Evolution of Insect Dorsoventral Patterning Mechanisms

M.W. Perry,[1] J.D. Cande,[2,4] A.N. Boettiger,[3] and M. Levine[2]

[1]*Department of Integrative Biology, University of California, Berkeley, California 94720-3140;*
[2]*Department of Molecular and Cell Biology, Division of Genetics, Genomics and Development, Center for Integrative Genomics, University of California, Berkeley, California 94720-3200;*
[3]*Biophysics Program, University of California, Berkeley, California 94720-3200*

Correspondence: mlevine@berkeley.edu

The dorsoventral (DV) patterning of the early *Drosophila* embryo depends on Dorsal, a maternal sequence-specific transcription factor related to mammalian NF-κB. Dorsal controls DV patterning through the differential regulation of ~50 target genes in a concentration-dependent manner. Whole-genome methods, including ChIP-chip and ChIP-seq assays, have identified ~100 Dorsal target enhancers, and more than one-third of these have been experimentally confirmed via transgenic embryo assays. Despite differences in DV patterning among divergent insects, a number of the Dorsal target enhancers are located in conserved positions relative to the associated transcription units. Thus, the evolution of novel patterns of gene expression might depend on the modification of old enhancers, rather than the invention of new ones. As many as half of all Dorsal target genes appear to contain "shadow" enhancers: a second enhancer that directs the same or similar expression pattern as the primary enhancer. Preliminary studies suggest that shadow enhancers might help to ensure resilience of gene expression in response to environmental and genetic perturbations. Finally, most Dorsal target genes appear to contain RNA polymerase II (pol II) prior to their activation. Stalled pol II fosters synchronous patterns of gene activation in the early embryo. In contrast, DV patterning genes lacking stalled pol II are initially activated in an erratic or stochastic fashion. It is possible that stalled pol II confers fitness to a population by ensuring coordinate deployment of the gene networks controlling embryogenesis.

DV patterning of the *Drosophila* embryo is controlled by Dorsal, a sequence-specific transcription factor related to mammalian nuclear factor κB (NF-κB) (Roth et al. 1989; Rushlow et al. 1989; Ip et al. 1991). The Dorsal protein is distributed in a broad nuclear gradient, with peak levels present in ventral nuclei and progressively lower levels in lateral and dorsal regions (Roth et al. 1989; Rushlow et al. 1989; Steward 1989). This Dorsal nuclear gradient initiates DV patterning by regulating 50–60 target genes in a concentration-dependent fashion (Stathopoulous et al. 2002; Zeitlinger et al. 2007a).

Whole-genome chromatin immunoprecipitation (ChIP)-chip assays (see below) identified ~100 potential Dorsal target enhancers, and more than 30 of these have been directly tested in transgenic embryos (see, e.g., Zeitlinger et al. 2007a; Hong et al. 2008a). Altogether, these enhancers direct six distinct patterns of gene expression across the DV axis of precellular embryos. Dorsal works in a highly combinatorial manner to generate these diverse patterns (for review, see Hong et al. 2008b). For example, Dorsal and SuH, a transcriptional effector of Notch signaling, activate *single-minded* (*sim*) expression in a single line of cells (central nervous system [CNS] ventral midline) on either side of the mesoderm (Cowden and Levine 2002; Morel et al. 2003). In contrast, Dorsal works together with a different sequence-specific transcription factor, Pointed (an effector of epidermal growth factor [EGF] signaling), to activate gene expression within lateral stripes in intermediate regions of the future ventral nerve cord (Gabay et al. 1996).

ENHANCER EVOLUTION

In principle, substitutions of "coactivator" binding sites within Dorsal target enhancers can alter the DV limits of gene expression. For example, replacing SuH-binding sites with Twist sites results in expanded expression of the modified enhancer within the presumptive neurogenic ectoderm (Gray and Levine 1996; Zinzen et al. 2006). Analysis of Dorsal target enhancers in divergent insects, including mosquitoes (*Anopheles gambiae*), flour beetles (*Tribolium castaneum*), and honeybees (*Apis mellifera*), suggests that such changes might occur during evolution to produce distinctive DV patterning mechanisms (Zinzen et al. 2006).

