THE INTERNATIONAL JOURNAL OF DEVELOPMENTAL BIOLOGY
www.intjdevbiol.com

doi: 10.1387/ijdb.160165jh

SUPPLEMENTARY MATERIAL

corresponding to:

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

DONG-HYEON SHIN and JOUNG-WOO HONG*

Table S1. DNA oligonucleotide sequences used in this study.

Enhancer Name	Sequence (5' to 3' direction)[1]		Note[2]
0.884 kb (full length)	F1:GAGAAGGAGGAGAAGTTGGT	R1:ATTTAATCGAAGGACTGCAA	Fig. 1
0.732 kb	F2:TACAGCGTATGGCGATTT	R1:ATTTAATCGAAGGACTGCAA	Fig. 1
0.616 kb	F3:AGAGCACTCTCACGCATC	R1:ATTTAATCGAAGGACTGCAA	Fig. 1
0.404 kb	F4:GCTCCTTATCCTTGCACA	R1:ATTTAATCGAAGGACTGCAA	Fig. 1
0.213 kb	F5:TATAGCCACATGTGTATGGTG	R1:ATTTAATCGAAGGACTGCAA	Fig. 1
0.731 kb	F1:GAGAAGGAGGAGAAGTTGGT	R6:GTTTCAGCGGAACAGGTAG	Fig. 1
0.542 kb	F1:GAGAAGGAGGAGAAGTTGGT	R7:CGCCATATCTGCTATTCCTA	Fig. 1
0.351 kb	F1:GAGAAGGAGGAGAAGTTGGT	R8:ATTCTACCTGTCCTGGGAAT	Fig. 1
0.686 kb	F1-1:GATTCAGCAGTTCCACAGAA	R6:GTTTCAGCGGAACAGGTAG	Fig. 1
0.610 kb	F1-1:GATTCAGCAGTTCCACAGAA	R6-1:GTGCGAAAACAGATGCAG	Fig. 1
0.502 kb	F2:TACAGCGTATGGCGATTT	R6-1:GTGCGAAAACAGATGCAG	Fig. 1
0.582 kb	F2:TACAGCGTATGGCGATTT	R6:GTTTCAGCGGAACAGGTAG	Fig. 1
0.450 kb	F3:AGAGCACTCTCACGCATC	R6:GTTTCAGCGGAACAGGTAG	Fig. 1
0.582ΔZld123	mutZld1-F:CAAATCGGAAATTCCCAGGACTCGTTGAATGCATTGGATATACGGGT[3]		Fig. 3B
	mutZld1-R:ACCCGTATATCCAATGCATTCAACGAGTCCTGGGAATTTCCGATTTG		Fig. 3B
	mutZld2-F:AGGATATCGCCATGGCCACGCTCGTTGGAATAGCAGATATGGCGGCA		Fig. 3B
	mutZld2-R:TGCCGCCATATCTGCTATTCCAACGAGCGTGGCCATGGCGATATCCT		Fig. 3B
	mutZld3-F:GGAATTTCCGATCGACCAGCCAACGAGTTCCGCTGAAACCCGGGAAT		Fig. 3B
	mutZld3-R:ATTCCCGGGTTTCAGCGGAACTCGTTGGCTGGTCGATCGGAAATTCC		Fig. 3B
0.582ΔZld1 3	mutZld1-F:CAAATCGGAAATTCCCAGGACTCGTTGAATGCATTGGATATACGGGT		Fig. 3C
	mutZld1-R:ACCCGTATATCCAATGCATTCAACGAGTCCTGGGAATTTCCGATTTG		Fig. 3C
	mutZld3-F:GGAATTTCCGATCGACCAGCCAACGAGTTCCGCTGAAACCCGGGAAT		Fig. 3C
	mutZld3-R:ATTCCCGGGTTTCAGCGGAACTCGTTGGCTGGTCGATCGGAAATTCC		Fig. 3C
0.582ΔZld1	mutZld1-F:CAAATCGGAAATTCCCAGGACTCGTTGAATGCATTGGATATACGGGT		Fig. 3D
	mutZld1-R:ACCCGTATATCCAATGCATTCAACGAGTCCTGGGAATTTCCGATTTG		Fig. 3D
0.582ΔZld 2	mutZld2-F:AGGATATCGCCATGGCCACGCTCGTTGGAATAGCAGATATGGCGGCA		Fig. 3E
	mutZld2-R:TGCCGCCATATCTGCTATTCCAACGAGCGTGGCCATGGCGATATCCT		Fig. 3E
0.582ΔZld 3	mutZld3-F:GGAATTTCCGATCGACCAGCCAACGAGTTCCGCTGAAACCCGGGAAT		Fig. 3F
	mutZld3-R:ATTCCCGGGTTTCAGCGGAACTCGTTGGCTGGTCGATCGGAAATTCC		Fig. 3F
0.582 Dl3-8bp-Zld1	F:GTGCGGCCAGATCGGAAATTCCCAGGACTCTAGGACAGGTAGAATGCATTGGATA		Fig. 3G
	R:TATCCAATGCATTCTACCTGTCCTAGAGTCCTGGGAATTTCCGATCTGGCCGCAC		Fig. 3G
0.582 Dl5-7bp-Zld3	F:GTGTATGGTGGGAATTTCCGATCGACCAGCATCGACCCTACCTGTTCCGCTGAA		Fig. 3H
	R:TTCAGCGGAACAGGTAGGGTCGATGCTGGTCGATCGGAAATTCCCACCATACAC		Fig. 3H
0.582ΔBcd234	mutBcd2-F:CGATGCGGCCGGAAAAACCCAAAGGCGAAAACCGAAAACCGAAAACCT		Fig. 4D
	mutBcd2-R:AGGTTTTCGGTTTTCGGTTTTCGCCTTTGGGTTTTTCCGGCCGCATCG		Fig. 4D
	mutBcd3-F:CAACATCGAATGCGGCACAACGGCCCGATTGTCTGCGAATACCCAC		Fig. 4D
	mutBcd3-R:GTGGGTATTCGCAGACAATCGGGCCGTTGTGCCGCATTCGATGTTG		Fig. 4D
	mutBcd4-F:ACGGATTGGAATTGGGATTGGCCAAGGGACTAGCACCAGGTTGCAA		Fig. 4D
	mutBcd4-R:TTGCAACCTGGTGCTAGTCCCTTGGCCAATCCCAATTCCAATCCGT		Fig. 4D
0.582ΔBcd2	mutBcd2-F:CGATGCGGCCGGAAAAACCCAAAGGCGAAAACCGAAAACCGAAAACCT		Fig. 4E
	mutBcd2-R:AGGTTTTCGGTTTTCGGTTTTCGCCTTTGGGTTTTTCCGGCCGCATCG		Fig. 4E
0.582ΔBcd3	mutBcd3-F:CAACATCGAATGCGGCACAACGGCCCGATTGTCTGCGAATACCCAC		Fig. 4F
	mutBcd3-R:GTGGGTATTCGCAGACAATCGGGCCGTTGTGCCGCATTCGATGTTG		Fig. 4F
0.582ΔBcd4	mutBcd4-F:ACGGATTGGAATTGGGATTGGCCAAGGGACTAGCACCAGGTTGCAA		Fig. 4G
	mutBcd4-R:TTGCAACCTGGTGCTAGTCCCTTGGCCAATCCCAATTCCAATCCGT		Fig. 4G
0.582 Dl1-7bp-Bcd2	F:ACTAGTGATTGCGGATCCGGAAAAACCCATCGACCAAAATCAAAAACCGAAAACC		Fig. 4H

