Immunome Research
ISSN: 1745-7580

Bansal and Upadhyay, Immunome Res 2017, 13:3
DOI: 10.4172/1745-7580.1000143

Research Article Open Access

Genome-Wide Identification and Analysis of Putative *Rhomboid* Gene Enhancers in Multiple *Drosophila* Species

Anu Bansal and Atul Kumar Upadhyay*
School of Bioengineering and Biosciences, Lovely Professional University, Phagwara, 144411, India

Abstract

Enhancers are DNA sequences containing multiple transcription factor binding sites, could be present upstream, downstream or within the gene. They are meant to enhance the expression of their target genes. Major features of enhancers are: clusters of transcription factor binding sites, evolutionary conserved non-coding DNA sequence and biochemical marks such as histone modification and chromatin structure. We have used genome-wide approach to identify sequence of the putative enhancer of *rhomboid* gene in eleven different species of *Drosophila* viz. *melanoagaster*, *yakuba*, *ananassae*, *erecta*, *grimshawi*, *mojavensis*, *persimilis*, *pseudoobscura*, *sechellia*, *virilise* and *willistoni* using genome browser ClusterDraw. Analyses of the cluster formation of dorsal, twist and snail motifs of the *rhomboid* gene suggested a range of 1 to 15 binding sites among the species. Among the eleven species used in this study for prediction of enhancer element, four have enhancer element at upstream and five have downstream whereas two species have on both side of the *rhomboid* gene.

Keywords: Enhancer, *Rhomboid* gene, Dorsal, Twist, Snail

Introduction

Animal development starts from a single fertilised egg cell, which gives rise to different cells, tissues and organs. A Single cell develops into a complex multicellular organism in a remarkable process of development mediated by extensive cell division and differentiation into diverse cell types. The specification of unique cell types, their subsequent differentiation, and their different functions are determined by carefully orchestrated regulation of gene expression. Transcription factors (TFs) regulate gene expression by acting on associated DNA regulatory elements in the genome [1]. These cis-regulatory elements (CRMs) include promoters, repressors, insulators, and enhancers. Enhancers are cis-regulatory elements that enable precise spatiotemporal patterns of gene expression during development, able to function at large distance from their target genes and comprise binding sites for multiple transcription factors [2]. Enhancers have sequence-specific transcription factors bound to it and nucleosome region of open chromatin. Poised enhancers are primed enhancers containing repressive epigenetic chromatin mark a state [3]. DNA sequencing technology coupled to both chromatin immunoprecipitation (ChIP-Seq) and DNAse hypersensitivity assays (DNAse-Seq) can predict an enhancers on a genome wide scale [4,5]. Enhancers of a given cell or tissue type can now be easily identified throughout the genome using these assays according to their epigenetic features such as histone and chromatin modifications. Nucleosomes of enhancer elements are enriched for mono-methylation at lysine 4 of histone 3 (H3K4me1) while nucleosomes near the promoters tend to have tri-methylated H3K4 (H3K4me3) [6].

The model organism for this study is *Drosophila,* which is smaller in size compare to housefly. It can be cultured easily in a mass and has a short life time. The complete genome of fly was sequenced in 2000 contains approximately 13,600 genes [7]. It has 4 pairs of chromosomes: 3autosomal and one sex chromosome.

Rhomboid (*rho*) gene encodes a putative transmembrane receptor that is required for the differentiation of the ventral epidermis. It is initially expressed in embryo before the completion of cellularization within the presumptive neuroectoderm. Maternal morphogen Dorsal (dl) including other gene called *twist* (*twi*) activate expression of *rho* in both lateral and ventral regions. Expression is blocked in ventral regions (the presumptive mesoderm) by *snail* (*sna*) [8].

