

The *short gastrulation* shadow enhancer employs dual modes of transcriptional synergy

DONG-HYEON SHIN and JOUNG-WOO HONG*

Graduate School of East-West Medical Science, Kyung Hee University, Yongin, Korea

ABSTRACT It remains unclear how a limited amount of maternal transcription factor Dorsal (DI) directs broad expression of *short gastrulation* (*sog*) throughout the presumptive neurogenic ectoderm in the *Drosophila* early embryo. Here, we present evidence that the *sog* shadow enhancer employs dual modes of transcriptional synergy to produce this broad pattern. Bioinformatics analyses indicated that a minimal enhancer region, systematically mapped *in vivo*, contains five DI-, three Zelda (Zld)-, and three Bicoid (Bcd)-binding sites; four of these five DI-binding sites are closed linked to two Zld- and two Bcd-binding sites. Mutations of either the linked Zld- or Bcd-binding sites led to severe reduction in lacZ expression width, length, and/or strength in transgenic embryos. In addition, alteration of the helical phasing in this enhancer region by insertion of spacer sequences between linked sites also resulted in aberrant lacZ expression. These results suggest that synergistic interactions between DI and Zld and between DI and Bcd are required for broad *sog* expression.

KEY WORDS: *enhancer, short gastrulation, transcriptional synergy, zelda, bicoid*

Introduction

The dorsal-ventral (DV) patterning of the *Drosophila* early embryo is regulated by the maternal transcription factor *dl*, which is related to mammalian NFκB (Rushlow *et al.*, 1989). The sequence-specific activator DI forms a broad nuclear gradient, with peak levels in ventral regions and progressively diminishing levels in lateral and dorsal regions of the embryo (Anderson *et al.*, 1985). Basically, the DI gradient produces three discrete patterns of gene expression along the DV axis of cellularizing embryos (Hong *et al.*, 2008b). The highest levels of the DI gradient activate many of the DI target genes such as *twist* (*twi*) and *snail* (*sna*) in the ventral one-third of the embryo that differentiates into the presumptive mesoderm (Jiang *et al.*, 1991; Ip *et al.*, 1992). Intermediate levels of gradient direct expression of target genes including *brinker* (*brk*) in the ventral part of the presumptive neurogenic ectoderm (Markstein *et al.*, 2002). The lowest levels of the gradient activate genes such as *sog* throughout the neurogenic ectoderm (Markstein *et al.*, 2002). Intriguingly, the same levels of the DI protein function as a repressor to restrict expression of *decapentaplegic* (*dpp*), *zerknüllt* (*zen*), and *tolloid* (*tld*) in the dorsal ectoderm (Jiang *et al.*, 1992). Altogether, the DI gradient activates 60-70 target genes to initiate DV patterning in a concentration-dependent manner

(Hong *et al.*, 2008b).

DI protein does not work alone to establish differential expression patterns across the DV axis of early embryos. For instance, the highest levels of the DI gradient in the mesoderm activate a basic helix-loop-helix (bHLH) transcription activator *twi* (Jiang *et al.*, 1991), whose product also forms a nuclear gradient that is steeper than the DI gradient. DI and Twi cooperatively direct expression of a zinc-finger repressor *Sna* that establishes the dorsal limit of the presumptive mesoderm (Ip *et al.*, 1992). Another type of transcriptional synergy is observed in neurogenic gene expression. Enhancers directing gene expression in the ventral regions of the neurogenic ectoderm contain a fixed arrangement of the closely linked DI- and Twi-binding sites. Intermediate and low levels of DI and Twi, respectively, cooperatively occupy the linked sites to delimit threshold responses of the enhancers to the DI gradient and synergistically activate neurogenic genes such as *rhomboid* (*rho*)

Abbreviations used in this paper: AP, anterior-posterior; B1H, bacterial one-hybrid; bcd, bicoid; bHLH, basic helix-loop-helix; brk, brinker; DIG, digoxigenin; dl, dorsal; dpp, decapentaplegic; DV, dorsal-ventra; eve, even-skipped; ftz, fushi tarazu; PCR, polymerase chain reaction; rho, rhomboid; SELEX, systematic evolution of ligands by exponential enrichment; sna, snail; sog, short gastrulation; tld, tolloid; twi, twist; vnd, ventral nervous system defective; zen, zerknüllt; zld, zelda; β-IFN, β-interferon.

*Address correspondence to: Joungh-Woo Hong, Graduate School of East-West Medical Science, Kyung Hee University, Yongin, 17104, Korea.
Tel: +82-31-201-3853. Fax: +82-31-204-8119. E-mail: jwhong46@khu.ac.kr

Supplementary Material (8 tables + 3 figures) for this paper is available at: <http://dx.doi.org/10.1387/ijdb.160165jh>

Submitted: 19 May, 2016. Accepted: 5 August, 2016.

and *ventral nervous system defective* (*vnd*) (Zinzen *et al.*, 2006). These findings raise the possibility that synergistic interactions with *Twi* may also allow the lowest levels of *DI* to direct broad *sog* expression throughout the neurogenic ectoderm. However, a couple of studies presented evidence that the possibility may not be the case (Szymanski and Levine, 1995; Liberman and Stathopoulos, 2009), suggesting that *DI* may interact synergistically with other cooperative partners besides *Twi* to generate broad *sog* expression throughout the neurogenic ectoderm. However, the details of the mechanism are unclear.

The broad pattern of endogenous *sog* expression in the presump-

tive neurogenic ectoderm is controlled by two separate enhancers, called the primary enhancer and shadow enhancer (Hong *et al.*, 2008a). About 400-bp primary enhancer has been identified by computational analyses of the genome-wide distribution of the *DI* recognition sequence (Markstein *et al.*, 2002). Subsequent chromatin immunoprecipitation (ChIP)-chip assays have shown that the *sog* locus contains a secondary enhancer to direct *sog* expression in the neurogenic ectoderm (Hong *et al.*, 2008a). The two redundant enhancers are referred to as the “primary” and “shadow” enhancer, respectively, based on the chronological order of their identification rather than any functional differences.

Here, evidence is presented that the *sog* shadow enhancer employs transcriptional synergies between *DI* and *Zld* and between *DI* and *Bcd* to produce the broad stripes of *sog* expression. The *sog* minimal enhancer, systematically determined in transgenic embryos, contains five *DI*-binding sites, four of which are closely linked to two *Zld*- and two *Bcd*-binding sites. Mutation of both the linked *Zld*- and *Bcd*-binding sites and extension of the distance between the *DI*- and linked sites led to a dramatic reduction in *lacZ* expression in transgenic embryos. It is conceivable that the broad *sog* expression observed in the neurogenic ectoderm is at least in part due to synergistic interactions between *DI* and either *Zld* or *Bcd*.