	R:GGTTTTCGGTTTTTGATTTTGGTCGATGGGTTTTTCCGGATCCGCAATCACTAGT	Fig. 4H
0.582 Dl2-7bp-Bcd3	F:ACTAGTGATTGCGGATCCGGAAAAACCCATCGACCAAAATCAAAAACCGAAAACC	Fig. 4I
	R:GGTTTTCGGTTTTTGATTTTGGTCGATGGGTTTTTCCGGATCCGCAATCACTAGT	Fig. 4I
***sog* primary enhancer**	F:GTTGCCAATGCCATTGCGCATACGCCGTGTCG	Fig. 5B
	R:GCTTTATGGTCCATGGTCCATACCAC	Fig. 5B
***sog* primary ΔBcd123**	mutBcd1-F:GGCGGGACCTGCTCGCACCTCTCCACCCGCCAGGGTTTTCGGGACAT	Fig. 5C
	mutBcd1-R:ATGTCCCGAAAACCCTGGCGGGTGGAGAGGTGCGAGCAGGTCCCGCC	Fig. 5C
	mutBcd2-F:TATTATTATTGTGTCCAGTTTTCCACCGGAAAGCGGGAATTCCCTTC	Fig. 5C
	mutBcd2-R:GAAGGGAATTCCCGCTTTCCGGTGGAAAACTGGACACAATAATAATA	Fig. 5C
	mutBcd3-F:GCGCAGACGCATCGGCGTCCGTGGGCCGCTTACCAAAAAGATACGGG	Fig. 5C
	mutBcd3-R:CCCGTATCTTTTTGGTAAGCGGCCACGGACGCCGATGCGTCTGCGC	Fig. 5C

