polycomb group genes are required for spatial control of the abdA and AbdB homeotic products. *Development* 114:493–505
103. Struhl G, Akam M (1985) Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of *Drosophila*. *EMBO J* 4:3259–3264
104. Brizuela BJ, Elfring L, Ballard J, Tamkun JW, Kennison JA (1994) Genetic analysis of the brahma gene of *Drosophila melanogaster* and polytene chromosome subdivisions 72AB. *Genetics* 137:803–813
105. Ingham PW (1985) A clonal analysis of the requirement for the trithorax gene in the diversification of segments in *Drosophila*. *J Embryol Exp Morphol* 89:349–365
106. McKeon J, Slade E, Sinclair DA, Cheng N, Couling M, Brock HW (1994) Mutations in some Polycomb group genes of *Drosophila* interfere with regulation of segmentation genes. *Mol Gen Genet* 244:474–483
107. Moazed D, O'Farrell PH (1992) Maintenance of the engrailed expression pattern by Polycomb group genes in *Drosophila*. *Development* 116:805–810
108. Pelegri F, Lehmann R (1994) A role of polycomb group genes in the regulation of gap gene expression in *Drosophila*. *Genetics* 136:1341–1353
109. Simon JA, Kingston RE (2009) Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* 10:697–708
110. Gupta RA et al (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464:1071–1076
111. Sanchez-Elsner T, Gou D, Kremmer E, Sauer F (2006) Non-coding RNAs of trithorax response elements recruit *Drosophila* Ash1 to Ultrabithorax. *Science* 311:1118–1123
112. Bolshakov VN, Topalis P, Blass C, Kokoza E, della Torre A, Kafatos FC, Louis C (2002) A comparative genomic analysis of two distant diptera, the fruit fly, *Drosophila melanogaster*, and the malaria mosquito, *Anopheles gambiae*. *Genome Res* 12:57–66
113. Zdobnov EM, Bork P (2007) Quantification of insect genome divergence. *Trends Genet* 23:16–20
114. Zdobnov EM et al (2002) Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* 298:149–159
115. Luke GN, Castro LF, McLay K, Bird C, Coulson A, Holland PW (2003) Dispersal of NK homeobox gene clusters in amphioxus and humans. *Proc Natl Acad Sci USA* 100:5292–5295
116. Jagla K, Frasch M, Jagla T, Dretzen G, Bellard F, Bellard M (1997) Ladybird, a new component of the cardiogenic pathway in *Drosophila* required for diversification of heart precursors. *Development* 124:3471–3479
117. Jagla K, Jagla T, Heitzler P, Dretzen G, Bellard F, Bellard M (1997) Ladybird, a tandem of homeobox genes that maintain late wingless expression in terminal and dorsal epidermis of the *Drosophila* embryo. *Development* 124:91–100
118. Barges S et al (2000) The Fab-8 boundary defines the distal limit of the bithorax complex *iab-7* domain and insulates *iab-7* from initiation elements and a PRE in the adjacent *iab-8* domain. *Development* 127:779–790
119. Mishra RK, Mihaly J, Barges S, Spierer A, Karch F, Hagstrom K, Schweinsberg SE, 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
120. Morata G, Kerridge S (1981) Sequential functions of the bithorax complex of *Drosophila*. *Nature* 290:778–781
 121. Sanchez-Herrero E, Vernos I, Marco R, Morata G (1985) Genetic organization of *Drosophila* bithorax complex. *Nature* 313:108–113
 122. Mishra RK, Yamagishi T, Vasanthi D, Ohtsuka C, Kondo T (2007) Involvement of polycomb-group genes in establishing HoxD temporal colinearity. *Genesis* 45:570–576
 123. Cai H, Levine M (1995) Modulation of enhancer–promoter interactions by insulators in the *Drosophila* embryo. *Nature* 376:533–536
 124. Zhou J, Barolo S, Szymanski P, Levine M (1996) The Fab-7 element of the bithorax complex attenuates enhancer–promoter interactions in the *Drosophila* embryo. *Genes Dev* 10:3195–3201
 125. Schweinsberg SE, Schedl P (2004) Developmental modulation of Fab-7 boundary function. *Development* 131:4743–4749
 126. Mishra K, Chopra VS, Srinivasan A, Mishra RK (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
 127. Vasanthi D, Mishra RK (2008) Epigenetic regulation of genes during development: a conserved theme from flies to mammals. *J Genet Genomics* 35:413–429
 128. Mihaly J et al (2006) Dissecting the regulatory landscape of the Abd-B gene of the bithorax complex. *Development* 133:2983–2993