Enhancers are the DNA sequences that contain multiple binding sites for TFs known as transcription factor binding sites (TFBSs). Cluster of transcription factors is identified by using scoring matrix called position-weight matrix (PWM) [9]. In this study we have used ClusterDraw, which is one of the programs aimed to identify binding sites and binding site clusters for transcription factors. It filters overlapping binding sites by selecting those producing the best statistical scores [10]. It searches for the best clusters in the parameter space defined by the motif match P-value and the significance of the cluster. Many of the studies had been done using this tool for instance distribution of enhancer specific binding motifs was predicted responsible for anterior-posterior and dorsal-ventral patterning of *Drosophila* embryo [11]. Normal gene expression is produce by overlapping activities of multiple enhancers present in the gap gene, which was analysed by ChIP-chip method. Multiple distal and proximal enhancer of *hunchback* (*hb*) gene in combination to gap gene produces authentic expression pattern in *Drosophila* embryo [12].

There are multiple enhancers found in the activation of gene transcription. Primary enhancer contains multiple binding sites for transcription factors enhance the activity of gene expression. Secondary enhancer located at more remote position from the target gene mediates the expression of target gene by overlapping the primary enhancer. This secondary enhancer is known as shadow enhancer. One of the first instances of genes having multiple enhancers was reported in case of *shavenbaby* gene which directs the development of larval trichomes in *Drosophila*. *Svb* possesses three known enhancers- 7, E and A, located 50kb upstream of the gene [13]. Additional *svb* enhancer were located by constructing two reporter targets, resulting in identification of Z

***Corresponding author:** Atul Kumar Upadhyay, School of Bioengineering and Biosciences, Lovely Professional University, Phagwara, 144411, India, Tel: 8088810529; E-mail: atul4606iabt@gmail.com

Received October 06, 2016; **Accepted** October 20, 2017; **Published** October 25, 2017

Citation: Bansal A, Upadhyay AK (2017) Genome-Wide Identification and Analysis of Putative *Rhomboid* Gene Enhancers in Multiple *Drosophila* Species. Immunome Res 13: 143. doi: 10.4172/17457580.1000143

and DG2 which overlaps the expression of 7, E and A, and allow these 3 enhancers to develop larval trichomes in embryo under optimal temperature [14].

Materials and Methods

Sequence retrieval

Rhomboid gene and flanking sequences FASTA format, which is an online database. were obtained from Flybase (http://flybase.org/) version FB2015_02 in

Prediction of putative enhancer

Gene sequence along with flanking upstream and downstream sequences retrieved from flybase was uploaded to online server ClusterDraw (http://line.bioinfolab.net/webgate/submit.cgi) version 2.55 to predict binding sites and binding sites clusters. Sequence of binding motifs "Dorsal", "Twist" and "Snail" were uploaded from the collection of "fly development" motifs (available at http://line.bioinfolab.net/webgate/help/dxp.htm) in FASTA format. For *D. melanogaster* ClusterDraw searches for minimum of one motif combination and 5 cluster significance. Uploaded gene and motifs sequence were submitted and result was visualized as a 2D graph, showing the location of gene, sequence of binding site. These results can be extracted in FASTA format and number of binding motifs are identified on submitted sequence.

Sequence alignment

Predicted enhancer sequence of *D.melanogaster rhomboid* gene from ClusterDraw was analysed with the sequence validated in Ip journal in online web server ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/). It is software to visualize multiple sequence alignment. Parameters used in this software are gap penalties and weight matrix (PAM250 and BLOSUM62 matrix) [15]. It aligns the most closely related sequence from the phylogeny by calculating the number of k – tuple (combination of small words of certain size) matches. K-tuples value stands for the length of the word used to search for identity. Typical values of k-tuple are 1 or 2 for proteins and 3 for DNA. Both sequences were uploaded in the given sequence enter bar in the prescribed format. Default setting for alignment is slow type and sequences were submitted to obtain the results.

Identification of epigenetic features

Epigenetic features such as histone modifications can be visualized in UCSC genome browser (https://genome.ucsc.edu/). It provides reference sequence for a large collection of genomes. *D. melanogaster* genome was selected (Apr.2004 (BDGPR4/dm2)), *rhomboid* gene was entered in the search bar and details were submitted, result appeared in graphic form showing promoter, Dorsal, Twist and Snail binding predicted by ChIP-chip method at different stages of *Drosophila* embryo.