One such example is seen for the ventral midline of *A. mellifera*. In *Drosophila*, the ventral midline is just two cells in width and arises from two lines of *sim*-expressing cells that straddle the mesoderm before gastrulation (Fig. 1). In contrast, the ventral midline of the *A. mellifera* CNS is considerably wider, encompassing about five to six cells. An expanded ventral midline is also seen in *T. castaneum*, suggesting that the broad pattern is ancestral, and the narrow midline of *Drosophila* (and *A. gambiae*) is a derived feature of the dipteran CNS (Zinzen et al. 2006).

Expansion of the *sim* expression pattern is sufficient to account for the broad ventral midlines of the *A. mellifera* and *T. castaneum* CNS. In *Drosophila*, ectopic activation of *sim* expression using the *eve* stripe-2 enhancer results in

[4]Present address: Developmental Biology Institute of Marseilles (IBDML), Marseille, France.

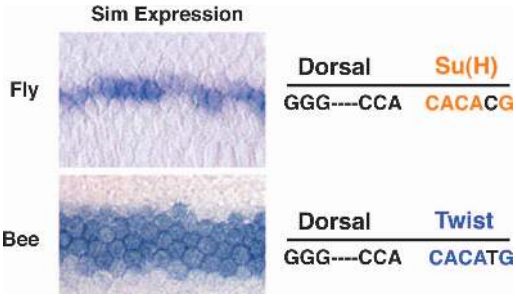

Figure 1. *sim* exhibits a broader pattern of expression in the honeybee CNS as compared with *Drosophila*. This expansion appears to result from the replacement of Suppressor of Hairless sites (Notch signaling) with Twist sites in the respective 5′ *sim* enhancers.

the formation of an ectopic ventral midline throughout the neurogenic ectoderm of transgenic embryos (Zinzen et al. 2006). The *sim* regulatory region contains two distinct enhancers: One mediates activation by Dorsal and Notch signaling (establishment enhancer), and the other mediates positive autofeedback through direct binding of the Sim transcription factor to the autoregulatory enhancer (Kasai et al. 1992). Once Sim is misexpressed, the expanded pattern is maintained by autofeedback.

Sim establishment enhancers were identified in the 5′-flanking regions of the *sim* loci in *A. gambiae*, *T. castaneum*, and *A. mellifera*. The *sim* enhancer from *A. gambiae* directs sharp lateral lines when expressed in transgenic *Drosophila* embryos. In contrast, the enhancers obtained from the *sim* loci of *T. castaneum* and *A. mellifera* produce broader expression patterns. The *A. gambiae* enhancer resembles the *Drosophila* enhancer in that it contains a series of Dorsal- and SuH-binding sites. However, the *T. castaneum* and *A. mellifera* enhancers contain Twist sites rather than SuH sites, and consequently, they direct broader patterns of gene expression (Zinzen et al. 2006; Cande 2009).

CONSTANCY OF ENHANCER LOCATION

The *sim* enhancers of flies, mosquitoes, flour beetles, and bees lack simple sequence similarity. Despite this extensive sequence divergence, comparable enhancers are located in the same relative positions: in the immediate 5′-flanking regions of the respective *sim* loci (e.g., Fig. 2).