Results

A 0.582 kb region within the shadow enhancer is sufficient to direct *sog* expression in the presumptive neurogenic ectoderm

To investigate how the *sog* shadow enhancer directs the broad pattern, the minimal region possessing enhancer activity was determined by 5' and 3' truncation analysis in transgenic embryos (Fig. 1). First, a 5' deletion set including four constructs (0.732, 0.616, 0.404, and

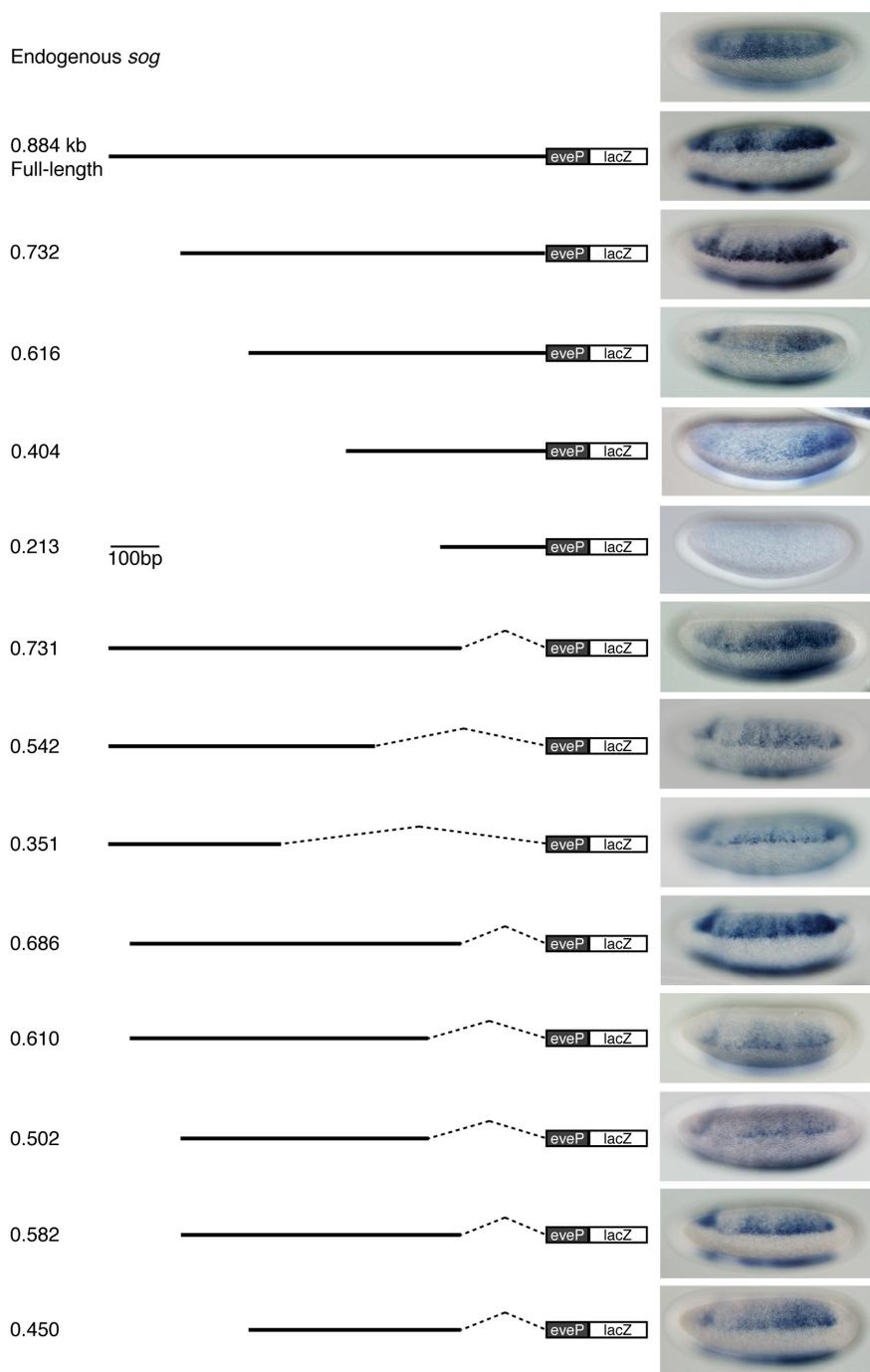


Fig. 1. A 0.582 kb region within the shadow enhancer is sufficient to direct *sog* expression in the presumptive neurogenic ectoderm. All genomic regions tested were PCR-amplified (Table S1), cloned upstream of a minimal *eve* promoter followed by a *lacZ* open reading frame (Small *et al.*, 1992) and injected into *Drosophila* early embryos for *Pe* element-mediated germline transformation. Cellularizing transgenic embryos (approximately early stage 5) are oriented with anterior to the left and dorsal up. Expression of *lacZ* in the transgenic embryos was visualized by in situ hybridization with an antisense *lacZ* RNA probe. The *lacZ* expression patterns directed by fragments of the *sog* shadow enhancers are shown at the right of each construct. Endogenous *sog* expression (top panel) is presented and was visualized with an antisense *sog* RNA probe. The *lacZ* expression patterns directed by several *sog* shadow enhancers seem to be in three bands. The banding pattern shows up from the late stage 5 of embryogenesis.

0.213 kb) was generated. The 0.732 kb construct directed broad and strong *lacZ* expression comparable to that mediated by the full-length (0.884 kb) construct. However, further truncation of the 0.732 kb construct (0.616 and 0.404 kb) led to severe reduction in *lacZ* expression, and the 0.213 kb construct did not activate the *lacZ* reporter. These results suggest that the 5' end of the minimal region is located somewhere between the 5' ends of the 0.732- and 0.616 kb constructs. A series of 3' deletion constructs (0.731, 0.542, and 0.351) were also created to map the 3' end of the minimal region. The 0.731 kb construct directed broad *lacZ* expression in the presumptive ectoderm. Two more deletion constructs (0.542 and 0.351 kb), however, showed severely impaired *lacZ* expression. These results suggest that the 3' end of the minimal region is located somewhere between the 3' ends of the 0.731- and 0.542 kb constructs. Five more constructs (0.686, 0.610, 0.502, 0.582, and 0.450) were produced to further refine the minimal region. The 0.686 kb construct directed *lacZ* expression similar to that mediated by the full-length, 0.732-, or 0.731 kb constructs.

Intriguingly, 3' ~70-bp deletion (0.610 kb) of the 0.686 kb construct resulted in a catastrophic reduction in *lacZ* expression, suggesting that this short region contains an essential element(s) required for *sog* enhancer activity. Consistently, the 0.502 kb construct derived from further 5' deletion of the 0.610 kb construct almost failed to activate the reporter gene. However, 3' ~70-bp extension of the 0.502 kb construct allowed restoration of the enhancer activity of the 0.582 kb construct. Although the 0.450 kb construct could still activate *lacZ*, *lacZ* expression was weaker in the anterior half of the embryo. Together, these results suggest that the 0.582 kb region within the *sog* shadow enhancer is necessary and sufficient to direct *lacZ* expression in a pattern almost congruent with the endogenous *sog* expression pattern.