[1]All primer sequences are presented in the 5' to 3' direction relative to the physiological orientation of *short gastrulation* (*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 the *sog* shadow 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
153	R	GGGTTTTTCC	4.76	152	R	GGTTTTTCCG	4.29	152	GGTTTTTCCG
175	R	TCGGTTTTCG	2.51						
203	R	TTGGTTTTCG	2.65						
252	R	GGGTATTCGC	3.09	251	R	GGGTATTCGC	3.88	251	GGGTATTCGC
328	R	GGGAATTTCC	5.56	327	F	GGAAATTCCC	4.57	327	GGAAATTCCC
				364	F	GGGTATACGG	2.83		
				499	F	GGATATCGCC	2.87		
589	F	GGGCTTTTCC	4.59	588	F	GGCTTTTCCC	3.86	588	GGCTTTTCCC
692	F	GGGAATTTCC	5.56	691	R	GGAAATTCCC	4.57	691	GGAAATTCCC
Match cutoff			3.06	Match cutoff			3.63		

Table S2. Dorsal (Dl)-binding motifs in the *sog* shadow enhancer were identified by ClusterDraw (1) fed with two different position frequency matrices (PFMs). One was built by motif sequences obtained from *in vitro* binding assays such as DNase footprinting assay and systematic evolution of ligands by exponential enrichment (SELEX) (http://line.bioinfolab.net/webgate/help.htm#mtfform) (1) and the other from *in vivo* bacterial one-hybrid (B1H) system (http://mccb.umassmed.edu/ffs/) (2). Dl-binding motifs that met the following two selection criteria were finally selected as putative functional Dl-binding motifs: 1) Dl-binding motifs commonly identified by two independent searches performed with two different PFMs, 2) Dl-binding motifs whose cumulative match probability (-log*P*) values were higher than the match cutoff value. The binding motifs whose match probabilities (-log*P*) were higher than the match cutoff value are underlined.

Table S3

Zelda binding motifs in the *sog* shadow enhancer.									
ChIP-chip				*In vivo* assays (Bacterial 1-hybrid)				Selected motifs	
Position	Orientation	Sequence	-log*P*	Position	Orientation	Sequence	-log*P*	Position	Sequence
342*	F	CAGGTAG	4.34	340*	F	ACAGGTAG	4.23	340*	ACAGGTAG
407	F	CAGGTTG	2.94						
519*	F	CAGGTAG	4.34	516*	F	GCAGGTAG	5.03	516*	GCAGGTAG
				530	F	GCAGATAT	2.71		
614	R	CAGATAA	2.55	613	R	GCAGATAA	2.84		
713*	R	CAGGTAG	4.34	712*	R	ACAGGTAG	4.23	712*	ACAGGTAG
Match cutoff			3.06	Match cutoff			3.63		

Table S3. Zelda (Zld)-binding motifs in the *sog* shadow enhancer were identified by ClusterDraw (1) algorithm fed with PFMs constructed from two independent motif alignments. One was built by motif sequences from chromatin immunoprecipitation-tiling array (ChIP-chip) analysis performed with anti-Zld antibody (3) and the other from an *in vivo* B1H system (http://mccb.umassmed.edu/ffs/) (2). Zld-binding motif alignments obtained 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 for Zld binding: 1) Zld-binding motifs commonly identified by two 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. The binding motifs whose match probabilities (-log*P*) were higher than the match cutoff value are underlined. Zld motifs overlapping with Sna-binding motifs are labeled with asterisks.