Because this is a relatively common location for developmental enhancers, additional studies were done to determine whether enhancer locations are conserved for other critical DV patterning genes (Cande et al. 2009). These studies identified enhancers for five additional genes: *cactus*, *sog*, *twist*, *brinker*, and *vnd*. *cactus* is a key component of the Toll signaling pathway that regulates Dorsal nuclear transport (Roth et al. 1991; Stein and

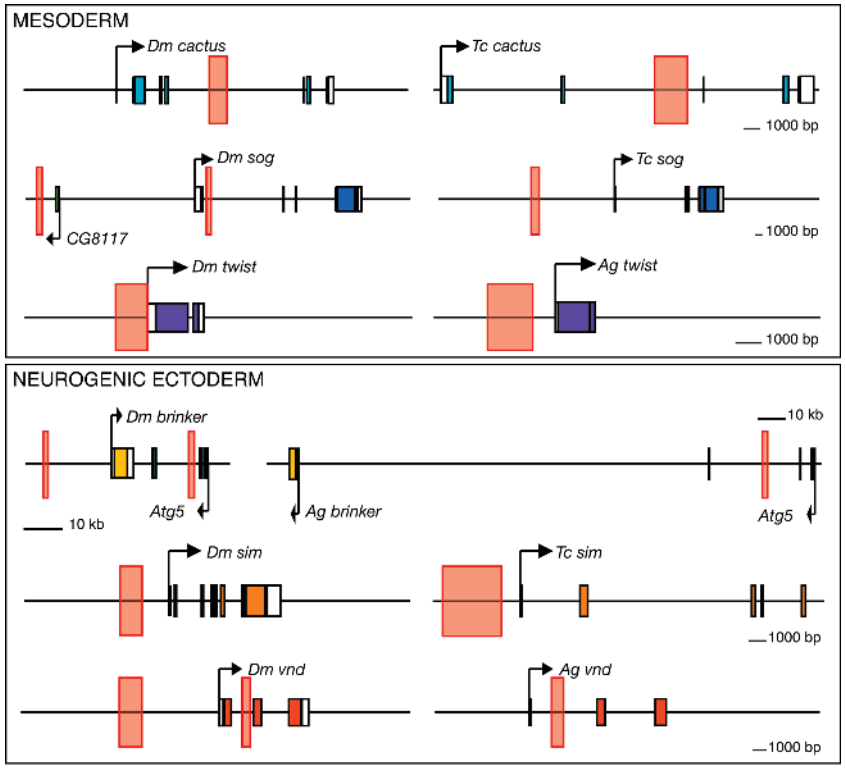

Figure 2. Conservation of enhancer location in divergent insects. (Pink boxes) Enhancers regulating the associated transcription units, (colored rectangles) coding exons. Note the conservation of a *brinker* enhancer within the intron of the neighboring *Atg5* loci of flies and mosquitoes. (Ag) *Anopheles gambiae*, (Dm) *Drosophila melanogaster*, (Tc) *Tribolium castaneum*, (*sim*) *single minded*, (*sog*) *short gastrulation*, (*vnd*) *ventral nervous system defective*. (Reprinted, with permission, from Cande et al. 2009 [© National Academy of Sciences].)

Nüsslein-Volhard 1992). It is activated by high levels of the Dorsal gradient in the presumptive mesoderm of both *Drosophila* and *T. castaneum* embryos (Maxton-Kuchenmeister et al. 1999; Nunes da Fonseca et al. 2008). The enhancers that are responsible for these expression patterns are located in 3′ introns of the respective *cactus* transcription units (Cande et al. 2009).

Enhancer conservation at the *brinker* (*brk*) locus is even more dramatic. *brk* encodes a sequence-specific transcriptional repressor that helps to restrict Dpp (bone morphogenetic protein [BMP]) signaling to the dorsal ectoderm (Jaźwińska et al. 1999). In *Drosophila*, two separate enhancers regulate *brk* expression in the presumptive neurogenic ectoderm of pregastrular embryos (Hong et al. 2008a). One of the enhancers is located ~10 kb 5′ of the *brk* transcription start site. The other is located 13 kb downstream from the start site, within the intron of a neighboring gene, *Atg5*. The major enhancer regulating *brk* expression in the *A. gambiae* embryo is located within the *Atg5* gene, even though the *brk* transcription unit is inverted relative to its orientation in *Drosophila* and *Atg5* is located quite far, ~100 kb, from *brk* in the mosquito genome (Fig. 2) (Cande et al. 2009).