The *sog* shadow minimal enhancer contains Doral (Dl)-binding sites closely linked to Zelda (Zld)- and Bicoid (Bcd)-binding sites

The previous findings that mutant embryos lacking either Zld or Bcd protein are defective in endogenous *sog* expression in their neurogenic ectoderm (Liang *et al.*, 2008; Papatsenko *et al.*, 2009) prompted us to hypothesize that Dl-binding sites form a statistically significant cluster(s) with Zld- and Bcd-binding sites within the minimal enhancer. *zld*, encoded by the X chromosomal gene *vielfältig (vfl)*, is a maternal C₂H₂ zinc finger transcription activator and expressed throughout early embryogenesis (Liang *et al.*, 2008). *bcd* is a maternal effect gene that encodes a transcriptional factor containing a homeodomain. Bcd protein forms a nuclear gradient with peak levels at the anterior pole and progressively diminishing levels in posterior regions and is involved in patterning along anterior-posterior (AP) axis of a developing early embryo (Porcher and Dostatni, 2010). To test this hypothesis, ClusterDraw

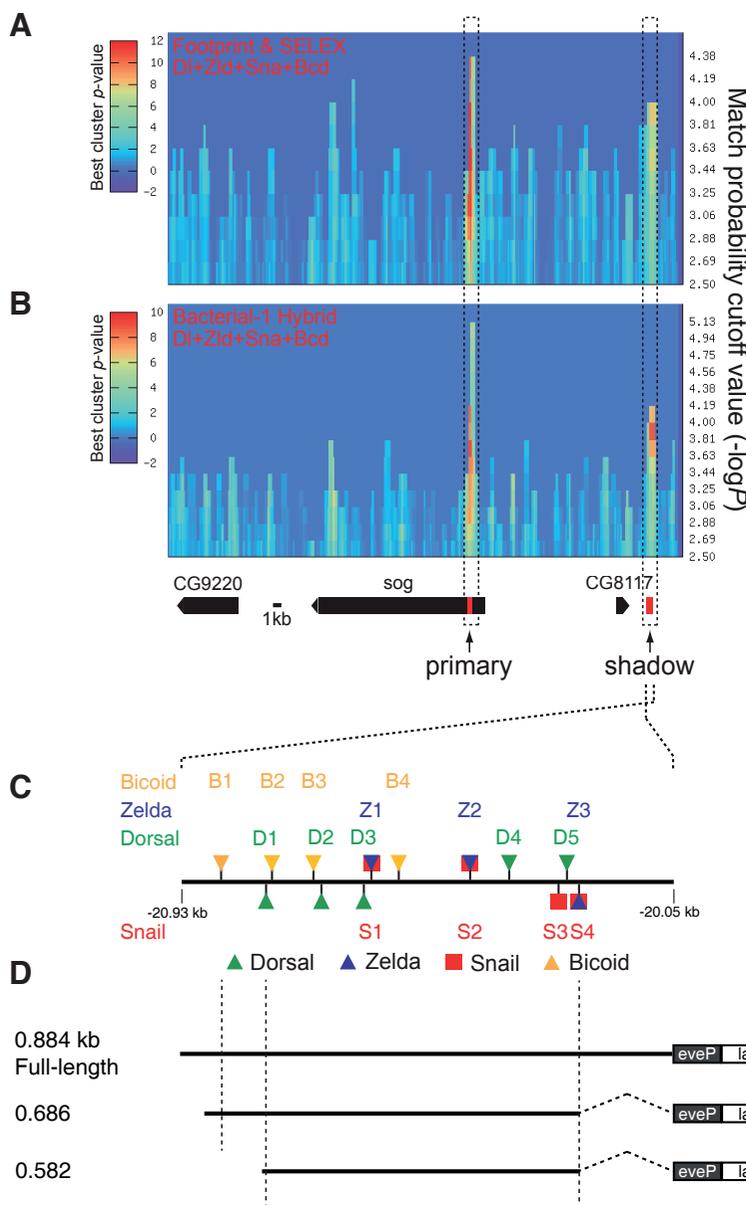


Fig. 2. ClusterDraw analysis of a ~62 kb genomic region encompassing the *sog* locus. ClusterDraw (Papatsenko, 2007) analyses were performed with two different types of PFMs for the Dl-, Zld-, Sna-, and Bcd-binding sequences. One was built by motif alignments obtained from in vitro binding data (<http://line.bioinfolab.net/webgate/help.htm#mttform>) (Papatsenko, 2007) (A) and the other from in vivo binding data (<http://mccb.umassmed.edu/ffs/>) (Zhu *et al.*, 2011) (B). Each analysis yielded two statistically significant best cluster p-values. Although the patterns of the best cluster p-values along the axis of match probability cutoff (-logP) (x axis) differed slightly between the two independent analyses, the two best clusters in each analysis coincided with the primary and shadow enhancers (dotted boxes, Table S8). Gene models over the 62 kb genomic region are depicted below (B). (C) ClusterDraw analyses also indicated the location and cumulative match probability (-logP) value of each motif found in the shadow enhancer (Table S2-S5). The thick line denotes ~880-bp of the full-length *sog* shadow enhancer (Table S8) (Hong *et al.*, 2008a). Triangles and squares shown above and below the line represent motifs identified in the sense and antisense strands relative to the transcription start sites of the *sog* gene, respectively. (D) Three fragments of the *sog* shadow enhancer whose enhancer activity were tested (Fig. 1) are shown to enable comparison of the locations of binding motifs for the four factors. All Dl-, Zld-, Sna-, and Bcd-binding motifs indicated by the ClusterDraw algorithm were located within the 0.582 kb minimal region. (C,D) have identical scales.

analysis were performed with position frequency matrices (PFMs) (Fig. S1 and S2) for DI-, Zld-, Sna-, and Bcd-binding sites (Fig. 2). The ClusterDraw program is an *r*-scan-based algorithm that has been used to identify the binding motifs and binding clusters of specific combinations of transcription factors (Papatsenko, 2007). To increase the statistical power of the *in silico* analysis, an identical ClusterDraw analysis was repeated twice with two different sets of PFMs for the four transcription factors. One set of PFMs was generated by motif alignments obtained from the *in vitro* binding data (Papatsenko, 2007) (Fig. 2A) and the other from the *in vivo* binding data (Zhu *et al.*, 2011) (Fig. 2B). Binding sites for the zinc-finger repressor Sna were included in these analyses because Sna restricts *sog* expression to the neurogenic ectoderm by repressing *sog* expression in the presumptive mesoderm (Cowden and Levine, 2002), suggesting that the minimal enhancer includes at least one high-quality Sna-binding site.

