BMB Reports - Manuscript Submission

Manuscript Draft

Manuscript Number: BMB-16-119

Title: Transcriptional activity of the short gastrulation primary enhancer in the ventral midline requires its early activity in the presumptive neurogenic ectoderm

Article Type: Article

Keywords: Drosophila; short gastrulation; primary enhancer; shadow enhancer; embryo

Corresponding Author: Joung-Woo Hong

Authors: Dong-Hyeon Shin[1], Joung-Woo Hong[1,*]

Institution: [1]Graduate School of East-West Medical Science, Kyung Hee University,

Transcriptional activity of the *short gastrulation* primary enhancer in the ventral midline requires its early activity in the presumptive neurogenic ectoderm

Dong-Hyeon Shin and Joung-Woo Hong*

Graduate School of East-West Medical Science, Kyung Hee University, Yongin, 17104, Korea

*Corresponding Author:

Joung-Woo Hong, Ph.D.
Graduate School of East-West Medical Science,
Kyung Hee University
Yongin, 17104, Korea
E-mail: jwhong46@khu.ac.kr
Office: +82-31-201-3853
Fax: +82-31-204-8119

Running Title: Dual transcriptional activities of the *sog* primary enhancer

ABSTRACT

The *short gastrulation* (*sog*) shadow enhancer directs early and late *sog* expression in the neurogenic ectoderm and the ventral midline of the developing *Drosophila* embryo, respectively. Here, evidence is presented that the *sog* primary enhancer also has both activities, with the late enhancer activity dependent on the early activity. Computational analyses showed that the *sog* primary enhancer contains five Dorsal (Dl)-, four Zelda (Zld)-, three Bicoid (Bcd)-, and no Single-minded (Sim)-binding sites. In contrast to many ventral midline enhancers, the primary enhancer can direct lacZ expression in the ventral midline as well as in the neurogenic ectoderm without a canonical Sim-binding site. Intriguingly, the impaired transcriptional synergy between Dl and either Zld or Bcd led to aberrant and abolished lacZ expression in the neurogenic ectoderm and in the ventral midline, respectively. These findings suggest that the two enhancer activities of the *sog* primary enhancer are functionally consolidated and geographically inseparable.

Keywords: *Drosophila*, *short gastrulation*, primary enhancer, shadow enhancer, embryo

INTRODUCTION

The *sog* gene is one of seven zygotically active genes [*decapentaplegic* (*dpp*), *zerknüllt* (*zen*), *sog*, *tolloid* (*tld*), *twisted gastrulation* (*tsg*), *screw* (*scw*) and *shrew* (*srw*)] required for dorsal-ventral (DV) patterning in the ectoderm of the *Drosophila* early embryo (1). Originally, *sog* was found as one of three X-lined zygotic genes required for specific morphogenetic events of gastrulation (2). The *sog* gene exhibits dynamic expression during embryonic development (1). *sog* transcripts are first observed in broad lateral stripes of the neurogenic ectoderm as early as nuclear cleavage cycle 13, the dorsal borders of which abut the ventral limits of the *dpp* expression domain in the dorsal ectoderm. At least by germ band extension, *sog* expression is restricted to the ventral midline, which comprises specialized glial cells that secrete signals essential for nerve cord patterning (3).

Early *sog* expression is dependent on the DV determinants, Dl and Snail (Sna) (4). Bioinformatics studies of the genome-wide distribution of the Dl recognition sequence have identified a ~400-bp genomic region that acts as an enhancer to direct the early broad pattern of *sog* expression in the neurogenic ectoderm (5). Subsequent chromatin immunoprecipitation followed by genomic tiling array (ChIP-chip) analyses indicated that many of the Dl target genes contain two independent and separate enhancers that control the same or similar expression patterns (6). Recently, it was shown that the *sog* locus has a secondary remote enhancer that directs its expression in the neurogenic ectoderm (7). The two enhancers are referred to as the "primary" and "shadow" enhancers, respectively, according to the chronological order of their identification, rather than functional differences. More recently, the shadow enhancer has been shown to direct late *sog* expression in the ventral midline after gastrulation (8). The finding that the shadow enhancer has dual enhancer activities in the early neurogenic ectoderm and the late ventral midline raised the possibility

that the primary enhancer also has the ability to control late *sog* expression in the ventral midline.

Here, we demonstrate that the *sog* primary enhancer also has enhancer activities in the neurogenic ectoderm and the ventral midline, and that the late enhancer activity requires the early enhancer activity. Computational analyses indicated that the primary enhancer contains five Dl-, four Zld-, and three Bcd-binding sites. Despite its midline enhancer activity, no Sim-binding site was identified within the primary enhancer. Intriguingly, removal of Dl-, Zld-, or Bcd-binding sites abolished lacZ expression in the ventral midline and led to its aberrant expression in the neurogenic ectoderm. These results suggest that late enhancer activity is potentiated by its early transcriptional activation.

RESULTS

The *sog* primary enhancer directs lacZ expression in the ventral midline of the late embryo

single-minded (*sim*) is a master regulatory gene that directly regulates the expression of many ventral midline genes (9). *sim* transcripts are first observed in a single row of cells, called the mesectoderm, found on along either side of the presumptive mesoderm (Fig. 1A). The symmetric lines of mesectodermal cells converge at the ventral midline during gastrulation (Fig. 1B). Once induced, *sim* expression is maintained via autoregulation during germ band elongation (Fig. 1C and D) and later stages of embryogenesis. The *sog* gene has a broad expression pattern in the neurogenic ectoderm as early as nuclear cleavage cycle 13 (Fig. 1E). Through gastrulation (Fig. 1F), *sog* expression is restricted to the ventral midline (Fig. 1G and H), which is comparable with the *sim* expression pattern in the corresponding developmental stages. Endogenous *sog* midline expression is thought to be controlled directly by the *sim* gene product. To test if *sog* is a target of the Sim protein, *sog* expression was examined in mutant embryos homozygous for a *sim* null allele (sim^{H9}) (Fig. 1I-L). Early *sog* expression was not affected by lack of Sim (Fig. 1I and J), while *sog* transcripts were not detected in the ventral midline, at least after completion of gastrulation (Fig. 1K and L). These results suggest that *sog* expression is directly controlled by Sim in the ventral midline of the developing embryo.

The previous finding that the *sog* shadow enhancer can direct its midline expression (8) prompted us to test if the *sog* primary enhancer has similar activity. An approximately 400-bp primary enhancer directed broad stripes of lacZ expression on either side of the neurogenic ectoderm in transgenic embryos (Fig. 1M) (5). The early broad domain of lacZ expression narrowed to the ventral regions of the neurogenic ectoderm during gastrulation (Fig. 1N).

From the onset of germ band elongation, lacZ transcripts were observed in the ventral midline (Fig. 1O and P), recapitulating endogenous *sim* and *sog* patterns in the ventral midline (compare with Fig. 1C, D, G, and H, respectively). These results suggested that, in addition to the *sog* shadow enhancer, the primary enhancer also contains ventral midline enhancer activity, and thus that the *sog* midline expression pattern is determined by both enhancers. The finding that both enhancers can direct both early and late *sog* expression explains why efforts to identify a *sog* ventral midline enhancer have long been unsuccessful.

The *sog* primary enhancer does not contain Sim-binding sites

The loss of *sog* expression in the ventral midline of a *sim* mutant embryo together with the ability of the *sog* primary enhancer to direct lacZ expression in the ventral midline raised the possibility that the primary enhancer contains Sim-binding sites. To test this possibility, ClusterDraw analyses were performed with position frequency matrices (PFMs) (Fig. S1 and S2 in Supplementary Material) for Dl-, Zld-, Sna-, Bcd-, and Sim-binding sites (Fig. 2). ClusterDraw is an *r*-scan-based program that has been used to identify binding motifs and binding clusters of specific combinations of transcription factors (10). To increase the statistical power of the computational analyses, an identical ClusterDraw analysis was performed twice with two different sets of PFMs for the five transcription factors. One set of PFMs was generated by motif alignments obtained from the *in vitro* binding data (10) (Fig. 2A) and the other from the *in vivo* binding data (11) (Fig. 2B). Zld- and Bcd-binding sites were included in those searches because lack of Zld and Bcd proteins led to impaired endogenous *sog* expression in the neurogenic ectoderm, suggesting that the primary enhancer contains at least one binding site for each transcription factor.

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 S6). These results suggest that the primary and shadow enhancers of *sog* contain the most significant clusters of Dl-, Zld-, Sna-, and Bcd-binding motifs across the ~62-kb genomic region. The ClusterDraw algorithm also displays the location and quality [cumulative match probability (-log*P*)] of each binding site found in the clusters that it identifies. 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 (Tables S2-S5, see Supplementary Material for more details). The cluster corresponding to the primary enhancer contained five Dl-, four Zld-, three Sna-, and three Bcd-binding sites, whereas no Sim-binding site was identified in either of the repeated analyses (Fig. 2C). The absence of Sim-binding sites within the ~400-bp primary enhancer is consistent with the previous observation that the *sog* shadow enhancer does not include the '5-ACGTG-3' Sim-binding site (8), which has been found in all ventral midline enhancers tested to date. These results strongly suggest that, like the shadow enhancer, the primary enhancer may also function as a HOT region (12) to control *sog* ventral midline expression.