Table S4

Snail binding motifs in the *sog* shadow 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
84	F	GCAAGGGT	2.63						
				104	F	GACAGGGGC	3.71		
112	F	GCAAGTGC	3.94						
				127	R	ATCTGGTGG	3.09		
				184	R	ACCAGGTTT	2.62		
341*	F	ACAGGTAG	3.56	339*	F	GACAGGTAG	4.03	339*	GACAGGTAG
406	F	CCAGGTTG	2.80	404	F	ACCAGGTTG	2.95		
518*	F	GCAGGTAG	3.86	517*	F	CGCAGGTAG	3.02	517*	CGCAGGTAG
				611	R	AAAAGGGGC	2.61		
639	R	ACAGATGC	2.96	638	R	AACAGATGC	3.15		
677	R	ACATGTGG	3.45	676	R	CACATGTGG	3.03	676	CACATGTGG
713*	R	ACAGGTAG	3.56	712*	R	AACAGGTAG	3.98	712*	AACAGGTAG
Match cutoff			3.06	Match cutoff			3.63		

Table S4. Snail (Sna)-binding motifs in the *sog* shadow enhancer were identified by ClusterDraw (1) fed with PFMs constructed from two independent binding sites alignments. One was built by site sequences from *in vitro* binding assays such as DNase footprinting and SELEX (http://line.bioinfolab.net/webgate/help.htm#mtfform) (1) and the other from an *in vivo* B1H system (http://mccb.umassmed.edu/ffs/) (2). Sna-binding motifs that met the following two selection criteria were finally selected as putative functional motifs that Sna binds: 1) Sna-binding motifs commonly identified by two independent searches performed with two different PFMs, 2) Sna-binding motifs whose cumulative match probability (-log*P*) values were higher than the match cutoff value. The binding motifs whose match probabilities (-log*P*) were higher than the match cutoff value are underlined. Sna-binding motifs overlapping Zld-binding motifs are labeled with asterisks. Even though only two Sna motifs met the two criteria, four Sna-binding motifs were finally selected as putative functional sites. Two motifs (5'-CGCAGGTAG-

3’ and 5’-CACATGTGG-3’) starting at the 517th and 676th nucleotides of the *sog* shadow enhancer were additionally selected for two reasons. First, they were commonly identified by two independent searches with different PFMs, although their match probability values were lower than the match cutoff. It seems apparent that the nucleotide sequences outside the Sna core motifs (5’-CANNTK-3’, N stands for any nucleotide and K stands for G or T nucleotides, respectively.) (2, 4) deviated from the Sna consensus sequence, thereby lowering their match probability scores. Second, the core sequences (5’-CAGGTA-3’ and 5’-CATGTG-3’) of the two Sna motifs were exactly same as those of the Sna motifs starting at the 518th and 677th nucleotides whose match probability values were higher than match cutoff, suggesting that the additionally selected Sna motifs starting at the 517th and 676th nucleotides may also be functionally significant.

Table S5

Bicoid binding motifs in the *sog* shadow 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
71	F	CGTAATCCT	3.89	72	F	GTAATCC	3.65	72	GTAATCC
								163	AAAATCA
236	F	AATAACCCG	2.75					237	ATAACCC
390	R	CCCAATCCC	2.90	391	R	CCAATCC	2.58	391	CCAATCC
396	R	GCTAGTCCC	2.80						
471	F	TATGATCCC	2.57						
Match cutoff			3.06	Match cutoff			3.63		

Table S5. Bicoid (Bcd)-binding motifs in the *sog* shadow enhancer were identified by ClusterDraw (1) fed with PFMs constructed from two independent binding sites alignments. One was built by site sequences from *in vitro* binding assays such as DNase footprinting and SELEX (http://line.bioinfolab.net/webgate/help.htm#mtfform) (1) and the other from an *in vivo* B1H system (http://mccb.umassmed.edu/ffs/) (2). Binding motifs whose match probabilities (-log*P*) were higher than the match cutoff value are underlined. Four Bcd-binding motifs were selected as putative functional motifs with some arbitrary determinations. Although the two motifs starting at the 396th and 471st nucleotides showed poor matches to the Bcd-binding consensus sequence (Fig. S1 and S2), the ClusterDraw search performed with the *in vitro* PFM picked them up as functional Bcd-binding motifs. In addition, although two motifs starting at the 163rd and 237th nucleotides show good matches to the Bcd consensus sequence (Fig. S1 and S2), searches with the *in vivo* PFM did not select them as functional motifs for unknown reasons. Thus, the former two motifs (starting at the 396th and 471st nucleotides) were excluded from and the latter two motifs (starting at the 163rd and 237th nucleotides) included in the list of functional Bcd motifs.