Results and Discussion

The sequence of *rho* gene was taken from Flybase (http://flybase.org/). The sequence of *rho* gene in *D. melanogaster* located is from 1463811-1468944 on 3L chromosome with its various regulatory elements. To predict putative enhancer for *rho* gene, 20 kb upstream and downstream sequence was analyzed. As previously described regulators for *rho* gene enhancer are Dorsal, Twist and Snail, cluster formation of these motifs at binding sites of gene enhancers was analysed using database ClusterDraw (http://line.bioinfolab.net/webgate/submit.cgi). Binding sites cluster was shown as colored peaks in which red color peak shows the highest cluster significance (Figure 1).

To analyse the results obtained from ClusterDraw, the predicted enhancer sequence of *Drosophila melanogaster,* were aligned with the validated enhancer of *rhomboid* gene of *Drosophila melanogaster* [16] by using ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Result showed that validated enhancer sequence of *rho* gene lies within the sequence of putative enhancer sequence resulted from ClusterDraw (Figure 2).

Enhancer sequences are consisting of multiple binding sites for transcription binding factors. Online genome browser UCSC (https://genome.ucsc.edu/) is used for visualization of binding sites for polymerase, Dorsal, Twist and Snail in predicted putative enhancer sequences of rhomboid gene (Figure 3).

Similarly, 20 kb upstream and downstream sequence of *rhomboid* gene of all other *Drosophila* species viz., *D.sechellia, D.erecta, D.grimshawi, D.pseudoobscura, D.yakuba, D.willistoni, D.virilis, D.mojavensis, D.ananassae* and *D.persimilis* were taken and cluster of binding sites of motifs were analysed. Clusters were seen as coloured peaks in which red peaks are more significant (Figure 4).