Early activity of the primary enhancer in the neurogenic ectoderm is required for its late activity in the ventral midline

Recently, it was shown that synergistic interactions between Dl and Zld and between Dl and Bcd in the shadow enhancer play a critical role in generating broad lacZ expression in the neurogenic ectoderm (13). Thus, the close proximities of Dl to both Zld and Bcd in the

primary enhancer (Fig. 2C) raised the possibility that the early broad stripes of lacZ expression directed by the primary enhancer (Fig. 3A) also require transcriptional synergy between D1 and Zld or Bcd. To examine this possibility, consensus sequences of Dl-, Zld-, and Bcd-binding sites were changed by site-directed mutagenesis (Fig. 3E, I, M and Table S1). Removal of four Dl-binding sites in the primary enhancer led to complete failure of lacZ expression in the neurogenic ectoderm (Fig. 3E). In addition, loss of either Zld- or Bcd-binding sites resulted in severe reduction in lacZ expression width (compare Fig. 3I and M with A). These lacZ expression patterns are reminiscent of those mediated by the mutant shadow enhancers where synergistic interaction between Dl and either Zld or Bcd is hampered (13). In addition, removal of only the first Zld-binding site (Z1) (Fig. 2C) created a narrow pattern of lacZ expression similar to that mediated by the mutant construct containing no Zld-binding site (Fig. S3). Furthermore, mutation of either the second (B2) or third (B3) Bcd-binding site (Fig. 2C) also resulted in a dramatic reduction in lacZ expression width (Fig. S3). Intriguingly, the remaining narrow lacZ expression patterns directed by constructs containing mutant Zld- and Bcd-binding sites (Z1, B2 and B3) were completely abolished by removal of the linked Dl-binding sites (Fig. S3). These results suggest that transcriptional synergy between Dl and either Zld or Bcd is required for creating the early broad pattern of *sog* expression throughout the neurogenic ectoderm.

One of the most intriguing features of the shadow enhancer is that linked Dl- and Zld-binding sites are required for its late ventral midline enhancer activity (8). Lack of those linked sites in the shadow enhancer results in catastrophic reduction in late enhancer activity in the ventral midline. Because *sog* expression is controlled by primary and shadow enhancers at the same time and in a identical space (7), we reasoned that the late activity of the primary enhancer may depend on early enhancer activity. To test this hypothesis, lacZ expression

patterns directed by the wild-type and mutant versions of the primary enhancers in the ventral midline were investigated. During late gastrulation, residual lacZ transcripts were still observed in wild-type and mutant embryos (Fig. 3B, F, J, and N). However, after gastrulation, none of the mutant primary enhancers activated lacZ in the ventral midline (compare Fig. 3G, H, K, L, O, and P with C and D). In contrast to the Dl mutant enhancer (Fig. 3E), the primary enhancers containing either mutant Zld or Bcd binding sites could still direct lacZ expression, even though there was severe reduction in lacZ expression width and strength. Nevertheless, the early impaired lacZ expression in the neurogenic ectoderm led to serious failure of late lacZ expression in the ventral midline.

DISCUSSION

Almost all of the developmental enhancers tested so far only control the transcription of their target genes at a particular time and in a defined space (14). However, most genes involved in development are repeatedly used at different times and in diverse spaces during an entire process of differentiation (15). Thus, each versatile gene needs various developmental enhancers that switch its transcription on and off at the correct time and in the proper location (16). Expression of *rhomboid* (*rho*) that is regulated by at least two discrete enhancers is one good example (17). *rho*, a DV patterning gene involved in the development of the central nervous system (CNS), is expressed in the neurogenic ectoderm and the ventral midline of the developing *Drosophila* embryo. The sequential expression in these two different locations is directed by two separate enhancers, the neurogenic ectoderm enhancer (NEE) and the ventral midline enhancer (VME). Although these enhancers are located close to each other, they are functionally autonomous and geographically separable. In contrast to this property of the *rho* enhancers, however, recent studies performed with transgenic embryos showed that the *sog* shadow enhancer can also direct *sog* expression in the ventral midline as well as in the presumptive neurogenic ectoderm (18). As far as we know, the *sog* shadow enhancer is the first developmental enhancer reported to direct its target gene expression at two different times and in two discrete spaces during *Drosophila* embryogenesis. The current study presents evidence that the *sog* primary enhancer shares three functional similarities with the shadow enhancer. First, the primary enhancer is also able to direct *sog* expression in the ventral midline of a developing embryo (Fig. 1). All of the ventral midline genes that have been examined so far have expression that is controlled by the ventral midline enhancers that are not able to allow expression of midline genes in locations other than the ventral midline in a developing embryo. Second, like the *sog* shadow enhancer, the primary enhancer also does not have a canonical Sim-binding site (Fig. 2 and Table S6). All of the ventral midline

enhancers have been shown to have at least one Sim-binding consensus sequence containing a '5'-ACGTG-3' core motif (19). However, both the primary and shadow *sog* enhancer do not have such a Sim-binding site (Table S6). These findings suggest that the two *sog* enhancers may direct ventral midline expression with an unknown mechanism to bypass involvement of Sim. This may be the reason why intense efforts to identify the *sog* midline enhancer with the canonical Sim-binding consensus sequences have not been successful. Third, the midline enhancer activity of the primary enhancer also requires its neurogenic ectoderm activity (Fig. 3). The early broad *sog* expression in the neurogenic ectoderm depends on transcriptional synergy between Dl, Zld and Bcd in the two *sog* enhancers (Fig. 3 and S3) (13). The impaired synergistic interaction between them led to severe reduction in the early neurogenic ectoderm enhancer activity, which in turn abolished the late midline enhancer activity in both enhancers (Fig. 3) (8). It is conceivable that in contrast to the two separate *rho* enhancers, the two enhancer activities embedded in the primary enhancer are functionally consolidated and geographically inseparable.

The early and late enhancer activities of the *sog* primary are determined by independent transcription factors in the neurogenic ectoderm and the ventral part of the developing embryo, respectively. For example, early and late *sog* expression was directed by Dl, Zld, and Sna in the neurogenic ectoderm and Sim in the ventral midline, respectively. Although their transcriptional activities are not coincidental in time or space during embryogenesis, changes in early transcriptional input of the enhancers catastrophically interfered with their late transcriptional outcome. The simplest interpretation for this paradoxical observation is that the early determinants may potentiate the enhancers for the subsequent transcriptional burst by late transcription factors. For example, the preceding interaction between the early transcription factors and enhancers may make the chromatin structure more accessible for the

late factors. Indeed, the maternal transcription factor Zld, which is involved in creating the broad *sog* expression in the neurogenic ectoderm (Fig. 3I), functions as a pioneer factor to increase chromatic accessibility (20). A pioneer factor is a transcription factor that confers transcriptional competency to inactive target enhancers by binding condensed chromatin prior to the binding of other transcription factors (21). It is plausible that early binding of Zld to the primary enhancer establishes an open chromatin environment in the neurogenic ectoderm and causes transcriptional synergy between Dl and Zld by facilitating their cooperative site occupancy. However, the functional dependency of the late enhancer activity on the early one observed in the primary enhancer does not seem to rely entirely on the pioneering activity of Zld, because Zld still binds the primary enhancer containing either mutant Dl- or Bcd-binding sites (Fig. 3). It is possible that there may exist unidentified pioneering factors that prime the developmental enhancers in addition to the zinc-finger transcription factor Zld. Pioneering activity has mostly been observed in transcription factors containing a forkhead box (FOX) or zinc-finger domain (22), which strongly supports the possibility that the fly genome encodes more than one pioneering factor involved in the regulation of early DV gene expression.

MATERIALS AND METHODS

Detailed information is provided in the online Supplementary Material.