ClusterDraw analyses over a ~62 kb genomic region encompassing the *sog* locus identified two clusters repeatedly (Fig. 2A and B). Although the patterns of the best cluster *p*-values along the axis of match probability cutoff ($-\log P$) were slightly different in the two independent analyses, the two best clusters in each analysis coincided with the primary and shadow enhancers (Fig. 2A and B, dotted boxes, Table S8). These results suggest that the primary and shadow enhancers of *sog* contain the most significant clusters of DI-, Zld-, Sna-, and Bcd-binding motifs across the ~62 kb genomic region. The ClusterDraw algorithm also provides information about the location and quality [cumulative match probability ($-\log P$)] of each binding site found in the clusters that ClusterDraw has identified (Table S2-S5). Binding sites commonly identified by the two repeated analyses and whose cumulative match probability values were higher than the match probability cutoff value were defined as functional binding motifs (Fig. 2C and Tables S2-S5, see Supplementary Material for more details). The cluster corresponding to the shadow enhancer contained four DI-, three Zld-, four Sna-, and four Bcd-binding sites. Intriguingly, two DI-binding sites (D3 and D5) were closely linked with the two Zld-binding sites (Z1 and Z3) and another two DI-binding sites (D1 and D2) were located nearby the two Bcd-binding sites (B2 and B3). In addition, all binding sites were included in the minimal region of the *sog* shadow enhancer except the first Bcd-binding site (B1) (Fig. 2D). These results suggest that there may be synergistic interactions between DI and Zld and between DI and Bcd.

Close linkages between DI- and Zld-binding sites in the minimal enhancer are required for the broad lacZ expression

Next, we tested if the Zld-binding sites are indispensable for minimal enhancer activity (Fig. 3). The 0.582 kb construct directed broad and strong lacZ expression that recapitulated endogenous *sog* expression in the neurogenic ectoderm (Fig. 3A, compare with Fig. 1, top panel). However, mutations of all the Zld-binding sites resulted in dramatic reduction in lacZ expression width (Fig. 3B). In addition, mutations in the two Zld-binding sites (Z1 and Z3) located adjacent to DI-binding sites (D3 and D5) created a narrow pattern of lacZ expression similar to that mediated by the 0.582ΔZld123 construct (Fig. 3C). Moreover, removal of only either the first (Z1) (Fig. 3D) or third (Z3) (Fig. 3F) Zld-binding site in the minimal enhancer led to even narrower lacZ expression along the DV axis than that directed by the 0.582ΔZld123 construct. Mutation of the second Zld-binding sites (Z2) alone, however, directed

lacZ expression much broader than any lacZ patterns mediated by the mutant enhancers of the Zld-binding sites (Fig. 3E). These results suggest that the first and third Zld-binding sites (Z1 and Z3) are required for broad *sog* expression, whereas loss of the second Zld-binding site (Z2) is tolerable.

The fact that the Zld-binding sites required for the broad pattern were closely linked to DI-binding sites prompted us to reason

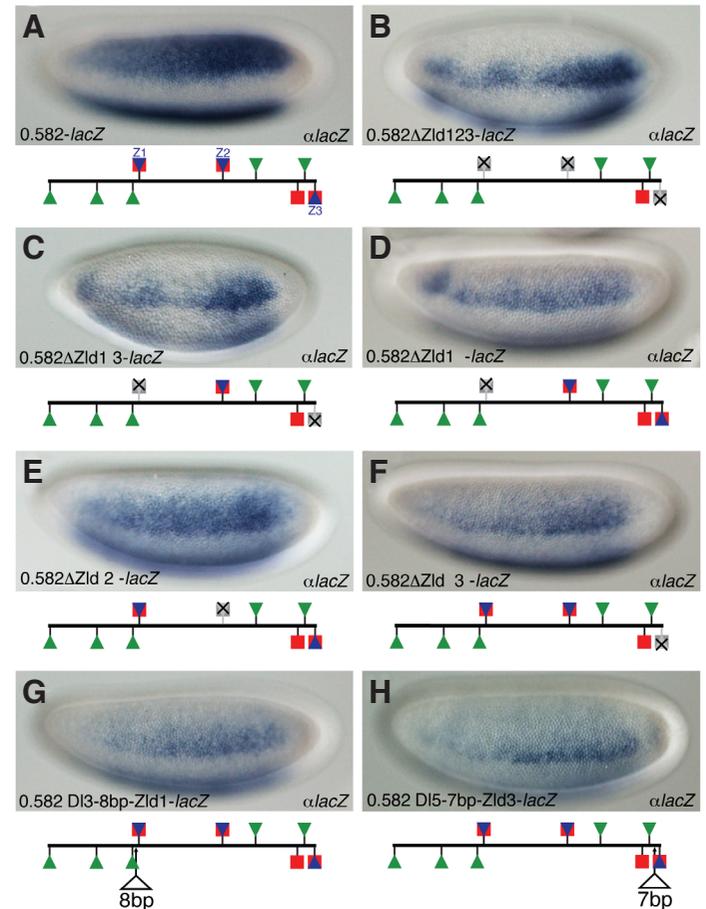


Fig. 3. Broad lacZ expression requires close linkages between DI- and Zld-binding sites in the *sog* shadow minimal enhancer. DI-, Zld-, and Sna-binding sites depicted in the 0.582 kb minimal region are shown below each panel. Changes in binding site sequence are shown with X marks. (A) The 0.582 kb construct activated lacZ expression that recapitulated endogenous *sog* expression in the presumptive neurogenic ectoderm. (B) Mutations in three Zld-binding sites (Z1, Z2 and Z3, Fig. 2C) led to a dramatic reduction in lacZ expression width along the DV axis. (C) Nucleotide substitutions in two Zld-binding sites (Z1 and Z3) linked with two nearby DI-binding sites (D3 and D5) directed lacZ expression analogous to that mediated by the 0.582ΔZld123 construct. (D-F) Loss of either the first (Z1) (D) or third (Z3) (F) Zld site in the 0.582 kb construct resulted in lacZ expression even narrower than that of the enhancer containing three impaired Zld sites (B). (E) However, mutation of the second Zld site (Z2) directed lacZ expression with a dorsal limit almost identical to that of the full-length enhancer. (G,H) To test if DI and Zld proteins interact cooperatively each other for transcriptional synergy, 8- and 7-bp nucleotides were inserted between D3 and Z1 (G) and between D5 and Z3 (H), respectively, to interfere with helical phasing. Extension of the distances between the two pairs of sites significantly reduced lacZ expression width to a similar extent as 0.582ΔZld1 and 0.582ΔZld3.

that DI and Zld proteins synergistically directed *lacZ* expression via cooperative occupation of linked binding sites. To test this reasoning, the helical phasing between the DI- and Zld-binding sites was altered by extending the distance between these sites. The relative stereospecific positioning, or binding face on helical DNA, of transcription activators is called “helical phasing” and is determined by the spacing between the binding sites for two factors. A strict spacing requirement between two factor binding sites for proper transcriptional activation indicates that proteins cooperatively occupy the binding sites and interact directly on the DNA (Zinzen *et al.*, 2006). Thus, even subtle changes in spacing can severely hamper cooperative site occupancy and direct interactions between two proteins. The centers of the two pairs of DI- and linked Zld-binding sites (D3-Z1 and D5-Z3) are separated by 12 bp and 21 bp, respectively, thereby raising the possibility that the two proteins interact on the almost same side of the DNA helix. Insertion of an 8-bp spacer sequence between the D3 and Z1 sites (Table S1) led to severe reduction in *lacZ* expression width along the DV axis, comparable to that directed by the enhancer containing either first or third mutant Zld-binding sites (Fig. 3G, compare with D and F). The addition of a 7-bp sequence between the D5 and Z3 sites resulted in *lacZ* expression similar to that directed by the 0.582DI3-8bp-Zld1 construct (Fig. 3H). These results suggest that a fixed arrangement of the DI- and linked Zld-binding sites is required for *lacZ* expression in the neurogenic ectoderm, where there are low levels of DI. There were no changes in the ventral limit of *lacZ* expression for any of the mutant enhancers tested, implying that the mutational changes did not interfere with short-range repression mediated by *Sna* in the presumptive mesoderm.