Binding-site turnover has been well documented in insect enhancers (Moses et al. 2006; for review, see Ludwig 2002). Despite this turnover within existing enhancers, there might be constraints on the de novo evolution of developmental enhancers. We suggest that the evolution of novel patterns of gene expression depends primarily on the modification of ancestral enhancers, rather than the invention of new ones.

SHADOW ENHANCERS

ChIP-chip assays led to the comprehensive identification of Dorsal target enhancers in the *Drosophila* genome (Zeitlinger et al. 2007a). These studies identified multiple enhancers at more than one-third of the target genes that are directly regulated by the Dorsal gradient. For example, the *vnd* gene encodes a sequence-specific transcription factor that specifies the ventral-most neuronal cell identities of the ventral nerve cord (see, e.g., Weiss et al. 1998). It is activated by enhancers located in both the 5′-flanking region and within the first intron of the transcription unit (Shao et al. 2002; Stathopoulous et al. 2002; Zeitlinger et al. 2007a). Similarly, *sog* is regulated by both a 5′ enhancer and an intronic enhancer (Fig. 3), and as discussed above, *brk* is activated by enhancers located in both 5′- and 3′-flanking regions (Zeitlinger et al. 2007a, Hong et al. 2008a).

We refer to the secondary enhancers located in remote 5′ or 3′ positions as shadow enhancers (Hong et al. 2008a). Preliminary studies suggest that they might help to confer resilience in gene expression in response to genetic and environmental perturbations. For example, *vnd* and *sog* exhibit normal patterns of transcriptional activation in embryos derived from dl/+ heterozygotes (half of the normal dose of the Dorsal gradient), whereas *Neu3* and *rho* display erratic patterns of activation (Fig. 4) (Boettiger and Levine 2009). *vnd* and *sog* contain shadow enhancers, whereas *Neu3* and *rho* do not. It is possible that dual enhancers for a common expression pattern ensure accurate and reproducible activation in large populations of embryos subject to environmental fluctuations.

It is possible that shadow enhancers arise from "cryptic" duplication events. Of course, other scenarios can be envisioned, but regardless of mechanism, once they arise, shadow enhancers might confer an adaptive advantage to a population by ensuring accurate activation of critical developmental control genes. Shadow enhancers offer an opportunity for producing novel patterns of gene expression without disrupting the core function of the primary enhancer and associated gene. According to this view, the evolution of shadow enhancers might come at a cost to the fitness of a population, but this cost could be compensated by the advantages conferred by the novel mode of gene expression.

TRANSCRIPTIONAL SYNCHRONY

Recent studies with mammalian progenitor cells, including stem cells, suggest that many critical developmental control genes (e.g., Hox genes) are repressed but poised for