ACKNOWLEDGEMENTS

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

REFERENCS

1. Francois V, Solloway M, O'Neill JW, Emery J and Bier E (1994) Dorsal-ventral patterning of the *Drosophila* embryo depends on a putative negative growth factor encoded by the short gastrulation gene. Genes Dev 8, 2602-2616

2. Wieschaus E, Nüsslein-Volhard C and Jurgens G (1984) Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster* III. Zygotic loci on the X-chromosome and fourth chromosome. Roux's Arch Dev Biol 193, 296-307

3. Menne TV, Luer K, Technau GM and Klambt C (1997) CNS midline cells in *Drosophila* induce the differentiation of lateral neural cells. Development 124, 4949-4958

4. Stathopoulos A and Levine M (2005) Genomic regulatory networks and animal development. Dev Cell 9, 449-462

5. Markstein M, Markstein P, Markstein V and 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

6. Zeitlinger J, Zinzen RP, Stark A, Kellis M, Zhang H, Young RA and 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

7. Hong JW, Hendrix DA and Levine, MS (2008) Shadow enhancers as a source of evolutionary novelty. Science 321, 1314

8. Shin DH and Hong JW (2015) Midline enhancer activity of the short gastrulation shadow enhancer is characterized by three unusual features for *cis*-regulatory DNA. BMB Rep 48, 589-594

9. Nambu JR, Franks RG, Hu S and Crews ST (1990) The *single-minded* gene of *Drosophila* is required for the expression of genes important for the development of CNS midline cells. Cell 63, 63-75

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. Zhu LJ, Christensen RG, Kazemian M, Hull CJ, Enuameh MS, Basciotta MD, Brasefield JA, Zhu C, Asriyan Y, Lapointe DS, Sinha S, Wolfe SA and 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

12. Kvon EZ, Stampfel G, Yanez-Cuna JO, Dickson BJ and Stark A (2012) HOT regions function as patterned developmental enhancers and have a distinct *cis*-regulatory signature. Genes Dev 26, 908-913

13. Shin DH and Hong JW (2016) The *short gastrulation* shadow enhancer employs dual modes of transcriptional synergy. Int J Dev Biol In press, http://dx.doi.org/10.1387/ijdb.160165jh.

14. Levine M (2010) Transcriptional enhancers in animal development and evolution. Curr Biol 20, R754-R763

15. Placzek M and Briscoe J (2005) The floor plate: multiple cells, multiple signals. Nat Rev Neurosci 6, 230-240

16. Jeong Y, El-Jaick K, Roessler E, Muenke M and Epstein DJ (2006) A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. Development 133, 761-772

17. Ip YT, Park RE, Kosman D, Bier E and 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

18. Shin DH and Hong JW (2015) The shadow enhancer of *short gastrulation* also directs its expression in the ventral midline of the *Drosophila* embryo. Genes Genom 37, 743-750

19. Hong JW, Park KW and 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

20. Foo SM, Sun Y, Lim B, Ziukaite R, O'Brien K, Nien CY, Kirov N, Shvartsman SY and Rushlow CA (2014) Zelda potentiates morphogen activity by increasing chromatin accessibility. Curr Biol 24, 1341-1346

21. Zaret KS and Carroll JS (2011) Pioneer transcription factors: establishing competence for gene expression. Genes Dev 25, 2227-2241

22. Magnani L, Eeckhoute J and Lupien M (2011) Pioneer factors: directing transcriptional regulators within the chromatin environment. Trends Genet 27, 465-474

23. Campos-Ortega JA and Hartenstein V (1985) *The Embryonic Development of Drosophila melanogaster*, Springer-Verlag, Berlin; Heidelberg

FIGURE LEGENDS

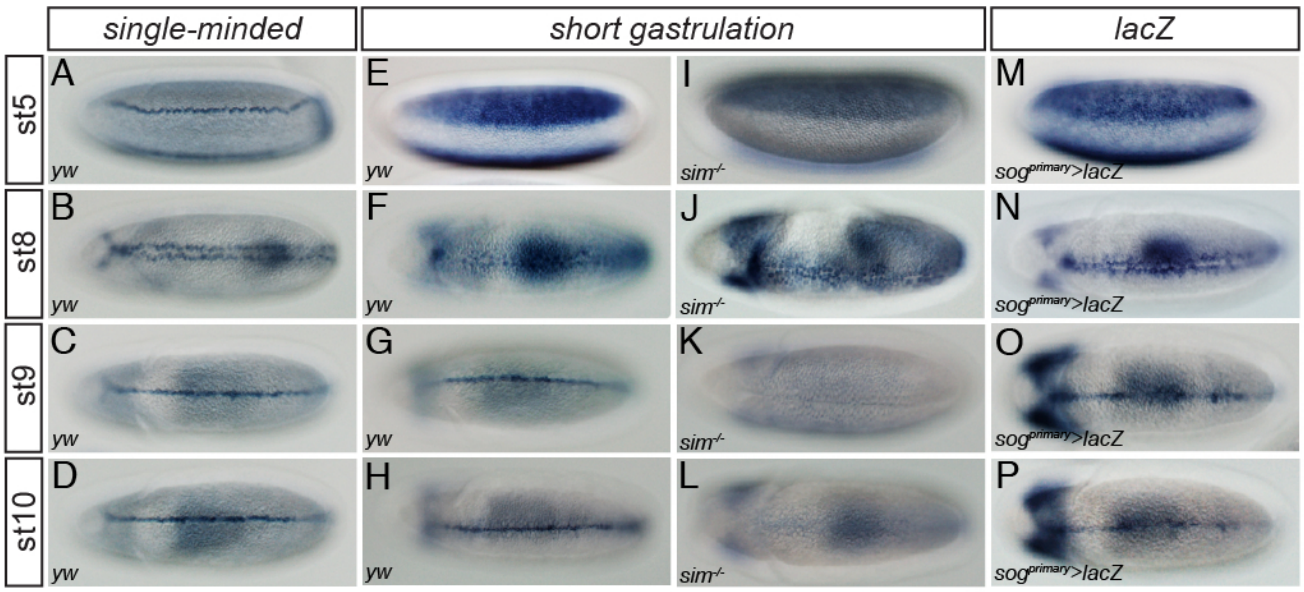
Figure 1. The *sog* primary enhancer directs expression in the ventral midline of the late embryo. Approximately 2-10 hours (h) after egg deposition (AED), embryos were collected, dechorionated, and fixed. Whole-mount *in situ* hybridization was performed with fixed embryos and digoxigenin (DIG)-UTP labeled antisense RNA probes complementary to *sim*, *sog*, and *lacZ*. Each probe used in the individual *in situ* hybridization is shown on the top of each column. Expression patterns of *sim* (A-D) were visualized in wild-type (*yw*) *Drosophila* embryos. An antisense *sog* RNA probe was used to target endogenous *sog* transcripts in both wild-type (*yw*) (E-H) and *sim* mutant ($sim^{-/-}$) (I-L) embryos. The *sim* mutant embryo was homozygous for the $sim^{2/H9}$ allele. (M-P) Expression of a *lacZ* fusion gene directed by a ~0.4-kb *sog* primary enhancer in a transgenic embryo recapitulated the endogenous pattern of *sog* expression in the neurogenic ectoderm and the ventral midline (compare with E-H). '*st*' indicates the developmental stage of *Drosophila* embryogenesis. Developmental stages were defined according to previously established criteria (23).

Figure 2. ClusterDraw analyses across a ~62-kb genomic region encompassing the *sog* locus. ClusterDraw analyses were performed with two different types of PFMs for the Dl-, Zld-, Sna-, Bcd-, and Sim-binding sequences. One was built by motif alignments obtained from *in vitro* binding data (10) (A) and the other from *in vivo* binding data (11) (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 (-log*P*) (*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 S6). No Sim-binding sites were identified in the *sog* primary enhancer. Gene models over the 62-kb genomic region are depicted below panel B. (C) ClusterDraw analyses also indicated the location and cumulative

match probability (-log*P*) value of each motif found in the primary enhancer (Tables S2-S5). The thick line denotes ~400-bp of the *sog* primary enhancer (Table S6) (5). 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.

Figure 3. Intact activity of the *sog* primary enhancer in the presumptive neurogenic ectoderm is required for its late activity in the ventral midline of the developing *Drosophila* embryo. Approximately 2-10 h AED embryos were collected, dechorionated, fixed, and hybridized with DIG-UTP labeled antisense *lacZ* RNA. Binding sites of Dl (green), Zld (blue), Sna (red), and Bcd (yellow) in the *sog* primary enhancer are depicted on top of each column. Triangles and squares represent binding sites for a transcriptional activator and repressor, respectively. Mutagenized Dl-, Zld-, Sna-, and Bcd-binding sites are marked with asterisks (*). (A-D) The wild-type (WT) *sog* primary enhancer directs l*acZ* expression in the early neurogenic ectoderm and the late ventral midline of *Drosophila* embryo. (E-H) Mutations in four Dl-binding sites abolished lacZ expression both early in the neurogenic ectoderm and late in the ventral midline. (I-L) Lack of Zld-binding sites in the primary enhancer led not only to severe reduction in lacZ expression width in the neurogenic ectoderm, but also to complete loss of lacZ expression in the ventral midline. Note that the remaining lacZ expression also gradually diminished along the anterior-posterior (AP) axis (I). (M-P) Removal of Bcd-binding sites in the primary enhancer produced similar lacZ expression patterns to those of the primary enhancer containing no Zld-binding site, except that narrow lacZ expression gradually increased from the anterior to posterior pole. This lacZ pattern appears to be a mirror image of that produced by the ΔZld1234 construct (compare M with I).