Close proximity between DI- and Bcd-binding sites in the minimal enhancer is required for broad *lacZ* expression

A total of four Bcd-binding sites were found in the full-length *sog* shadow enhancer, three of which are shared with the 0.582 kb minimal enhancer (Figs. 2C, D and Table S5). However, the 0.686- and 0.582 kb constructs generated similar patterns of *lacZ* expression (Figs. 4A and B), suggesting that the three Bcd-binding sites in the minimal enhancer are sufficient to mediate the typical pattern of *sog* expression, even if Bcd is involved in enhancer activity. Two (B2 and B3) of the three Bcd-binding sites in the minimal enhancer are located adjacent to two DI-binding sites (D1 and D2), while the remaining site (B4) is located beside a nearby Zld-binding site (Z1). We first tested if Bcd is necessary for *sog* expression in the neurogenic ectoderm. To do this, we altered the core consensus sequences of the three Bcd-binding sites by site-directed mutagenesis. Mutations in the three Bcd sites resulted in catastrophic reduction in *lacZ* expression (Fig. 4C). Lack of either B2 (Fig. 4D) or B3 (Fig. 4E) gradually diminished *lacZ* expression in the anterior half of the embryos, whereas mutation of the fourth Bcd site (B4) did not appear to be deleterious to *lacZ* expression (Fig. 4F). These results suggest that the two Bcd-binding sites (B2 and B3) linked to the two DI-binding sites (D1 and D2) are required for normal *lacZ* expression.

To test for a possible synergistic interaction between the DI- and linked Bcd-binding sites, the helical phasing between the DI- and Bcd-binding sites was changed by inserting spacer sequences between them. A 7-bp spacer sequence was inserted between D1 and B2 sites (Table S1) to interfere with helical phasing between DI and Bcd proteins on the DNA. The elongated distance between the sites resulted in a significant and gradual decrease

in *lacZ* expression in the anterior half of the embryo (Fig. 4G). The 0.582 DI2-7bp-Bcd3 construct produced a similar *lacZ* expression pattern to that directed by the 0.582 DI1-7bp-Bcd2 construct (Fig. 4H). These results suggest that like the transcriptional synergy between DI and Zld, there is also synergistic interaction between DI- and Bcd-binding sites, and that this is required for normal *sog* expression in the neurogenic ectoderm.

Bcd-binding sites are also required for *lacZ* expression directed by the *sog* primary enhancer

It is believed that *sog* expression in the neurogenic ectoderm of the cellularizing blastoderm is directed by the collaboration

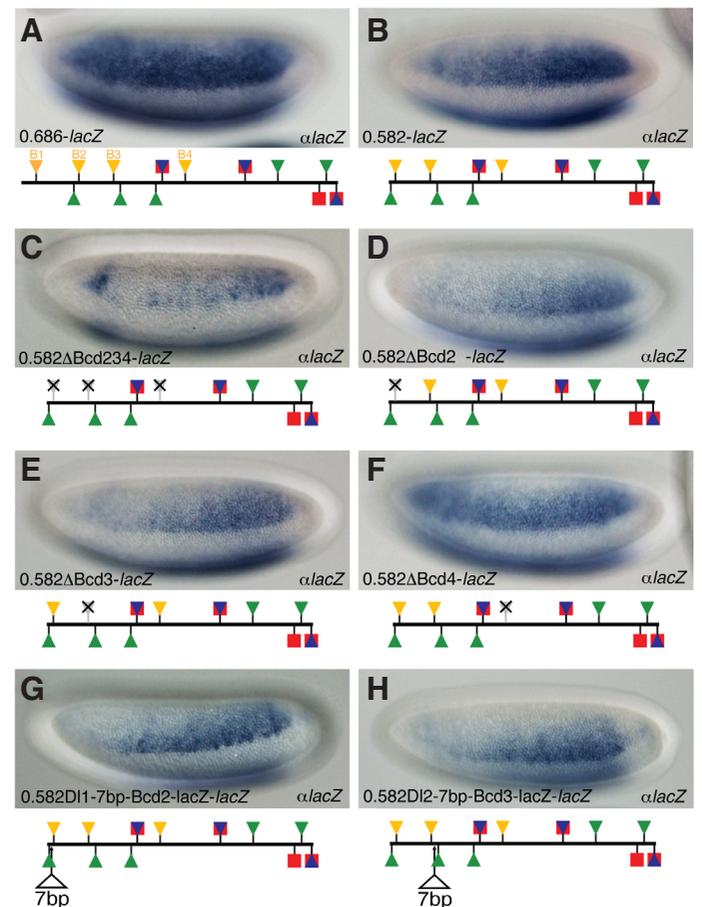


Fig. 4. Close proximity between DI- and Bcd-binding sites in the minimal enhancer is required for *lacZ* expression in the neurogenic ectoderm. (A,B) Consistent with the previous mapping analysis (Fig. 1), the 0.686- and 0.582 kb constructs directed *lacZ* expression similarly, faithfully recapitulating endogenous *sog* expression. (C) To examine if Bcd is required for *sog* expression, the three Bcd-binding sites were removed simultaneously by site-directed mutagenesis. Loss of the three Bcd sites resulted in a catastrophic reduction in *lacZ* expression along the AP and DV axes. Removal of either B2 (D) or B3 (E) also resulted in a gradual decrease in *lacZ* expression in the anterior half of the embryo in contrast to loss of the B4 site (F). To test for synergistic interactions between DI- and Bcd-binding sites, 7 nucleotides were inserted between D1 and B2 (G) and between D2 and B3 (H) to interfere with possible helical phasing between DI and Bcd proteins. Extension of the distance between DI and Bcd sites caused a gradual reduction in *lacZ* expression in both cases (G,H).