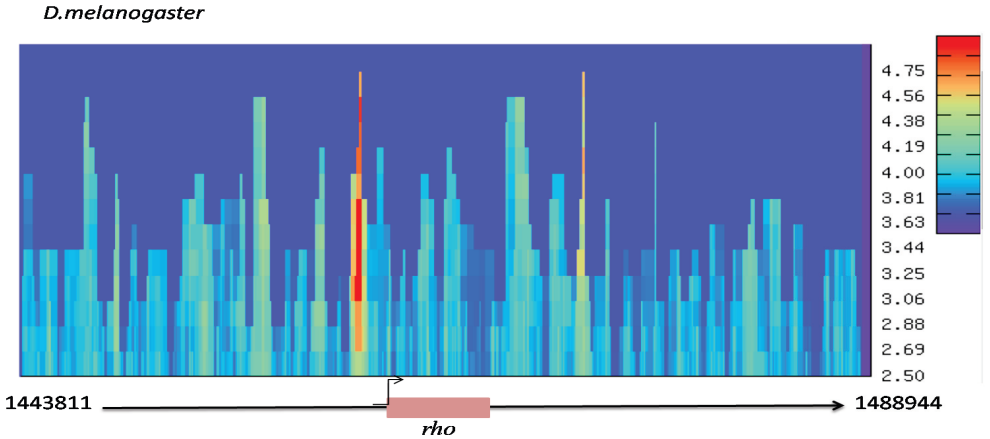

Figure 1: ClusterDraw output: 2D plot of cluster E-values using *drosophila* binding motifs Dorsal, Twist and Snail. X-axis shows the location of gene and 20kb flanking sequence and Y-axis defined the match P–value. Red peaks show the position of putative enhancer in all *Drosophila melanogaster.*

```
CLUSTAL 2.1 multiple sequence alignment

Dmel_rho_E1     TGGCATTTCCATTGTTATTTGTTTGTTTGCAAATTTGTTCAAGAAGTTGTCGTTATATTC 60
Ip_1992         ------------------------------------------------------------

Dmel_rho_E1     TATTTCGCAAGCTTTTCCTCTGCTCAAAATCAAAATGATTAAAACAACAGTTTGATACGA 120
Ip_1992         ------------------------------------------------------------

Dmel_rho_E1     ATTTTAATTCCCCTTTTTGCTGCGGAGTCAGTTAAGTGAGTCGCTTTCAGGACTCAGGGC 180
Ip_1992         ------------------------------------------------------------

Dmel_rho_E1     ATCATCCAGATCGCACGATCCCATTTGCATCTGCCTTCTCAGAAGCTGCTTGAAAGACGC 240
Ip_1992         ------------------------------------------------------------

Dmel_rho_E1     GCCCCTGTCGGATGATTAGTGCTAAGATCCTTGGGCAGGATGGAAAAATGGGAAAACATG 300
Ip_1992         ----------------------------CTTGGGCAGGATGGAAAAATGGGAAAACATG 31
                                             ******************************

Dmel_rho_E1     CGGTGGGAAAAACACACATCGCGAAACATTTGGCGCAACTTGCGGAAGACAAGTGCGGCT 360
Ip_1992         CGGTGGGAAAAACACACATCGCGAAACATTTGGCGCAACTTGCGGAAGACAAGTGCGGCT 91
                ************************************************************

Dmel_rho_E1     GCAACAAAAAGTCGCGAAACGAAACTCTGGGAAGCGGAAAAAGGACACCTTGCTGTGCGG 420
Ip_1992         GCAACAAAAAGTCGCGAAACGAAACTCTGGGAAGCGGAAAAAGGACACCTTGCTGTGCGG 151
                ************************************************************

Dmel_rho_E1     CGGGAAGCGCAAGTGGCGGGCGGAATTTCCTGATTCGCGATGCCATGAGGCACTCGCATA 480
Ip_1992         CGGGAAGCGCAAGTGGCGGGCGGAATTTCCTGATTCGCGATGCCATGAGGCACTCGCATA 211
                ************************************************************

Dmel_rho_E1     TGTTGAGCACATGTTTTGGGGGAAATTCCCGGGCGACGGGCCAGGAATCAACGTCCTGTC 540
Ip_1992         TGTTGAGCACATGTTTTGGGGGAAATTCCCGGGCGACGGGCCAGGAATCAACGTCCTGTC 271
                ************************************************************

Dmel_rho_E1     CTGCGTGGGAAAAGCCCAC------------------------------------ 559
Ip_1992         CTGCGTGGGAAAAGCCCACGTCCTACCCACGCCCACTCGGTTACCTGAATTCGAGCT 328
                ******************
```

Figure 2: Screenshot of pairwise sequence alignment of putative enhancer after ClusterDraw analysis of 20kb upstream and downstream Drosophila melanogaster *ho* gene with the rho NEE identified in Ip et al., 1992. Alignment was done by ClustalW.