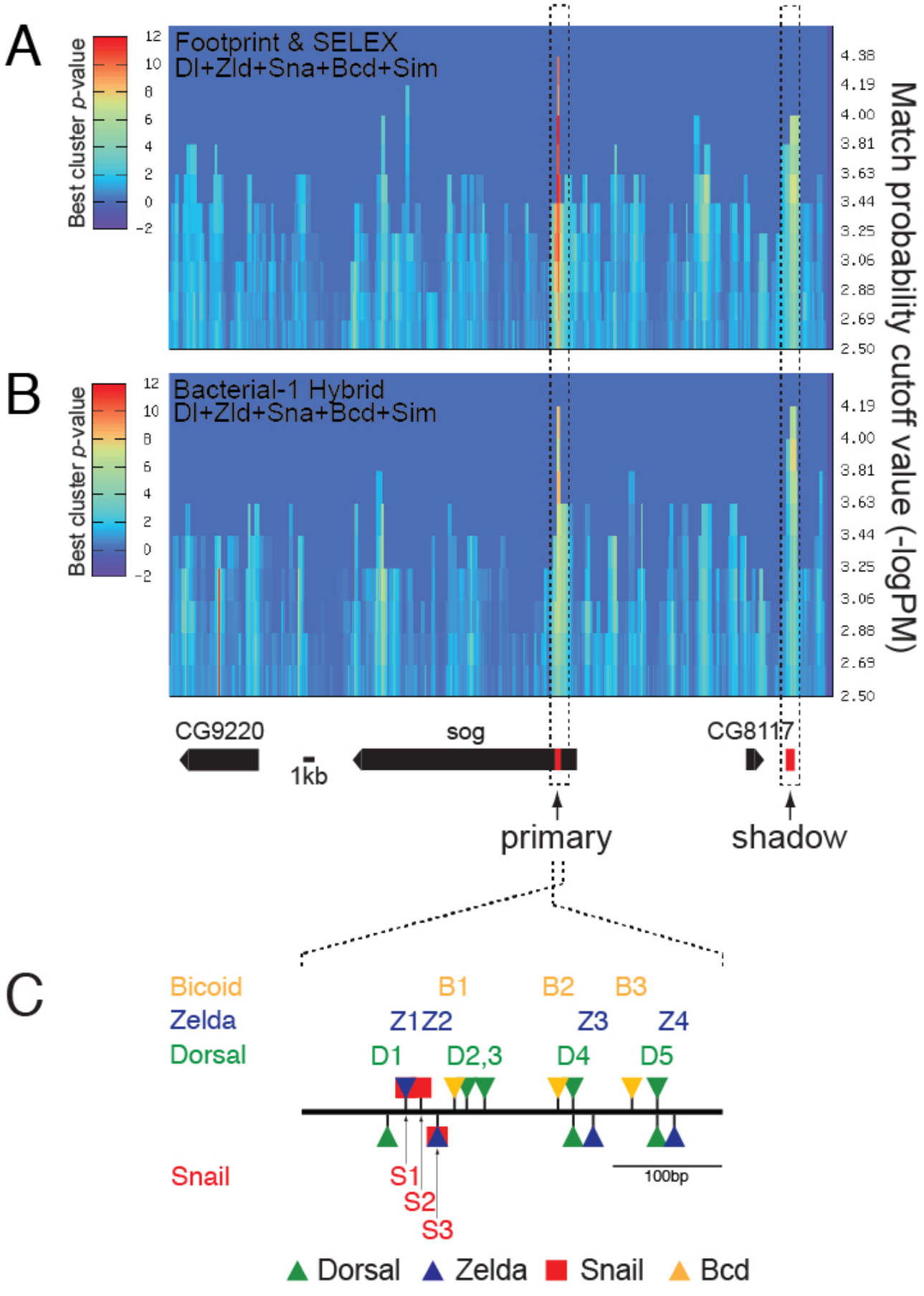
Figure 1



Shin & Hong

Fig. 1

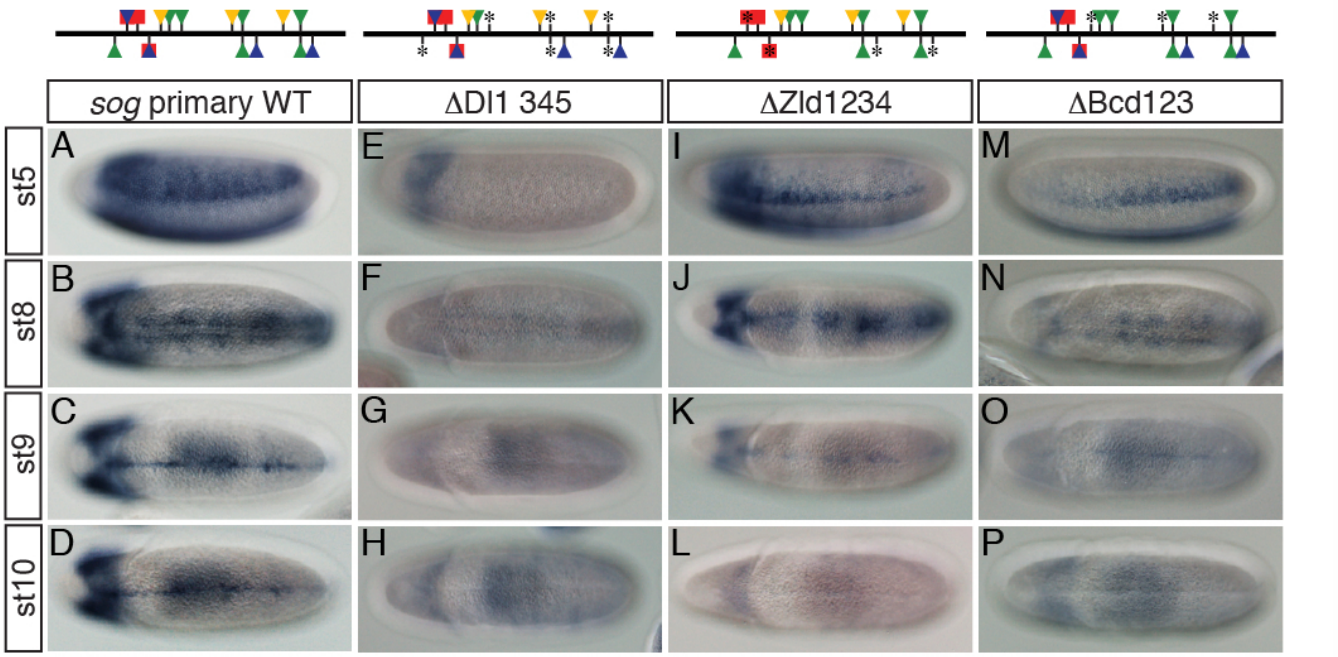
Figure 2



Shin & Hong

Fig. 2

Figure 3



Shin & Hong

Fig. 3

Supplementary Material

Transcriptional activity of the *short gastrulation* primary enhancer in the ventral midline requires its early activity in the presumptive neurogenic ectoderm

Dong-Hyeon Shin and Joung-Woo Hong*

Graduate School of East-West Medical Science, Kyung Hee University, Yongin, 17104, Korea

*Corresponding Authors:

Joung-Woo Hong, Ph.D.
Graduate School of East-West Medical Science,
Kyung Hee University
Yongin, 17104, Korea
E-mail: jwhong46@khu.ac.kr
Office: +82-31-201-3853
Fax: +82-31-204-8119

Running Title: Late *sog* expression requires its early expression

Supplementary MATERIALS AND METHODS

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

Genomic DNA was isolated from *yw*^{67c23} embryos collected 2-4 hours after egg deposition (AED) using previously described methods (Hendrix, *et al*., 2008). All genomic regions used for P-element-mediated transformation were prepared by genomic polymerase chain reaction (PCR) amplification (Table S1 in Supplementary Material). PCR-amplified genomic fragments were cloned into the Promega™ pGEM®-T Easy vector, and sequences of the cloned fragments were verified by DNA sequencing reactions. Cloned fragments were digested by NotI and inserted into a [(-42)-*eve*P-lacZ]-pCaSpeR vector (Small, *et al*., 1992) that was modified to contain a unique NotI site upstream of the *even-skipped* (*eve*) promoter (*eve*P) in place of the EcoRI site originally present in this vector (Stathopoulos and Levine, 2005). Enhancer sequences were all oriented in a 5' to 3' direction relative to the chromosomal transcription start-site. Site-directed mutagenesis was performed with the Stratagene™ QuikChange® Multi Site-Directed Mutagenesis Kit and site-specific primers (Table S1). Transformation constructs were introduced into the germline of *Drosophila melanogaster*, as described previously (Rubin and Spradling, 1982; Ip, *et al*., 1992; Jiang and Levine, 1993).