 129. Hogga I, Mihaly J, Barges S, Karch F (2001) Replacement of Fab-7 by the gypsy or scs insulator disrupts long-distance regulatory interactions in the Abd-B gene of the bithorax complex. *Mol Cell* 8:1145–1151
 130. Venken KJ, He Y, Hoskins RA, Bellen HJ (2006) P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science* 314:1747–1751
 131. Jiang C, Pugh BF (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* 10:161–172
 132. Heintzman ND et al (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459:108–112
 133. Heintzman ND, Ren B (2009) Finding distal regulatory elements in the human genome. *Curr Opin Genet Dev* 19:541–549
 134. Farnham PJ (2009) Insights from genomic profiling of transcription factors. *Nat Rev Genet* 10:605–616
 135. Song, L. and Crawford, G.E. (2010) DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *CSH Protoc* 2010, pdb prot5384
 136. Gargiulo G et al (2009) NA-Seq: a discovery tool for the analysis of chromatin structure and dynamics during differentiation. *Dev Cell* 16:466–481
 137. Negre N et al (2010) A comprehensive map of insulator elements for the *Drosophila* genome. *PLoS Genet* 6:e1000814
 138. Bushey AM, Ramos E, Corces VG (2009) Three subclasses of a *Drosophila* insulator show distinct and cell type-specific genomic distributions. *Genes Dev* 23:1338–1350
 139. Schwartz YB, Kahn TG, Nix DA, Li XY, Bourgon R, Biggin M, Pirrotta V (2006) Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nat Genet* 38:700–705
 140. Zeitlinger J, Zinzen RP, Stark A, Kellis M, Zhang H, Young RA, Levine M (2007) Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the *Drosophila* embryo. *Genes Dev* 21:385–390
 141. Stathopoulos A, Van Drenth M, Erives A, Markstein M, Levine M (2002) Whole-genome analysis of dorsal–ventral patterning in the *Drosophila* embryo. *Cell* 111:687–701
 142. Visel A et al (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457:854–858
 143. Core LJ, Lis JT (2009) Paused Pol II captures enhancer activity and acts as a potent insulator. *Genes Dev* 23:1606–1612
 144. Cleard F, Moshkin Y, Karch F, Maeda RK (2006) Probing long-distance regulatory interactions in the *Drosophila melanogaster* bithorax complex using Dam identification. *Nat Genet* 38:931–935
 145. Lanzuolo C, Roure V, Dekker J, Bantignies F, Orlando V (2007) Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat Cell Biol* 9:1167–1174
 146. Jiang N, Emberly E, Cuvier O, Hart CM (2009) Genome-wide mapping of BEAF binding sites in *Drosophila* links BEAF to transcription. *Mol Cell Biol* 29:3556–3568
 147. Raab JR, Kamakaka RT (2010) Insulators and promoters: closer than we think. *Nat Rev Genet* 11:439–446
 148. Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. *Science* 295:1306–1311
 149. Dostie J et al (2006) Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res* 16:1299–1309
 150. Lieberman-Aiden E et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326:289–293
 151. Li G et al (2010) ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol* 11:R22
 152. Gaulton KJ et al (2010) A map of open chromatin in human pancreatic islets. *Nat Genet* 42:255–259
 153. Ip YT, Park RE, Kosman D, Bier E, Levine M (1992) The dorsal gradient morphogen regulates stripes of rhomboid expression in the presumptive neuroectoderm of the *Drosophila* embryo. *Genes Dev* 6:1728–1739
 154. Chan CS, Rastelli L, Pirrotta V (1994) A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression. *EMBO J* 13:2553–2564
 155. Gindhart JG Jr, Kaufman TC (1995) Identification of polycomb and trithorax group responsive elements in the regulatory region of the *Drosophila* homeotic gene sex combs reduced. *Genetics* 139:797–814
 156. Hagstrom K, Muller M, Schedl P (1997) A Polycomb and GAGA dependent silencer adjoins the Fab-7 boundary in the *Drosophila* bithorax complex. *Genetics* 146:1365–1380
 157. Kassis JA (1994) Unusual properties of regulatory DNA from the *Drosophila* engrailed gene: three “pairing-sensitive” sites within a 1.6-kb region. *Genetics* 136:1025–1038
 158. Pirrotta V, Rastelli L (1994) White gene expression, repressive chromatin domains and homeotic gene regulation in *Drosophila*. *Bioessays* 16:549–556