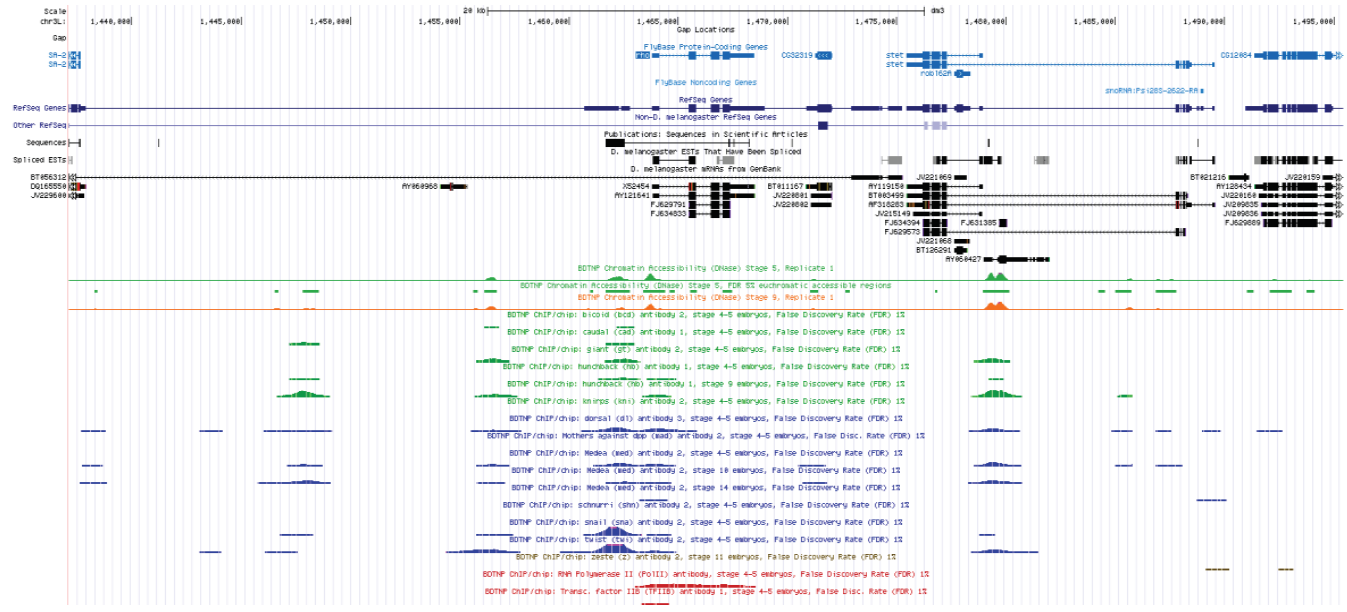

Figure 3: Screenshot of *rho* gene from UCSC browser visualizing binding sites for Polymerase, Dorsal, Twist and Snail at different embryonic stages.

Number of putative enhancers and binding sites ranges from 1-3 and 1-15, respectively (Table 1). *D. ananassae, D. persimilis, D. yakuba* have two putative enhancers. *D. erecta* have three pupative enhancer whereas other species viz., *D. melanogaster, D. grimshawi, D. mojavensis* and *D. pseudoobscura* have single putative enhancer element.