Bioinformatics

Position frequency matrices (PFMs) of Dorsal (Dl)-, Zelda (Zld)-, Snail (Sna)-, Bicoid (Bcd)-, and Single-minded (Sim)-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 Dl, Sna, Bicoid, and Sim (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 Dl, Zld, Sna, Bcd, and Sim. 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 Dl-, Zld-, Sna-, and Bcd-binding sites in the *sog* primary enhancer (Tables S2-S5). Primary and shadow enhancer sequences in a ~62-kb genomic region that direct *sog* expression in the neurogenic ectoderm of early *Drosophila* embryos are shown in Table S6.

Whole-mount *in situ* hybridization

Embryos were collected 2-10 h AED, dechorinated, fixed, and hybridized with digoxigenin (DIG)-UTP labeled antisense RNA probes, as described previously (Hong, *et al*., 2013). An antisense RNA probe was used to investigate staining patterns of transgenic embryos. The hybridized antisense RNA probe was stained with nitro-blue tetrazolium chloride (NBT) and 5'-bromo-4-chloro-3-indolyl phosphate (BCIP) for exactly 20 minutes to normalize expression levels. To examine the endogenous expression patterns of *sim*, *short gastrulation* (*sog*), and *intermediate neuroblasts defective* (*ind*), templates for probes were produced by genomic PCR, introduced into the pGEM®-T Easy vector, and used for *in vitro* transcription reactions. Between 1-2 kb of coding sequences were

used as a template for each probe. Developmental stages during embryogenesis were defined according to previously established criteria (Campos-Ortega and Hartenstein, 1985).

Table S1. DNA oligonucleotide sequences used in this study

Enhancer Name	Sequence (5' to 3' direction)[1]	Note[2]
sog primary 0.392	F:GTTGCCAATGCCATTGCGCATACGCCGTGTCG	Fig. 1M-P
	R:GCTTTATGGTCCATGGTCCATACCAC	Fig. 1M-P
0.392 ΔDl 1 345	mutDl1-F:CTATATGGCTGTATGGTGCGTTGAAATCAACGTAATCGCAGGTAGAATTC	Fig. 3E-H
	mutDl1-R:GAATTCTACCTGCGATTACGTTGATTTCAACGCACCATACAGCCATATAG	Fig. 3E-H
	mutDl3-F:TCGCACCTCTAATCCCGCCATTGTTTTCAAGACATGGGATATTCCCGACG	Fig. 3E-H
	mutDl3-R:GTCGGGAATATCCCATGTCTTGAAAACAATGGCGGGATTAGAGGTGCGA	Fig. 3E-H
	mutDl4-F:CCAGTTTTAATCCGGAAAGCTTGAATTCAATTCCGCTCGCTGCCTGCACT	Fig. 3E-H
	mutDl4-R:AGTGCAGGCAGCGAGCGGAATTGAATTCAAGCTTTCCGGATTAAAACTGG	Fig. 3E-H
	mutDl5-F:GCCGCTTACCAAAAAGATACTTGTATACAAAAATGGATGCCTGCCCATGT	Fig. 3E-H
	mutDl5-R:ACATGGGCAGGCATCCATTTTTGTATACAAGTATCTTTTTGGTAAGCGGC	Fig. 3E-H
0.392 ΔZld 1234	mutZld1-F:CCGGGGAAATCCCCGTAATCGCTCGTTGAATTCCAGCCGGTGCCGAGG	Fig. 3I-L
	mutZld1-R:CCTCGGCACCGGCTGGAATTCAACGAGCGATTACGGGGATTTCCCCGG	Fig. 3I-L
	mutZld2-F:TTCCAGCCGGTGCCGAGGCGGGAGGGGCTCGCACCTCTAATCCCGCC	Fig. 3I-L
	mutZld2-R:GGCGGGATTAGAGGTGCGAGCCCCTCCCGCCTCGGCACCGGCTGGAA	Fig. 3I-L
	mutZld3-F:CGGGAATTCCCTTCCGCTCGCAGCGAGCACTGCGCTGCGCAGACGCA	Fig. 3I-L
	mutZld3-R:TGCGTCTGCGCAGCGCAGTGCTCGCTGCGAGCGGAAGGGAATTCCCG	Fig. 3I-L
	mutZld4-F:ATACGGGTATACCCAAATGGAAGCGAGCCCATGTATATAGACCATTG	Fig. 3I-L
	mutZld4-R:CAATGGTCTATATACATGGGCTCGCTTCCATTTGGGTATACCCGTAT	Fig. 3I-L
***sog* primary ΔBcd123**	mutBcd1-F:GGCGGGACCTGCTCGCACCTCTCCACCCGCCAGGGTTTTCGGGACAT	Fig. 3M-P
	mutBcd1-R:ATGTCCCGAAAACCCTGGCGGGTGGAGAGGTGCGAGCAGGTCCCGCC	Fig. 3M-P
	mutBcd2-F:TATTATTATTGTGTCCAGTTTTCCACCGGAAAGCGGGAATTCCCTTC	Fig. 3M-P
	mutBcd2-R:GAAGGGAATTCCCGCTTTCCGGTGGAAAACTGGACACAATAATAATA	Fig. 3M-P
	mutBcd3-F:GCGCAGACGCATCGGCGTCCGTGGGCCGCTTACCAAAAAGATACGGG	Fig. 3M-P
	mutBcd3-R:CCCGTATCTTTTTGGTAAGCGGCCACGGACGCCGATGCGTCTGCGC	Fig. 3M-P
0.4ΔDl1-lacZ	mutDl1-F:CTATATGGCTGTATGGTGCGTTGAAATCAACGTAATCGCAGGTAGAATTC	Fig. S3A
	mutDl1-R:GAATTCTACCTGCGATTACGTTGATTTCAACGCACCATACAGCCATATAG	Fig. S3A
0.4ΔDl4-lacZ	mutDl4-F:CCAGTTTTAATCCGGAAAGCTTGAATTCAATTCCGCTCGCTGCCTGCACT	Fig. S3B
	mutDl4-R:AGTGCAGGCAGCGAGCGGAATTGAATTCAAGCTTTCCGGATTAAAACTGG	Fig. S3B
0.4ΔZld1-lacZ	mutZld1-F:CCGGGGAAATCCCCGTAATCGCTCGTTGAATTCCAGCCGGTGCCGAGG	Fig. S3C
	mutZld1-R:CCTCGGCACCGGCTGGAATTCAACGAGCGATTACGGGGATTTCCCCGG	Fig. S3C
0.4ΔZld3-lacZ	mutZld3-F:CGGGAATTCCCTTCCGCTCGCAGCGAGCACTGCGCTGCGCAGACGCA	Fig. S3D
	mutZld3-R:TGCGTCTGCGCAGCGCAGTGCTCGCTGCGAGCGGAAGGGAATTCCCG	Fig. S3D
0.4ΔBcd2-lacZ	mutBcd2-F:TATTATTATTGTGTCCAGTTTTCCACCGGAAAGCGGGAATTCCCTTC	Fig. S3E
	mutBcd2-R:GAAGGGAATTCCCGCTTTCCGGTGGAAAACTGGACACAATAATAATA	Fig. S3E
0.4ΔBcd3-lacZ	mutBcd3-F:GCGCAGACGCATCGGCGTCCGTGGGCCGCTTACCAAAAAGATACGGG	Fig. S3F
	mutBcd3-R:CCCGTATCTTTTTGGTAAGCGGCCACGGACGCCGATGCGTCTGCGC	Fig. S3F
0.4ΔDl1+Zld1-lacZ	mutDl1-F:CTATATGGCTGTATGGTGCGTTGAAATCAACGTAATCGCAGGTAGAATTC	Fig. S3G
	mutDl1-R:GAATTCTACCTGCGATTACGTTGATTTCAACGCACCATACAGCCATATAG	Fig. S3G
	mutZld1-F:CCGGGGAAATCCCCGTAATCGCTCGTTGAATTCCAGCCGGTGCCGAGG	Fig. S3G
	mutZld1-R:CCTCGGCACCGGCTGGAATTCAACGAGCGATTACGGGGATTTCCCCGG	Fig. S3G
0.4ΔDl4+Bcd2-lacZ	mutDl4-F:CCAGTTTTAATCCGGAAAGCTTGAATTCAATTCCGCTCGCTGCCTGCACT	Fig. S3H
	mutDl4-R:AGTGCAGGCAGCGAGCGGAATTGAATTCAAGCTTTCCGGATTAAAACTGG	Fig. S3H
	mutBcd2-F:TATTATTATTGTGTCCAGTTTTCCACCGGAAAGCGGGAATTCCCTTC	Fig. S3H
	mutBcd2-R:GAAGGGAATTCCCGCTTTCCGGTGGAAAACTGGACACAATAATAATA	Fig. S3H

[1]All primer sequences are presented in the 5' to 3' direction relative to the physiological orientation of *sog* transcription.

[2]Indicates section(s) in the manuscript where the primer was used.

[3]For primers used in site-directed mutagenesis, the nucleotides used to introduce mutations are designated in red.

Key: F = forward; R = reverse.