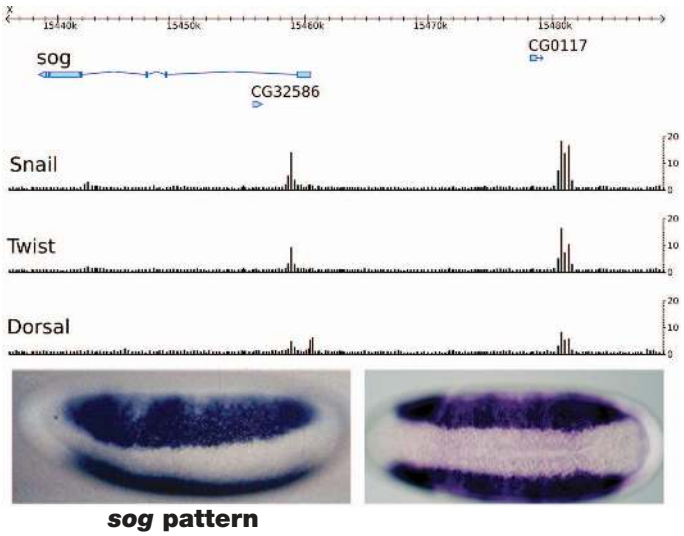

Figure 3. ChIP-chip assays identified two enhancers for the early *sog* expression pattern. (Light yellow) *sog* transcription unit. The locations of Dorsal-, Twist-, and Snail-binding sites are indicated below. There are two clusters of binding sites: in the first intron and more than 20 kb 5′ of the start site. The intronic cluster was previously shown to function as an enhancer for the *sog* expression pattern (*left*, embryo stained to show the endogenous *sog* expression pattern). The distal cluster generates a similar pattern of expression when attached to a *lacZ* reporter gene and expressed in transgenic embryos (*right*). (Modified, with permission, from Hong et al. 2008a [© AAAS].)

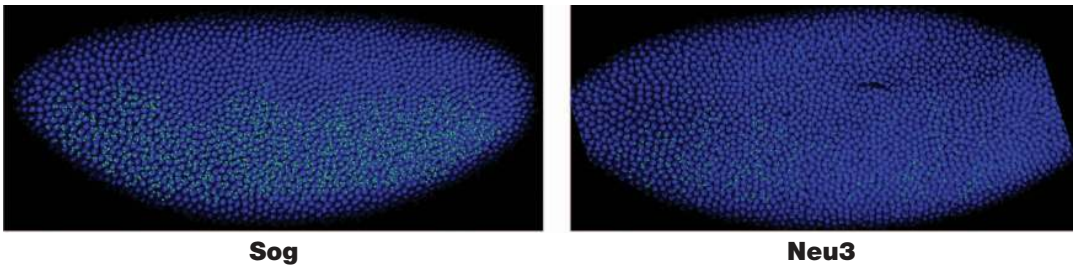

Figure 4. Onset of *sog* and *Neu3* expression in precellular embryos at the early phases of nuclear cleavage cycle 14. The embryos were collected from *dorsal*/+ females and therefore contain half of the normal levels of the Dorsal nuclear gradient. The *sog* pattern is normal, but *Neu3* displays erractic activation. *sog* contains a shadow enhancer, whereas *Neu3* does not. (Modified, with permission, from Boettiger and Levine 2009 [© AAAS].)

rapid induction (Guenther et al. 2007). Many such genes contain bivalent histone marks, H3K4 trimethylation and H3K27 methylation, which are indicative of genes that are active and repressed, respectively (Bernstein et al. 2006).

ChIP-chip assays were done in *Drosophila* using a mixture of antibodies directed against pol II (Muse et al. 2007; Zeitlinger et al. 2007b). These studies suggest that most DV patterning genes contain stalled pol II at the core promoter before their activation in response to the Dorsal gradient. Like the bivalent marks seen in mammalian progenitor cells, stalled pol II is likely to render the associated genes repressed but poised for rapid activation.

Classical studies on *Drosophila* heat shock genes have documented that stalled, or paused, pol II accelerates their activation in response to stress as compared with comparable promoters lacking paused pol II (Lis and Wu 1993; Conaway et al. 2000; Saunders et al. 2006). This paradigm of gene expression was seen as a specialized stress response. However, the finding that many developmental control genes contain stalled pol II in the early *Drosophila* embryo raises the possibility that the control of transcription elongation is an important strategy for differential gene regulation during development (Lis 2007; Zeitlinger et al. 2007b; Hendrix et al. 2008; Chopra et al. 2009).

It is possible that stalled pol II suppresses transcriptional noise during development. In principle, a major source of such noise is variability in pol II recruitment and promoter escape (Raser and O'Shea 2004, 2005; Raj et al. 2006; Darzacq et al. 2007; Raj and van Oudenaarden 2008). For example, just a fraction of the pol II that interacts with a promoter succeeds in melting the double-stranded DNA and launching transcription. In contrast, stalled pol II has already passed the "checkpoint" (promoter escape) and is more likely to succeed in transcribing the associated gene as compared with a naïve pol II complex that is newly recruited to the DNA template.