Table S6

Dorsal binding motifs in the *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 S6. Dl-binding motifs in the *sog* primary enhancer were identified by ClusterDraw (1) fed with PFMs constructed from two independent binding sites alignments. One was built by site sequences from *in vitro* binding assays such as DNase footprinting and SELEX (http://line.bioinfolab.net/webgate/help.htm#mtfform) (1) and the other from an *in vivo* B1H system (http://mccb.umassmed.edu/ffs/) (2). Dl-binding motifs that met two selection criteria were finally selected as putative functional motifs that Dl binds: 1) Dl-binding motifs that were commonly identified by two independent searches performed with two different PFMs, 2) Dl-binding motifs whose cumulative match probability (-log*P*) values were higher than the match cutoff value. The 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 putatively functional Dl-binding sites.

Table S7

Bicoid binding motifs in the *sog* primary enhancer.									
Dimitri's Collection (Footprint)				FlyFactorSurvey Collection (B1H)				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 S7. Bcd-binding motifs in the *sog* primary enhancer were identified by ClusterDraw (1) fed with PFMs constructed from two independent binding sites alignments. One was built by site sequences from *in vitro* binding assays such as DNase footprinting and SELEX (http://line.bioinfolab.net/webgate/help.htm#mtfform) (1) and the other from an *in vivo* B1H system (http://mccb.umassmed.edu/ffs/) (2). Bcd-binding motifs that met two selection criteria were finally selected as putative functional motifs that Bcd binds: 1) Bcd-binding motifs commonly identified by two independent searches performed with two different PFMs, 2) Bcd-binding motifs whose cumulative match probability (-log*P*) values were 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 S8.

Enhancer DNA sequence (5’ to 3’ direction).
>sog_primary_392bp GTTGCCAATGCCATTGCGCATACGCCGTGTCGTCTATATGGCTATATGGCTATATGGCTGTATGGTGCGGGGAAATC CCCGTAATCGCAGGTAGAATTCCAGCCGGTGCCGAGGCGGGACCTGCTCGCACCTCTAATCCCGCCAGGGTTTTCGG GACATGGGATATTCCCGACGGCACAGCATAGCACTCCGTTTTCTTTTTTTTTTTTTATTATTATTGTGTCCAGTTTT AATCCGGAAAGCGGGAATTCCCTTCCGCTCGCTGCCTGCACTGCGCTGCGCAGACGCATCGGCGTCCGTAAGCCGCT TACCAAAAAGATACGGGTATACCCAAATGGATGCCTGCCCATGTATATAGACCATTGGGTGGTATGGACCATGGACC ATAAAGC
>sog_shadow_884bp GAGAAGGAGGAGAAGTTGGTTGAGAGGTCATCGTTGCGATTCTGCGATTCAGCAGTTCCACAGAAGGTGTCGTAATC CTGGACGCAAGGGTGCACGGACCAACTGACAGGGGCAAGTGCGTCCTGTGCCACCAGATGACGCACGATGCGGCCGG AAAAACCCAAAATCAAAAACCGAAAACCGAAAACCTGGTCAGAGTTTCCGAAAACCAAAGAGCCAACATCGAATGCG GCACAATAACCCGATTGTCTGCGAATACCCACGATGATCTAGAATCGCACGGAGAGCACTCTCACGCATCCGTGGCC ATATGGGTGCGGCCAAATCGGAAATTCCCAGGACAGGTAGAATGCATTGGATATACGGGTATACGGATTGGAATTGG GATTGGGATTGGGACTAGCACCAGGTTGCAACGCCCGCCAAGAAGCCAATTTAAATAAGCAGCATAAACAAAAGCGA CAGCGTTTTATGATCCCCGCTCCTTATCCTTGCACAAGGATATCGCCATGGCCACGCAGGTAGGAATAGCAGATATG GCGGCAATGATGCGCCAACCGCACTGCTTCGTCCTGGTCCTGGTCGGATGGGCTTTTCCCACGCAACCGCGACCTTA TCTGCGCCCCTTTTATGAGGCTGCATCTGTTTTCGCACCTCGATGCCGTTGGCATTATAGCCACATGTGTATGGTGG GAATTTCCGATCGACCAGCCTACCTGTTCCGCTGAAACCCGGGAATCTGTCCATCCTGAGCTTCCACACACACACAC ACACACACACAGGTCAGTCGGCATCAATTGGCTGCCATAAACATATAACAATCAATATTGAATCCTTTATCGTAGAA TTTGTTGTATATGCCCATTGCAGTCCTTCGATTAAAT

Table S8. Primary and shadow enhancer sequences that direct *sog* expression in the neurogenic ectoderm of *Drosophila* early embryos. Primary enhancer for *sog* expression was initially found by a computational algorithm to search the *Drosophila* genome for clusters that contain three or more optimal Dl-binding sites (5). ChIP-chip assays predicted that many of the Dl target genes contain 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 away from the transcription start sites of Dl target genes. The *sog* shadow enhancer was predicted to be one of those secondary enhancers and its transcriptional activity was tested by P element-mediated germline transformation followed by whole-mount *in situ* hybridization in *Drosophila* early embryo (6). 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. The primer

sequences used for genomic PCR amplification are underlined.