Table S2

Dorsal binding motifs in *sog* primary enhancer									
in vitro assays (Footprint or SELEX)				*in vivo* assays (Bacterial 1-hybrid)				Selected motifs	
Position	Orientation	Sequence	-log*P*	Position	Orientation	Sequence	-log*P*	Position	Sequence
1	R	GGGATTTCCC	6.25	1	R	GGGATTTCCC	5.77	1	GGGATTTCCC
75	D	AGGGTTTTCG	2.89	76	D	GGGTTTTCGG	3.37	76	GGGTTTTCGG
91	D	GGGATATTCC	4.28	92	D	GGATATTCCC	4.48	92	GGATATTCCC
175	R	GGGAATTCCC	5.95	175	D	GGGAATTCCC	6.25	175	GGGAATTCCC
254	D	GGGTATACCC	3.15	254	D	GGGTATACCC	4.63	254	GGGTATACCC
Match cutoff			2.50	Match cutoff			2.69		

Table S2. Dl-binding motifs in the *sog* primary enhancer were identified by ClusterDraw (Papatsenko, 2007) fed with PFMs produced from two independent binding sites alignments. One was built using motif sequences obtained from *in vitro* binding assays such as DNase footprinting and SELEX (Papatsenko, 2007) and the other from *in vivo* bacterial one-hybrid system data (Zhu, *et al*., 2011). Dl-binding motifs that met the following two selection criteria were finally selected as putative functional motifs: 1) Dl-binding motifs identified in both of the independent searches performed with two different PFMs, 2) Dl-binding motifs with cumulative match probability (-log*P*) values higher than the match cutoff values (2.50 and 2.69). Binding motifs whose match probabilities (-log*P*) were higher than the match cutoff value are underlined. Five Dl-binding motifs in the *sog* primary enhancer were selected as functional Dl-binding sites.

Table S3

Zelda binding motifs in *sog* primary enhancer									
in vitro assays (Footprint or SELEX)				*in vivo* assays (Bacterial 1-hybrid)				Selected motifs	
Position	Orientation	Sequence	-log*P*	Position	Orientation	Sequence	-log*P*	Position	Sequence
19	D	CAGGTAG	4.34	18	D	GCAGGTAG	5.03	18	GCAGGTAG
48	R	CAGGTCC	3.42	48	R	GCAGGTCC	3.13	48	GCAGGTCC
194	R	CAGGCAG	3.85	194	R	GCAGGCAG	3.70	194	GCAGGCAG
270	R	CAGGCAT	3.24	269	D	GCAGGCAT	2.88	269	GCAGGCAT
Match cutoff			2.50	Match cutoff			2.69		

Table S3. Zld-binding motifs in the *sog* primary enhancer were identified by ClusterDraw (Papatsenko, 2007) fed with PFMs produced from two independent binding site alignments. One was built by motif sequences from ChIP-chip analysis performed with anti-Zld antibody (Nien *et al*., 2011) and the other from *in vivo* bacterial one-hybrid system data (Zhu, *et al*., 2011). Zld-binding motif alignments from *in vitro* binding assays are not currently available. Zld-binding motifs that met the following two selection criteria were finally selected as putative functional motifs: 1) Zld-binding motifs identified in both independent searches performed with two different PFMs, 2) Zld-binding motifs whose cumulative match probability (-log*P*) values were higher than the match cutoff value. Binding motifs with match probabilities (-log*P*) higher than the match cutoff value are underlined. Four Zld-binding motifs in the *sog* primary enhancer were selected as putatively functional Zld-binding sites.

Table S4

Snail binding motifs in *sog* primary enhancer									
in vitro assays (Footprint or SELEX)				*in vivo* assays (Bacterial 1-hybrid)				Selected motifs	
Position	Orientation	Sequence	-log*P*	Position	Orientation	Sequence	-log*P*	Position	Sequence
18	D	GCAGGTAG	3.86	17	D	CGCAGGTAG	3.02	18	GCAGGTAG
33	D	GCCGGTGC	2.59	32	D	AGCCGGTGC	3.05	33	GCCGGTGC
48	R	GCAGGTCC	3.59	48	R	TGCAGGTCC	3.13	48	GCAGGTCC
Match cutoff			2.50	Match cutoff			2.69		

Table S4. Sna-binding motifs in the *sog* primary enhancer were identified by ClusterDraw (Papatsenko, 2007) fed with PFMs produced from two independent binding sites alignments. One was built using motif sequences obtained from *in vitro* binding assays such as DNase footprinting and SELEX (Papatsenko, 2007) and the other from *in vivo* bacterial one-hybrid system data (Zhu, *et al*., 2011). Sna-binding motifs that met the following two selection criteria were finally selected as putative functional motifs: 1) Sna-binding motifs identified in both independent searches performed with two different PFMs, 2) Sna-binding motifs with cumulative match probability (-log*P*) values higher than the match cutoff value. Binding motifs whose match probabilities (-log*P*) were higher than the match cutoff value are underlined. Three Sna-binding motifs in the *sog* primary enhancer were selected as putatively functional Sna-binding sites.

Table S5

Bicoid binding motifs in *sog* primary enhancer									
in vitro assays (Footprint or SELEX)				*in vivo* assays (Bacterial 1-hybrid)				Finally selected	
Position	Orientation	Sequence	-log*P*	Position	Orientation	Sequence	-log*P*	Position	Sequence
11	D	CGTAATCGC	3.18						
63	D	TCTAATCCC	4.27	64	D	CTAATCC	3.80	64	CTAATCC
160	D	TTTAATCCG	3.96	161	D	TTAATCC	4.03	161	TTAATCC
229	D	CGTAAGCCG	3.39	230	D	GTAAGCC	2.80	230	GTAAGCC
Match cutoff			2.50	Match cutoff			2.69		

Table S5. Bcd-binding motifs in the *sog* primary enhancer were identified by ClusterDraw (Papatsenko, 2007) fed with PFMs produced from two independent binding sites alignments. One was built using motif sequences obtained from *in vitro* binding assays such as DNase footprinting and SELEX (Papatsenko, 2007) and the other from *in vivo* bacterial one-hybrid system data (Zhu, *et al*., 2011). Bcd-binding motifs that met the following two selection criteria were finally selected as putative functional motifs: 1) Bcd-binding motifs identified in both independent searches performed with two different PFMs, 2) Bcd-binding motifs with cumulative match probability (-log*P*) values higher than the match cutoff value. Bcd-binding motifs whose match probabilities (-log*P*) were higher than the match cutoff value are underlined. Three Bcd-binding motifs in the *sog* primary enhancer were selected as putatively functional Bcd-binding sites.

Table S6. Primary and shadow enhancer sequences that direct *sog* expression in the neurogenic ectoderm of early *Drosophila* embryos

Enhancer DNA Sequence (5' to 3' direction)
>sog_primary_392bp GTTGCCAATGCCATTGCGCATACGCCGTGTCGTCTATATGGCTATATGGCTATATGGCTGTATGGTGCGGGGAAATCCCC GTAATCGCAGGTAGAATTCCAGCCGGTGCCGAGGCGGGACCTGCTCGCACCTCTAATCCCGCCAGGGTTTTCGGGACATG GGATATTCCCGACGGCACAGCATAGCACTCCGTTTTCTTTTTTTTTTTTTATTATTATTGTGTCCAGTTTTAATCCGGAA AGCGGGAATTCCCTTCCGCTCGCTGCCTGCACTGCGCTGCGCAGACGCATCGGCGTCCGTAAGCCGCTTACCAAAAAGAT ACGGGTATACCCAAATGGATGCCTGCCCATGTATATAGACCATTGGGTGGTATGGACCATGGACCATAAAGC
>sog_shadow_884bp GAGAAGGAGGAGAAGTTGGTTGAGAGGTCATCGTTGCGATTCTGCGATTCAGCAGTTCCACAGAAGGTGTCGTAATCCTG GACGCAAGGGTGCACGGACCAACTGACAGGGGCAAGTGCGTCCTGTGCCACCAGATGACGCACGATGCGGCCGGAAAAAC CCAAAATCAAAAACCGAAAACCGAAAACCTGGTCAGAGTTTCCGAAAACCAAAGAGCCAACATCGAATGCGGCACAATAA CCCGATTGTCTGCGAATACCCACGATGATCTAGAATCGCACGGAGAGCACTCTCACGCATCCGTGGCCATATGGGTGCGG CCAAATCGGAAATTCCCAGGACAGGTAGAATGCATTGGATATACGGGTATACGGATTGGAATTGGGATTGGGATTGGGAC TAGCACCAGGTTGCAACGCCCGCCAAGAAGCCAATTTAAATAAGCAGCATAAACAAAAGCGACAGCGTTTTATGATCCCC GCTCCTTATCCTTGCACAAGGATATCGCCATGGCCACGCAGGTAGGAATAGCAGATATGGCGGCAATGATGCGCCAACCG CACTGCTTCGTCCTGGTCCTGGTCGGATGGGCTTTTCCCACGCAACCGCGACCTTATCTGCGCCCCTTTTATGAGGCTGC ATCTGTTTTCGCACCTCGATGCCGTTGGCATTATAGCCACATGTGTATGGTGGGAATTTCCGATCGACCAGCCTACCTGT TCCGCTGAAACCCGGGAATCTGTCCATCCTGAGCTTCCACACACACACACACACACACACAGGTCAGTCGGCATCAATTG GCTGCCATAAACATATAACAATCAATATTGAATCCTTTATCGTAGAATTTGTTGTATATGCCCATTGCAGTCCTTCGATT AAAT