between primary and shadow enhancers (Table S8) (Markstein *et al.*, 2002; Hong *et al.*, 2008a). Critical involvement of Bcd in *sog* shadow enhancer activity (Fig. 4) raised the question of whether the primary enhancer also requires Bcd for its transcriptional activity. To evaluate this, core sequences of the Bcd binding consensus sequence in the primary enhancer were changed by site-directed mutagenesis (Fig. 5 and Table S1). ClusterDraw analyses also identified a cluster of Dl-, Zld-, Sna-, and Bcd-binding sites that precisely corresponded to the previously found primary enhancer in the first intron within the *sog* transcription unit (Figs. 2A, B, Tables S6 and S7). The primary enhancer contains three Bcd-binding sites, at least two of which are closely linked to Dl-binding sites (Fig. 5A). Initially, the primary enhancer was found by a computational algorithm designed to search the *Drosophila* genome for clusters of three or four optimal Dl-binding sites (Markstein *et al.*, 2002). The ~400-bp primary enhancer directed lateral broad stripes of *lacZ* expression (Fig. 5B), recapitulating the pattern of expression of *sog* (Fig. 1, top panel). Consistent with the previous mutagenesis study of Bcd-binding sites in the shadow enhancer (Fig. 4), loss of all Bcd-binding sites resulted in a narrow *lacZ* expression

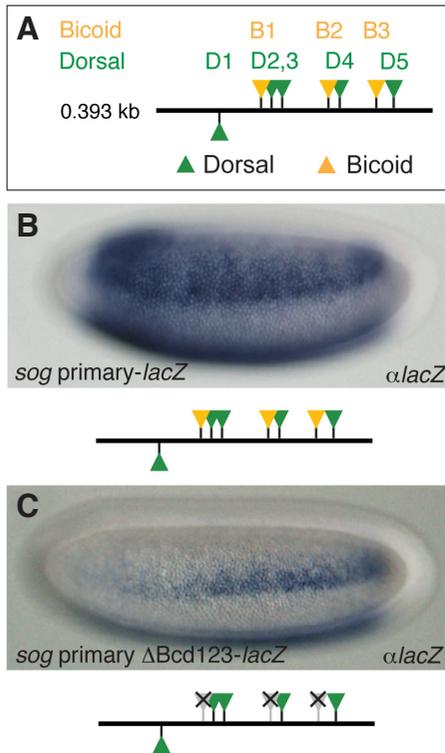


Fig. 5. Bcd-binding sites are also required for *lacZ* expression directed by the *sog* primary enhancer. Earlier ClusterDraw analyses also identified Dl- and Bcd-binding sites in the intronic primary enhancer (Fig. 2A, B; Tables S6, S7). (A) The primary enhancer was determined to contain five Dl- and three Bcd-binding sites, at least two pairs of which are closely linked to each other. The ~400-bp primary enhancer (Markstein *et al.*, 2002) directs *lacZ* expression (B) that recapitulate endogenous *sog* expression (Fig. 1, top panel), although the posterior half of the *lacZ* expression is a little narrower than that of endogenous *sog* expression. (C) Mutations of the three Bcd-binding sites in the primary enhancer resulted in a significant reduction in *lacZ* expression width and intensity. In particular, the mutant version of the primary enhancer almost failed to direct *lacZ* expression in the anterior one third of embryo.

pattern (Fig. 5C). In particular, the mutant version of the primary enhancer almost failed to activate *lacZ* in the anterior half of the transgenic embryo. The dependence of both enhancers on Bcd activity strongly implies that the broad *sog* lateral stripes in the neurogenic ectoderm require transcriptional synergy between Dl and Bcd in addition to transcriptional synergy between Dl and Zld.

Discussion

The current study presents evidence that transcriptional synergy between Dl and Zld and between Dl and Bcd is required for the broad pattern of *sog* expression observed in the neurogenic ectoderm of cellularizing blastoderm. It has been proposed that linked heterotypic activator sites can amplify transcriptional signals in an analogous manner to bipolar transistors in electronic circuits (Papatsenko and Levine, 2007). One activator in the pair functions as the signal carrier to produce a very specific expression profile. The other activator, expressed at higher levels, exhibits a broad or ubiquitous distribution. Synergistic interaction between the signal carrier and the broad or ubiquitous activator converts the weak specific signal to a functional transcription outcome via signal amplification. Dl and Zld very likely function as the signal carrier and broad amplifier, respectively. The ubiquitous maternal Zld amplifies diminishing concentrations of nuclear Dl in lateral and dorsal parts of the neurogenic ectoderm to delineate the broad profile of *sog* expression. The experimentally demonstrated transcriptional synergy between Dl and Zld in the *sog* shadow minimal enhancer (Fig. 3) indicates that signal amplification mediated by the heterotypic site combination is one of the mechanisms that delimits broad *sog* expression in the neurogenic ectoderm.

The transcriptional synergy between Dl- and Bcd-binding sites, another example of signal amplification mediated by the linked heterotypic site combination (Papatsenko *et al.*, 2009), appears to play a critical role in typical *sog* expression. *dl* and *bcd* are representative maternal effect genes that initiate AP and DV patterning of the early embryo, respectively (Berleth *et al.*, 1988; Reeves and Stathopoulos, 2009). Although classic genetic studies suggested that pattern formation is independently regulated by two distinct sets of maternal effect genes expressed along AP and DV axes (St Johnston and Nüsslein-Volhard, 1992), it is believed that AP gene expression can be modulated by DV genes and vice versa during early embryogenesis (Li *et al.*, 2008). However, the mechanism by which AP and DV genes cooperate to direct gene expression has remained obscure. Here, we present the first evidence that synergistic interactions between AP and DV genes delimit the expression profile of a patterning gene in the cellularizing blastoderm. Loss of Bcd-binding sites in the *sog* shadow minimal enhancer resulted in two characteristic *lacZ* expression patterns. First, there was a gradual loss of *lacZ* expression, at least in the anterior half of the early embryo (for instance Fig. 4D and E), when enhancers were defective in a Bcd-binding site. The aberrant *lacZ* expression pattern is likely spatially complementary to the Bcd nuclear gradient with peak levels at the anterior pole and progressively diminishing levels in posterior regions (Porcher and Dostatni, 2010), implying that the characteristic *lacZ* pattern is due to absence of binding of Bcd to its dedicated binding sites. The finding that a mutation in either the D1 Dl- or B2 Bcd-binding site in the minimal enhancer resulted in similar patterns of *lacZ* expression (Fig. S3 and 4D) also indicates that Dl and Bcd cooperatively occupy their linked

binding sites for transcriptional synergy. The second feature of Bcd-binding site mutations was that lacZ expression was much narrower than wild-type lacZ expression (compare Fig. 4C with B) when driven by enhancers defective in Bcd-binding sites. The mutant version of the primary enhancer containing no Bcd-binding site showed the most dramatic reduction in lacZ expression width (Fig. 5C). Why did the lack of Bcd-binding sites in both *sog* enhancers cause a catastrophic reduction in lacZ expression width despite intact linkages between the DI- and Zld-binding sites? It is plausible that the ability of the enhancer to interpret positional information for establishing DV pattern formation was hampered by removal of Bcd-binding sites. Perhaps cells also respond to AP positional information when initiating early patterning gene expression along the DV axis under the control of an unknown principle. In fact, dorsal-type spacing of a segmentation gene *fushi tarazu* (*ftz*) occurs all sides of the embryos lacking any of 10 maternal-effect DV polarity genes (Carroll *et al.*, 1987). These results suggest the existence of more sophisticated contacts between the AP and DV positional information.