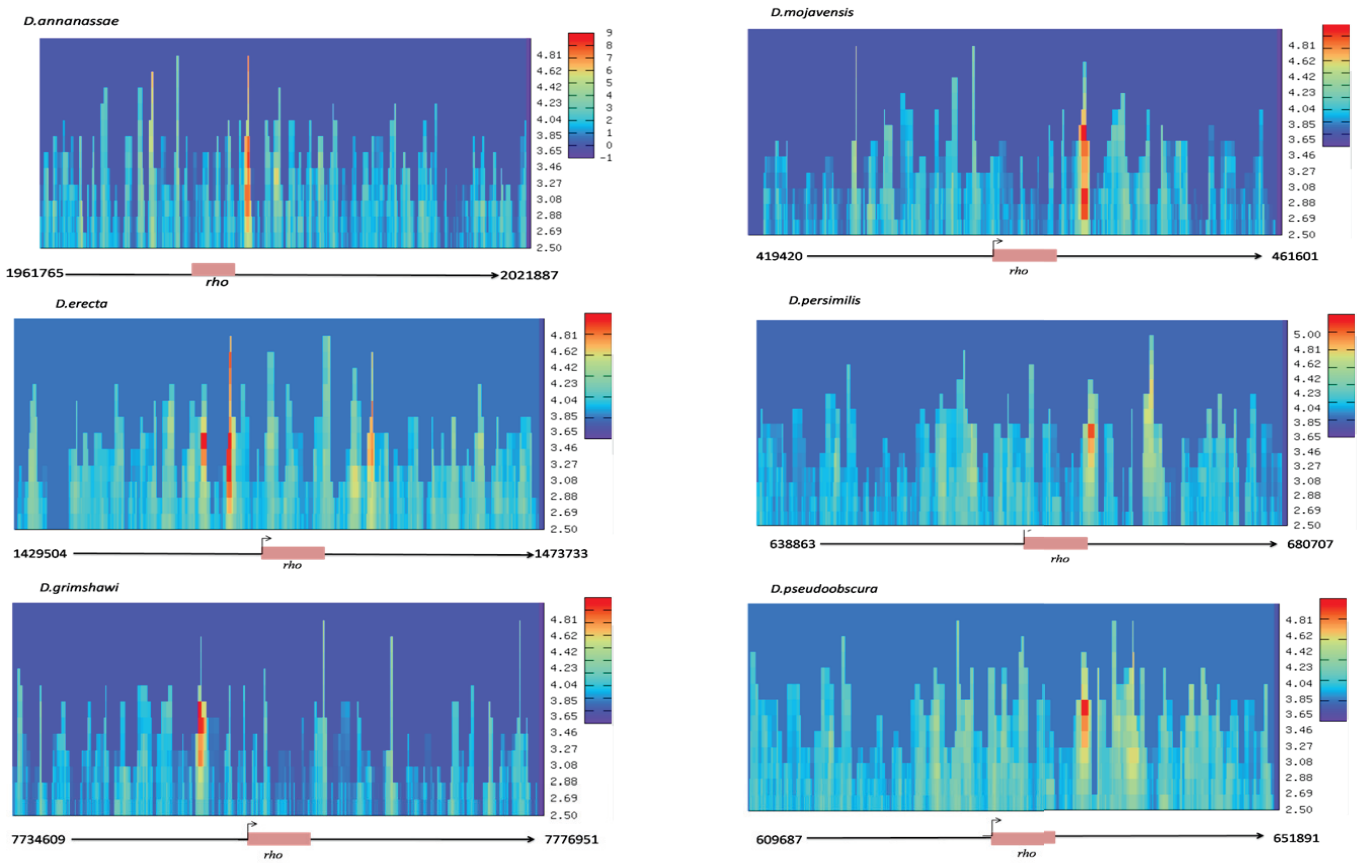

Figure 4: ClusterDraw output: 2D plot of cluster E-values using *drosophila* binding motifs Dorsal, Twist and Snail. X-axis shows the location of gene and 20 kb flanking sequence and Y-axis defined the match P – value. Red peaks show the position of putative enhancer in all *Drosophila* species.

Species	No. of putative Enhancers	Dorsal binding sites	Twist binding site	Snail binding site	Distance of enhancers peak from *rho* gene (in kb)
D. melanogaster	1	6	4	8	-2.272
D. ananassae	2	E_1–5 E_2–4	E_1– 6 E_2-3	E_1–9 E_2–8	E_1– +5.308 E_2 --6.575
D. erecta	3	E_1–5 E_2-4 E_3–5	E_1–2 E_2-4 E_3-4	E_1–12 E_2-7 E_3-10	E_1– -4.279 E_2 - -2.076 E_3 - +9.776
D. grimshawi	1	5	3	4	-5.093
D. mojavensis	1	9	7	11	+6.401
D. persimilis	2	E_1–15 E_2-9	E_1–7 E_2-1	E_1–15 E_2-11	E_1 +5.951 E_2 +10.793
D. pseudoobscura	1	12	4	10	+6.578
D. sechellia	1	6	4	9	-3.258
D. virilis	1	6	5	2	-5.602
D. willistoni	1	3	2	3	+7.985
D. yakuba	2	E_1–12 E_2–4	E_1.5 E_2-4	E_1–15 E_2–13	E_1 -2.416 E_2- -4.538

Table 1: Table shows the number of putative enhancer predicted from Cluster Draw in multiple species consisting of multiple Dorsal, Twist and Snail binding sites and distance of enhancer from *rho* gene.

From the analysis it is found that there are multiple binding sites for Dorsal, Twist and Snail and one or more than one enhancer are essential for the activation of expression of the gene and the estimated distance of enhancer shows that enhancer can be located upstream or downstream from the gene.