Table S6. The primer sequences used for genomic PCR amplification are underlined. Primary enhancer for *sog* expression was initially cloned by a computational algorithm to search the *Drosophila* genome for clusters that contain three or more optimal sites (Markstein, *et al*., 2002). ChIP-chip assays predicted that many of the Dl target genes contained two separate enhancers for the same or similar expression pattern, and some of the potential secondary enhancers identified by the ChIP-chip assays were predicted to be located a long distance from the transcription start sites of Dl target genes. The *sog* shadow enhancer was one of the predicted enhancers, and its transcriptional activity was tested by P element-mediated germline transformation followed by whole-mount *in situ* hybridization (Hong, *et al*., 2008). The genomic coordinates for the primary and shadow enhancers are X: 15,622,698-15,627,089 and X: 15,647,477-15,646,594, respectively.

Figure S1

```
>dorsal_PFM
A:    0.018 0.000 0.196 0.652 0.420 0.071 0.000 0.000 0.000 0.018
C:    0.125 0.018 0.045 0.063 0.018 0.000 0.054 0.580 1.000 0.768
G:    0.607 0.938 0.750 0.161 0.063 0.000 0.045 0.018 0.000 0.134
T:    0.250 0.045 0.009 0.125 0.500 0.929 0.902 0.402 0.000 0.080

>snail_PFM
A:    0.417 0.000 1.000 0.333 0.083 0.083 0.083 0.083
C:    0.000 1.000 0.000 0.000 0.000 0.000 0.083 0.250
G:    0.500 0.000 0.000 0.583 0.917 0.083 0.500 0.583
T:    0.083 0.000 0.000 0.083 0.000 0.833 0.333 0.083

>zelda_ChIP-chip_PFM
A:    0.000 1.000 0.000 0.000 0.000 1.000 0.333
C:    0.815 0.000 0.000 0.000 0.148 0.000 0.074
G:    0.000 0.000 1.000 1.000 0.000 0.000 0.519
T:    0.185 0.000 0.000 0.000 0.852 0.000 0.074

>bicoid-DNase footprinting_PFM
A:    0.147 0.176 0.176 0.971 1.000 0.000 0.000 0.029 0.118
C:    0.353 0.353 0.000 0.000 0.000 0.059 1.000 0.794 0.353
G:    0.324 0.147 0.000 0.029 0.000 0.088 0.000 0.059 0.324
T:    0.176 0.324 0.824 0.000 0.000 0.853 0.000 0.118 0.206

>sim PFM
A:    0.385 0.308 0.769 0.000 0.000 0.000 0.000
C:    0.000 0.000 0.000 1.000 0.000 0.000 0.000
G:    0.538 0.077 0.231 0.000 1.000 0.000 1.000
T:    0.077 0.615 0.000 0.000 0.000 1.000 0.000
```

Figure S1. *In vitro* PFMs of Dl-, Zld-, Sna-, Bcd-, and Sim-binding DNA motifs. The Dl, Sna, Bcd, and Sim PFMs were obtained from *in vitro* binding assays such as DNase footprinting and SELEX (Papatsenko, 2007). The Zld PFM was obtained from motif alignment of ChIP-chip analyses performed with anti-Zld antibody (Nien, *et al*., 2011). Zld-binding motif alignments from *in vitro* binding assays are not currently available.

Figure S2

```
>dl_NBT_PFM
A:    0     0     0.063 0.563 0.563 0.25  0.063 0     0     0.219
C:    0     0     0     0     0     0     0.25  1     0.938 0.625
G:    1     1     0.75  0     0     0     0     0     0     0.156
T:    0     0     0.188 0.438 0.438 0.75  0.688 0     0.063 0

>sna_SOLEXA_PFM
A:    0.5   0.733 0.029 0.999 0.007 0.003 0.011 0.026 0.054
C:    0.035 0     0.965 0     0.001 0.004 0.01  0.002 0.148
G:    0.373 0.174 0.005 0     0.987 0.992 0.016 0.92  0.586
T:    0.092 0.093 0     0.001 0.005 0.002 0.963 0.052 0.212

>zld_SOLEXA_PFM
A:    0.071 0.009 0.993 0     0.001 0.005 0.993 0.1
C:    0     0.984 0.001 0     0     0     0.007 0.099
G:    0.628 0     0.001 0.961 0.986 0.004 0     0.744
T:    0.301 0.007 0.004 0.039 0.013 0.991 0     0.058

>bcd_SOLEXA_PFM
A:    0.121 0.102 1     0.986 0     0.005 0.127
C:    0.285 0.014 0     0.014 0.009 0.875 0.619
G:    0.126 0.005 0     0     0.093 0.014 0.083
T:    0.467 0.879 0     0     0.898 0.106 0.171

>sim_tgo_SANGER_PFM
A:    0.227 0.273 1     0     0     0     0     0.864 0     0
C:    0.182 0     0     1     0     0     0     0.136 0.81  0.619
G:    0.591 0.091 0     0     1     0     1     0     0.048 0.143
T:    0     0.636 0     0     0     1     0     0     0.143 0.238
```

Figure S2. *In vivo* PFMs of Dl-, Zld-, Sna-, Bcd, and Sim-binding DNA motifs. The Dl, Sna, Bcd, and Sim PFMs were obtained from *in vivo* bacterial one-hybrid system data obtained from the FlyFactorSurvey database (Zhu, *et al*., 2011).

Figure S3

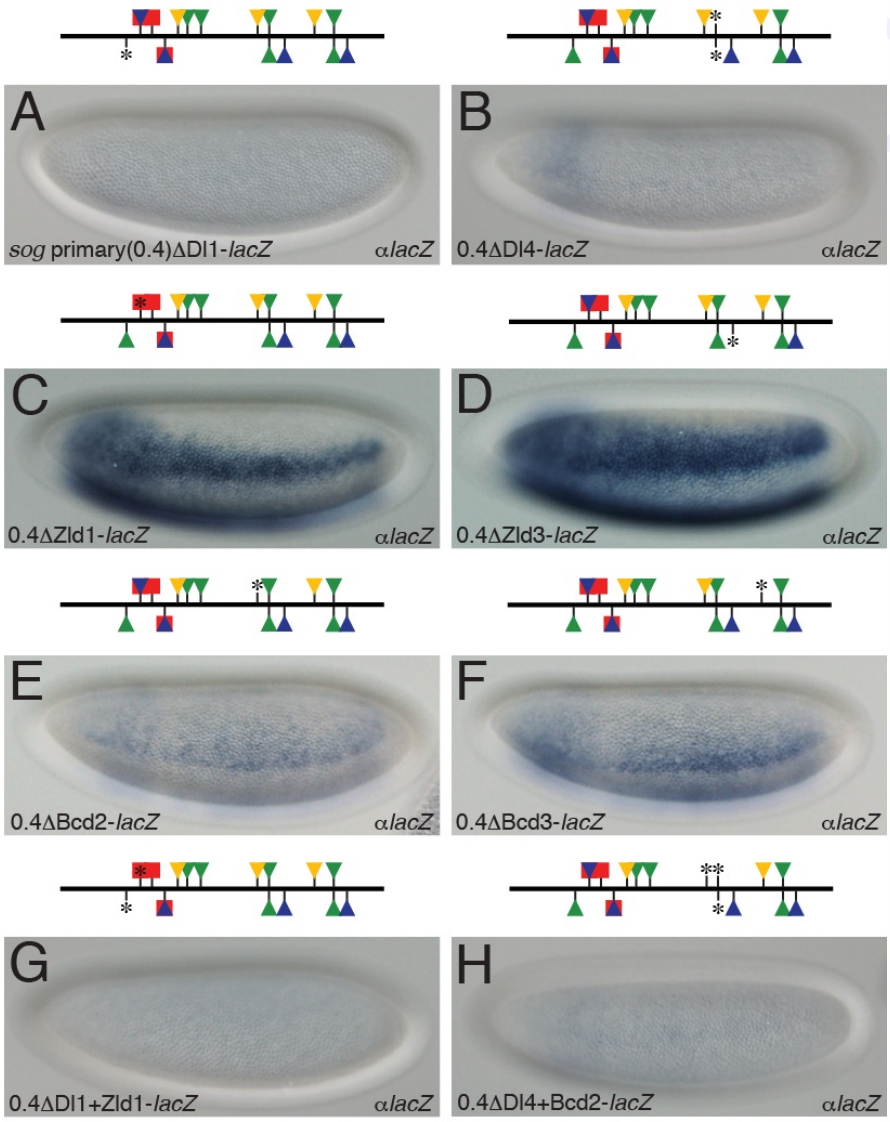


Figure S3. Broad lacZ expression in the presumptive neurogenic ectoderm requires close linkages between Dl- and either Zld- or Bcd-binding sites in the *sog* primary enhancer. Removal of all of the Zld- and Bcd-binding sites in the *sog* primary enhancer led to catastrophic reduction in lacZ expression width in the neurogenic ectoderm (Fig. 3I and M). Additional mutagenesis studies were performed to further investigate the effect of transcriptional interaction between Dl and either Zld or