The minimal enhancer identified here contains two pairs of DI- and linked Zld-binding sites; DI3-Zld1 and DI5-Zld3 (Fig. 2C). Centers of the DI3 and DI5 sites are located 12 and 21 bp away from those of the Zld1 and Zld3 sites, respectively (Tables S2, S3, S5, and S8). Inserting an 8-bp “spacer” sequence between DI3 and Zld1 sites separated them by 20 bp, which is almost identical to the distance between DI5 and Zld3 sites (Fig. 3G). Alteration in spacing of DI3 and Zld1 sites caused narrow lacZ expression (Fig. 3G) similar to that directed by enhancers containing a mutant Zld-binding site (Figs. 3D and F), implying that the phasing mutant pair of DI3-Zld1 sites separated by the 20-bp spacer did not function normally, whereas the DI5 and Zld3 sites separated by 21-bp spacer worked properly. The same was true for the DI1 and Bcd2 site pair interrupted by a 7-bp extra spacer (Fig. 4G). How can two pairs of binding sites separated by comparable spacers show differential transcriptional competency? The simplest interpretation is that the overall structure of a protein complex built on the minimal enhancer is more important than the strict spacing requirement among the transcription factors for synergistic transcriptional activation. Proper binding of DI and Zld (or Bcd) to their linked sites constitutes a critical platform to build a higher order cofactor complex for communicating with the RNA polymerase complex. Perhaps lack of a factor in the complex disrupts the structural topology of the platform, thereby leading to failure to assemble a complete cofactor complex. Functional impairment mediated by an alteration in phasing is evocative of the mammalian enhanceosome, which controls the expression of the β -interferon gene (β -IFN) (Thanos and Maniatis, 1995). The enhanceosome includes a series of closely linked binding sites for several transcriptional activators including NF κ B. Changes in the spacing of these binding sites lead to a severe disruption in structure and function (Panne *et al.*, 2007). However, it is hard to believe that every single nucleotide in the enhancer is sensitive to subtle changes. Rather, stabilizing selection pressure likely continually optimizes the overall arrangement of the *cis*-regulatory elements to discriminate unchangeable sequences from the flexible over evolutionary time (Ludwig *et al.*, 2000). It is conceivable that the two pairs of DI- and Bcd-binding sites and the two sets of composite elements comprising DI-, Zld-, and Sna-binding sites might be examples of invariant elements of the *sog* shadow minimal enhancer.

Materials and Methods

Plasmid construction, mutagenesis, and P element-mediated germline transformation

Genomic DNA was isolated from *yw*^{67c23} embryos aged 2-4 hours after egg deposition (AED) using previously described methods. All genomic regions used for P element-mediated germline transformation were prepared by polymerase chain reaction (PCR) amplification (Table S1 in Supplementary Material) of genomic DNA. PCR-amplified genomic fragments were cloned into the Promega™ pGEM™-T Easy vector, and sequences of the cloned fragments were verified by DNA sequencing. Cloned fragments were digested with NotI and inserted into a modified version of the [(-42)-*eve*P-lacZ]-pCaSpeR vector (Small *et al.*, 1992), in which the EcoRI site upstream of the *even-skipped* (*eve*) promoter (*eve*P) is replaced with a NotI site. All enhancer sequences were oriented in a 5' to 3' direction relative to the *sog* transcription start site. Site-directed mutagenesis was performed with a Stratagene™ QuikChange® Multi Site-Directed Mutagenesis Kit and oligonucleotide primers for introducing site-specific mutations (Table S1). Transformation constructs were introduced into the germline of *Drosophila melanogaster* as described previously (Rubin and Spradling, 1982). At least five independent lines were generated and tested for each construct.

Bioinformatics

PFMs of DI-, Zld-, Sna-, and Bcd-binding sites were obtained from *in vitro* (Fig. S1 in Supplementary Material) (Papatsenko, 2007) and *in vivo* (Fig. S2) (Zhu *et al.*, 2011) DNA binding assays. The *in vitro* data were generated by DNase footprinting analyses and systematic evolution of ligands by exponential enrichment (SELEX) experiments performed with recombinant DI, Sna, and Bcd proteins (<http://line.bioinfolab.net/webgate/help.htm#mtfform>). The *in vitro* data for Zld-binding sequences were produced by chromatin-immunoprecipitation followed by genomic DNA tiling array (ChIP-chip) analysis (Nien *et al.*, 2011), because DNase footprinting and SELEX data for Zld-binding sites were not available at the time of this study. The *in vivo* data were derived from the FlyFactorSurvey database, which is a library of the binding site preferences of transcription factors in *D. melanogaster* generated by the high-throughput bacterial one-hybrid (B1H) system (<http://pgfe.umassmed.edu/ffs/>). The ClusterDraw algorithm was given ~62 kb of the genomic sequence of the *sog* locus (X:15,588,413-15,650,156) and either *in vitro* or *in vivo* PFMs of DI, Zld, Sna, and Bcd. The *D. melanogaster* genomic sequence was obtained from the Flybase GBrowse database (<http://flybase.org/cgi-bin/gbrowse2/dmel/>, BDGP genome assembly 5 and *D. melanogaster* annotation 5.56). ClusterDraw analysis with either *in vitro* or *in vivo* PFMs indicated the presence of DI-, Zld-, Sna-, and Bcd-binding sites (Table S2-S7) in *sog* primary and shadow enhancers (Table S8) in the ~62 kb genomic region. See Supplementary Material for more details regarding determination of putative functional DI-, Zld-, Sna-, and Bcd-binding motifs in *sog* primary and shadow enhancers.

Whole-mount *in situ* hybridization

Whole-mount *in situ* hybridization was performed as described in a previous study (Hong *et al.*, 2013). Briefly, embryos were collected 2-4 hours AED, dechorinated, fixed, and then hybridized with digoxigenin (DIG) UTP-labeled antisense lacZ RNA probes. The antisense lacZ RNA probe was produced by *in vitro* transcription with PCR-generated DNA templates.

Acknowledgements

This work was supported by grants from the National Research Foundation of Korea (NRF) funded by the Korean Government (MSIP and MOE) (NRF-2010-0002792 and NRF-2012R1A1 A2038502).

Author Contributions

D.H.S. and J.W.H. conceived and supervised this study. D.H.S. and J.W.H. performed all experiments and bioinformatics analyses. D.H.S. and J.W.H. contributed to manuscript preparation.