Conclusion

The *rhomboid* gene was first isolated from the follicle cell of the embryo and is activated initially at the early stage of the drosophila embryo (i.e. 0-4hr embryo) as regulatory elements required for the gene expression are found in follicle stage or regulatory elements which represses the gene expression found in late developmental stage of drosophila embryo in all the species [17] and expression is seen in the development of neuroectoderm in dorsal – ventral region. Dorsal, twist and snail together regulates the expression of gene.

In this study we have predicted enhancer elements binding motifs of the *rhomboid* gene. Analysis of clusters of binding sites of Dorsal, Twist and Snail in multiple species was done by using software ClusterDraw. It predicts the cluster of best binding sites. We have performed precition of these regulatory elements of the *rhomboid* gene among 11 species of *Drosophila*.

As a result of these predictions we reported that there are multiple binding sites for Dorsal, Twist and Snail motifs on the enhancer element of the genes. We also found that there are one or more than one enhancer elements are required for the activation of expression of the gene and the estimated distance of enhancer shows that enhancer can be located upstream or downstream from the gene.

Above discussed method predicts the binding sites of proteins on DNA segment at genomic level but to validate that a DNA segment on which transcription factors binds and regulates the expression of target

gene requires various experimental techniques such as: in vitro assays and transgenic assays in which activity of enhancer is determined by using reporter gene which is detected by luciferase or beta- galactosidase [18]. Other method used for validating enhancer activity are enhancer-FACS seq (eFC) method [19] and sequencing of transcribed barcode using plasmid based systems.

References

1. Hardison RC, Taylor J (2012) Genomic approaches towards finding cis-regulatory modules in animals. Nat Rev Genet 13: 469-483.
2. Smith E, Shilatifard A (2014) Enhancer biology and enhanceropathies. Nat Struct Mol Biol 21: 210-219.
3. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, et al. (2003) Super-enhancers in the control of cell identity and disease. Cell 155: 934-947.
4. Johnson DS, Mortazavi, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein- DNA interactions. Science 316: 1497-1502.
5. Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, et al. (2012) Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes Dev 26: 2604-2620.
6. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, et al. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 39: 311-318.
7. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, et al. (2004) The genome sequence of D.melanogaster. Science 287: 2185-2195.
8. Huang AM, Rusch J, Levine M (1997) An anteriorposterior Dorsal gradient in the Drosophila embryo. Genes Dev 15: 1963-1973.
9. Stormo GD (2000) DNA binding sites: representation and discovery. Bioinformatics 16: 16-23.
10. Papatsenko D (2007) ClusterDraw web server: a tool to identify and visualize clusters of binding motifs for transcription factors. Bioinformatics 23: 1032-1034.
11. Papatsenko D, Goltsev Y, Levine M (2009) Organization of developmental enhancers in the Drosophila embryo. Nucleic Acids Res 37: 5665-5667.
12. Perry MW, Boettiger AN, Levine M (2011) Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. PNAS 108: 13570-13575.
13. McGregor AP, Orgogozo V, Delon I, Zanet J, Srinivasan DG, et al. (2007) Morphological evolution through multiple cis regulatory mutations at a single gene. Nature 448: 587-590.
14. Nicolas F, Erezyilmaz DF, McGregor AP, Wang S, Payre F, et al. (2011) Morphological evolution caused by many subtle- effect substitutions in regulatory DNA. Nature 474: 598-603.
15. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680.
16. 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.
17. Nakamura Y, Kagesawa T, Nishikawa M, Hayashi Y, Kobayashi S, et al. (2007) Soma-dependent modulations contribute to divergence of rhomboid expression during evolution of Drosophila eggshell morphology. Development 134: 1529-1537.
18. Gisselbrecht SS, Barrera LA, Porsch M, Aboukhalil A, Estep PW, et al. (2013) Highly parallel assays of tissue-specific enhancers in whole Drosophila embryos. Nat Method 10: 774-780.
19. Sharon E, Kalma Y, Sharp A, Raveh-Sadka T, Levo M, et al. (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nat Biotech 30: 521-530.