Bcd on generating the broad and normal *sog* expression pattern. All of the mutant *sog* primary enhancers 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 P element-mediated germline transformation (Rubin and Spradling, 1982). Cellularizing transgenic embryos (approximately early stage 5) are oriented with the anterior to the left and dorsal side up. Expression of lacZ in the transgenic embryos was visualized by *in situ* hybridization with an antisense lacZ RNA probe labeled with digoxigenin (Dig). Binding sites of Dl (green), Zld (blue), Sna (red), and Bcd (yellow) in the *sog* primary enhancer are depicted on the top of each panel. Triangles and squares represent binding sites for a transcriptional activator and repressor, respectively. Mutagenized Dl-, Zld-, Sna-, and Bcd-binding sites are marked with asterisks (*).

The *sog* primary enhancer contains five Dl-binding sites (Fig. 2C and Table S2), the first (Dl) and forth (D4) of which exhibit the highest cumulative match probabilities (-log*P*) (Table S2), suggesting that those Dl-binding sites may play a critical role in *sog* expression. To test this hypothesis, the two Dl-binding sites were mutagenized one at a time and the lacZ expression patterns directed by the mutant *sog* primary enhancer were investigated in the transgenic embryos (Fig. S3A and B). The mutant version of the *sog* primary enhancer containing either the mutant D1- or D4-binding sites completely failed to activate lacZ in the neurogenic ectoderm, which is very analogous to the lack of expression mediated by the construct containing no Dl-binding site (Fig. 3E). This result suggests that all *sog* expression depends on Dl-binding sites and among them, both the first and forth Dl-binding sites are indispensable for broad *sog* expression.

A recent study showed that the minimal *sog* shadow enhancer contains two clusters of closed linked Dl-, Zld- and Sna-binding sites, where the Zld- and Sna-binding sites almost completely coincide (Shin and Hong, 2016). Among the clusters of binding sites, Zld-binding sites play a critical role in the transcriptional synergy between Dl and Zld to produce the broad *sog* expression pattern in the neurogenic ectoderm in an early developing embryo. The fact that the *sog* primary enhancer also contains one cluster of Dl-, Zld- and Sna-binding sites (Fig. 2C) raised the possibility that the first Zld-binding site (Z1) is also necessary for the transcriptional synergy with Dl. To test this possibility, the consensus sequence of the first Dl-binding site was mutagenized (Fig. S3C). The *sog* primary enhancer with a mutant Z1 site directed narrow lacZ expression in the neurogenic ectoderm, although the lacZ expression directed by the enhancer containing four mutant Zld-binding sites (Fig. 3I) is somewhat narrower than that of the Z1 mutant enhancer. However, lack of the third Zld-binding site that shows the second highest match probability among the four Zld-binding sites appeared to be tolerable in terms of lacZ expression (Fig. S3D). These results suggest that the first Zld-binding site (Z1) synergistically interacts with the nearby linked Dl-binding site (D1) to facilitate DNA-binding affinity of the very limited amount of nuclear Dl protein in the dorsal half of the neurogenic ectoderm, and that much narrower lacZ expression directed by the primary enhancer containing four mutant Zld-binding sites (Fig. 3I) is attributed at least in part to the first Zld-binding site.

The *sog* primary enhancer contains two clusters composed of Dl-, Bcd- and Zld-binding sites (Fig. 2C), which are not identified in the *sog* shadow enhancer (Shin and Hong, 2016). The fact that the third Zld-binding site (Z3) located in one of the clusters is dispensable (Fig. S3D) raised the possibility that the remaining Bcd- and Dl-binding sites (B2 and D4) within the cluster may

synergistically interact with each other. As expected, the loss of the second Bcd-binding site (B2) led to a severe reduction in lacZ expression width and strength (Fig. S3E). In addition, removal of the third Bcd-binding site (B3) also resulted in decreased lacZ expression width and strength (Fig. S3F). These results suggest that Bcd-binding sites liked with Dl-binding sites also participate in the transcriptional synergy to ensure broad *sog* expression. Removal of the Dl-binding sites (D1 and D4) linked with the first Zld- and the second Bcd-binding sites (Z1 and B2) completely abolished lacZ expression in the transgenic embryo (Fig. S3G and H), suggesting that the residual lacZ expression restricted to the ventral part of the neurogenic ectoderm in the absence of Z1, B2 or B3 sites was generated by the nuclear Dl gradient alone.

Taken altogether, at least two pairs of Dl-, Zld- and Bcd-binding sites (D1-Z1 and D4-B2) seem to synergistically interact with each other to enhance affinity between Dl-binding sites and the very low amount of nuclear Dl gradient so that even a small amount of Dl protein can produce a transcriptionally functional outcome throughout the neurogenic ectoderm. The reduced lacZ expression widths in the absence of Zld- and Bcd-binding sites are proof that broad *sog* expression is generated by at least a dual model of transcriptional synergy (Dl-Zld and Dl-Bcd) in the *sog* primary enhancer.

Supplementary REFERENCES

Campos-Ortega, J.A. and Hartenstein, V. (1985) The embryonic development of *Drosophila melanogaster*. Springer-Verlag, Berlin; Heidelberg.

Hendrix, D.A., Hong, J.W., Zeitlinger, J., Rokhsar, D.S. and Levine, M.S. (2008) Promoter elements associated with RNA Pol II stalling in the *Drosophila* embryo. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7762-7767.

Hong, J.W., Hendrix, D.A. and Levine, M.S. (2008) Shadow enhancers as a source of evolutionary novelty. *Science* 321, 1314.

Hong, J.W., Park, K.W. and Levine, M.S. (2013) Temporal regulation of *single-minded* target genes in the ventral midline of the *Drosophila* central nervous system. *Dev. Biol.* 380, 335-343.

Ip, Y.T., Park, R.E., Kosman, D., Yazdanbakhsh, K. and Levine, M. (1992) *dorsal-twist* interactions establish *snail* expression in the presumptive mesoderm of the *Drosophila* embryo. *Genes Dev.* 6, 1518-1530.

Jiang, J. and Levine, M. (1993) Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell* 72, 741-752.

Markstein, M., Markstein, P., Markstein, V. and Levine, M.S. (2002) Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the *Drosophila* embryo. *Proc. Natl. Acad. Sci. U.S.A.* 99, 763-768.

Nien, C.Y., Liang, H.L., Butcher, S., Sun, Y., Fu, S., Gocha, T., Kirov, N., Manak, J.R. and Rushlow, C. (2011) Temporal coordination of gene networks by Zelda in the early *Drosophila* embryo. *PLoS Genet.* 7, e1002339.

Papatsenko, D. (2007) ClusterDraw web server: a tool to identify and visualize clusters of binding motifs for transcription factors. *Bioinformatics* 23, 1032-1034.

Rubin, G.M. and Spradling, A.C. (1982) Genetic transformation of *Drosophila* with transposable elements vectors. *Science* 218, 348-353.

Shin, D.H. and Hong, J.W. (2016) The *short gastrulation* shadow enhancer employs dual modes of transcriptional synergy. *International Journal of Developmental Biology* In press (http://dx.doi.org/10.1387/ijdb.160165jh)

Small, S., Blair, A. and Levine, M. (1992) Regulation of *even-skipped* stripe 2 in the *Drosophila* embryo. *EMBO J.* 11, 4047-4057.

Stathopoulos, A. and Levine, M. (2005) Localized repressors delineate the neurogenic ectoderm in the early *Drosophila* embryo *Dev. Biol.* 280, 482-493.

Stathopoulos, A., Van Drenth, M., Ervies, A., Markstein, M. and Levine, M. (2002) Whole-genome analysis of dorsal-ventral patterning in the *Drosophila* embryo. *Cell* 111, 687-701.

Zhu, L.J., Christensen, R.G., Kazemian, M., Hull, C.J., Enuameh, M.S., Basciotta, M.D., Brasefield, J.A., Zhu, C., Asriyan, Y., Lapointe, D.S., Sinha, S., Wolfe, S.A. and Brodsky, M.H. (2011) FlyFactorSurvey: a database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res*. 39, D111-117.