References

- ANDERSON KV, JURGENS G, NÜSSLEIN-VOLHARD C (1985). Establishment of dorsal-ventral polarity in the *Drosophila* embryo: Genetic studies on the role of the Toll gene product. *Cell* 42: 779-789.
- BERLETH T, BURRI M, THOMA G, BOPP D, RICHSTEIN S, FRIGERIO G, NOLL M, NÜSSLEIN-VOLHARD C (1988). The role of localization of *bicoid* RNA in organizing the anterior pattern of the *Drosophila* embryo. *EMBO J* 7: 1749-1756.
- CARROLL SB, WINSLOW GM, TWOMBLY VJ, SCOTT MP (1987). Genes that control dorsoventral polarity affect gene expression along the anteroposterior axis of the *Drosophila* embryo. *Development* 99: 327-332.
- COWDEN J, LEVINE M (2002). The snail repressor positions Notch signaling in the *Drosophila* embryo. *Development* 129: 1785-1793.
- HONG JW, PARK KW, LEVINE MS (2013). Temporal regulation of *single-minded* target genes in the ventral midline of the *Drosophila* central nervous system. *Dev Biol* 380: 335-343.
- HONG JW, HENDRIX DA, LEVINE MS (2008a). Shadow enhancers as a source of evolutionary novelty. *Science* 321: 1314.
- HONG JW, HENDRIX DA, PAPATSENKO D, LEVINE MS (2008b). How the Dorsal gradient works: Insights from postgenome technologies. *Proc Natl Acad Sci USA* 105: 20072-20076.
- IP YT, PARK RE, KOSMAN D, YAZDANBAKHSH K, LEVINE M (1992). Dorsal-Twist interactions establish *snail* expression in the presumptive mesoderm of the *Drosophila* embryo. *Genes Dev* 6: 1518-1530.
- JIANG J, RUSHLOW CA, ZHOU Q, SMALL S, LEVINE M (1992). Individual Dorsal morphogen binding sites mediate activation and repression in the *Drosophila* embryo. *EMBO J* 11: 3147-3154.
- JIANG J, KOSMAN D, IP YT, LEVINE M (1991). The Dorsal morphogen gradient regulates the mesoderm determinant *twist* in early *Drosophila* embryos. *Genes Dev* 5: 1881-1891.
- LI XY, MACARTHUR S, BOURGON R, NIX D, POLLARD DA, IYER VN, HECHMER A, SIMIRENKO L, STAPLETON M, LUENGO HENDRIKS CL, CHU HC, OGAWA N, INWOOD W, SEMENTCHENKO V, BEATON A, WEISZMANN R, CELNIKER SE, KNOWLES DW, GINGERAS T, SPEED TP, EISEN MB, BIGGIN MD (2008). Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol* 6: e27.
- LIANG HL, NIEN CY, LIU HY, METZSTEIN MM, KIROV N, RUSHLOW C (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. *Nature* 456: 400-403.
- LIBERMAN LM, STATHOPOULOSA (2009). Design flexibility in *cis*-regulatory control of gene expression: Synthetic and comparative evidence. *Dev Biol* 327: 578-589.
- LUDWIG MZ, BERGMAN C, PATELNH, KREITMAN M (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403: 564-567.
- MARKSTEIN M, MARKSTEIN P, MARKSTEIN V, LEVINE MS (2002). Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the *Drosophila* embryo. *Proc Natl Acad Sci USA* 99: 763-768.
- NIEN CY, LIANG HL, BUTCHER S, SUN Y, FU S, GOCHA T, KIROV N, MANAK JR, RUSHLOW C (2011). Temporal coordination of gene networks by Zelda in the early *Drosophila* embryo. *PLoS Genet* 7: e1002339.
- PANNE D, MANIATIS T, HARRISON SC (2007). An atomic model of the interferon- β enhanceosome. *Cell* 129: 1111-1123.
- PAPATSENKO D, GOLTSEV Y, LEVINE M (2009). Organization of developmental enhancers in the *Drosophila* embryo. *Nucleic Acids Res* 37: 5665-5677.
- PAPATSENKO D (2007). ClusterDraw web server: A tool to identify and visualize clusters of binding motifs for transcription factors. *Bioinformatics* 23: 1032-1034.
- PAPATSENKO D, LEVINE M (2007). A rationale for the enhanceosome and other evolutionarily constrained enhancers. *Curr Biol* 17: R955-R957.
- PORCHER A, DOSTATNI N (2010). The Bicoid morphogen system. *Curr Biol* 20: R249-R254.
- REEVES GT, STATHOPOULOS A (2009). Graded Dorsal and differential gene regulation in the *Drosophila* embryo. *Cold Spring Harb Perspect Biol* 1: a000836.
- RUBIN GM, SPRADLING AC (1982). Genetic transformation of *Drosophila* with transposable element vectors. *Science* 218: 348-353.
- RUSHLOW CA, HAN K, MANLEY JL, LEVINE M (1989). The graded distribution of the Dorsal morphogen is initiated by selective nuclear transport in *Drosophila*. *Cell* 59: 1165-1177.
- SMALL S, BLAIR A, LEVINE M (1992). Regulation of *even-skipped* stripe 2 in the *Drosophila* embryo. *EMBO J* 11: 4047-4057.
- ST JOHNSTON D, NÜSSLEIN-VOLHARD C (1992). The origin of pattern and polarity in the *Drosophila* embryo. *Cell* 68: 201-219.
- SZYMANSKI P, LEVINE M (1995). Multiple modes of Dorsal-bHLH transcriptional synergy in the *Drosophila* embryo. *EMBO J* 14: 2229-2238.
- THANOS D, MANIATIS T (1995). Virus induction of human IFN β gene expression requires the assembly of an enhanceosome. *Cell* 83: 1091-1100.
- ZHU LJ, CHRISTENSEN RG, KAZEMIAN M, HULL CJ, ENUAMEH MS, BASCIOTTA MD, BRASEFIELD JA, ZHU C, ASRIYAN Y, LAPOINTE DS, SINHA S, WOLFE SA, BRODSKY MH (2011). FlyFactorSurvey: A database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res* 39: D111-D117.
- ZINZEN RP, SENGER K, LEVINE M, PAPATSENKO D (2006). Computational models for neurogenic gene expression in the *Drosophila* embryo. *Curr Biol* 16: 1358-1365.

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

Regulation of Merkel cell development by Pax6

Ida Parisi and J. Martin Collinson
Int. J. Dev. Biol. (2012) 56: 341-350
<http://www.intjdevbiol.com/web/paper/113406ip>

Evolution of cis-regulatory regions versus coding regions

Francisco Rodríguez-Trelles, Rosa Tarrío and Francisco J Ayala
Int. J. Dev. Biol. (2003) 47: 665-673
<http://www.intjdevbiol.com/web/paper/14756342>

The circadian gene *Clock* is required for the correct early expression of the head specific gene *Otx2*

Richard Morgan
Int. J. Dev. Biol. (2002) 46: 999-1004
<http://www.intjdevbiol.com/web/paper/12533023>

Twist functions in mouse development

Meredith P O'Rourke and Patrick P L Tam
Int. J. Dev. Biol. (2002) 46: 401-413
<http://www.intjdevbiol.com/web/paper/12141426>

The role of the homeodomain protein *Bozozok* in zebrafish axis formation

L Solnica-Krezel and W Driever
Int. J. Dev. Biol. (2001) 45: 299-310
<http://www.intjdevbiol.com/web/paper/11291860>

5 yr ISI Impact Factor (2013) = 2.879