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

>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

>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

>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
```

Figure S1. PFMs of Dl-, Zld-, Sna-, and Bcd-binding DNA motifs obtained from *in vitro* binding assays such as DNase footprinting and SELEX (1). The Zld PFM was obtained from motif alignment of chromatin immunoprecipitation-DNA tiling assay (ChIP-chip) analyses performed with anti-Zld antibody (3), because Zld-binding motif alignments obtained from *in vitro* binding assays are not currently available. Those PFMs were used to calculate position weight matrices (PWMs) for each motif found in the *sog* shadow enhancer.

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

>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

>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

>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
```

Figure S2. PFMs of Dl-, Zld-, Sna-, and Bcd-binding DNA motifs obtained from an *in vivo* B1H system in FlyFactorSurvey database (2). Those PFMs were used to calculate position weight matrices (PWMs) for each motif found in the *sog* shadow enhancer.

Figure S3

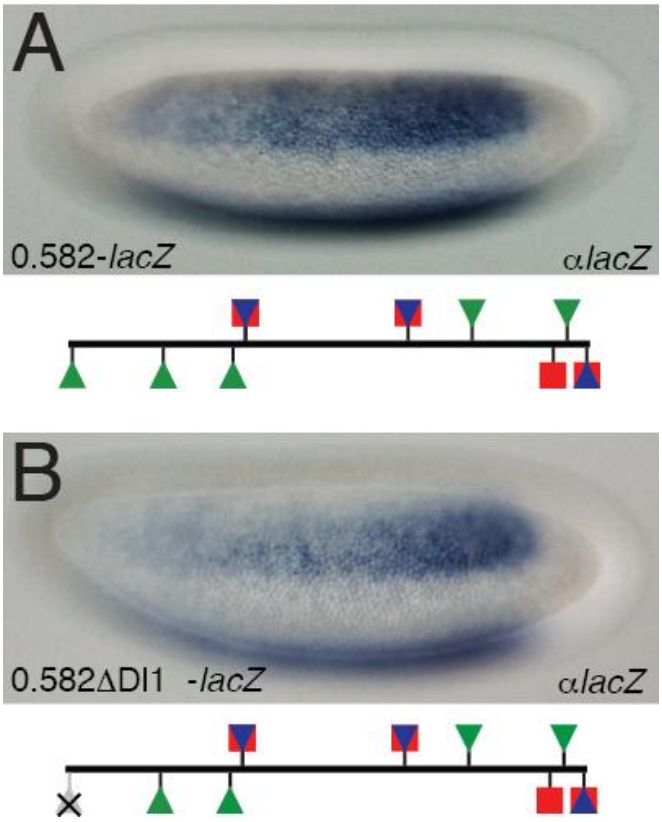

Figure S3. A mutation in either D1 Dl- or B2 Bcd-binding sites in the minimal enhancer resulted in similar lacZ expression patterns. (A) The 0.582-kb wild-type sog shadow minimal enhancer directed broad lacZ expression in the neurogenic ectoderm of the cellularizing blastoderm. (B) A mutant version of the minimal enhancer containing a mutant Dl-binding site (Dl1) resulted in reduced lacZ expression in the anterior half of the embryo; this pattern of expression is similar to that directed by the enhancer containing a mutant Bcd-binding site (B2).

Supplementary References

1. Papatsenko, D. (2007) ClusterDraw web server: a tool to identify and visualize clusters of binding motifs for transcription factors. *Bioinformatics* 23, 1032-1034.
2. 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 research* 39, D111-D117.
3. 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 Genetics* 7, e1002339.
4. Mauhin, V., Lutz, Y., Dennefeld, C. and Alberga, A. (1993) Definition of the DNA-binding site repertoire for the Drosophila transcription factor SNAIL. *Nucleic acids research* 21, 3951-3957.
5. 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.
6. Hong, J.W., Hendrix, D.A. and Levine, M.S. (2008) Shadow enhancers as a source of evolutionary novelty. *Science* 321, 1314.