These considerations raise the possibility that genes containing stalled pol II might be activated in a synchronous fashion due to diminished nucleus-to-nucleus variation in de novo transcription upon induction. To test this possibility, a quantitative in situ hybridization method was developed to identify the initial de novo transcripts in all of the nuclei of a large number of embryos (Boettiger and Levine 2009). These studies suggest that genes containing stalled pol II are activated in a coordinated fashion throughout the field of nuclei where the gene is expressed (Fig. 5). In contrast, genes lacking stalled pol II are activated in an erratic fashion, whereby nuclei displaying de novo transcripts are surrounded by those lacking expression (Boettiger and Levine 2009).

It is possible that transcriptional synchrony is a manifestation of metazoan development, whereby groups of cells function in a highly coordinated fashion. Stalled pol II and transcriptional synchrony might help to foster such coordinate behavior. We propose that stalled pol II contributes to population fitness, in that it helps to ensure the accurate and reproducible regulation of key developmental control genes.

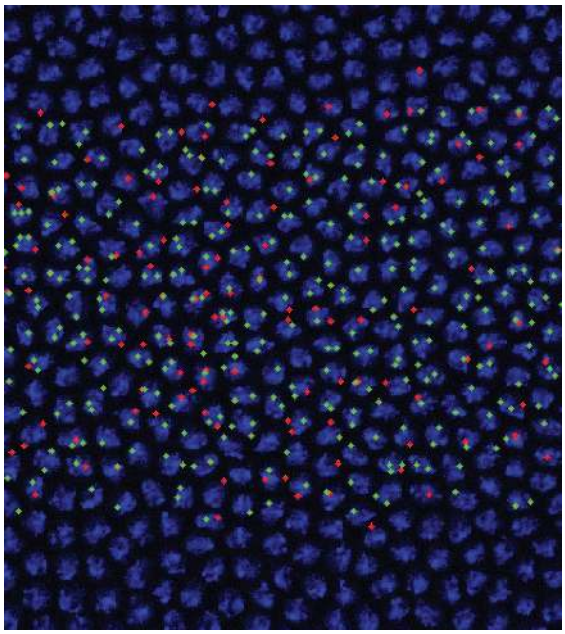

Figure 5. High-magnification view of the presumptive mesoderm of a precellular embryo at the early phase of nuclear cleavage cycle 14. Intronic probes were used to visualize nascent transcripts from the *Mes2* (green) and *Mes4* (red) genes. *Mes2* displays expression in most of the mesodermal nuclei, whereas *Mes4* is expressed in less than half of the nuclei. *Mes2* contains stalled pol II, whereas *Mes4* does not. (Modified, with permission, from Boettiger and Levine 2009 [© AAAS].)

ACKNOWLEDGMENTS

This work was supported in part by the National Institutes of Health grants GM46638 and GM34431. A.N.B. and M.W.P. are supported by National Science Foundation graduate research fellowships and J.D.C. is supported by a Chang Ling Tien graduate fellowship.

REFERENCES

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125:** 315–326.

Boettiger AN, Levine M. 2009. Synchronous and stochastic patterns of gene activation in the *Drosophila* embryo. *Science* **325:** 471–473.

Cande J, Goltsev Y, Levine M. 2009. Conservation of enhancer positions in divergent insects. *Proc Natl Acad Sci* **106:** 14414–14419.

Chopra VS, Hong JW, Levine M. 2009. Regulation of Hox gene activity by transcriptional elongation in *Drosophila*. *Curr Biol* **19:** 688–693.

Conaway JW, Shilatifard A, Dvir A, Conaway RC. 2000. Control of elongation by RNA polymerase II. *Trends Biochem Sci* **25:** 375–380.

Cowden J, Levine M. 2002. The Snail repressor positions Notch signaling in the *Drosophila* embryo. *Development* **129:** 1785–1793.

Darzacq X, Shav-Tal Y, de Turris V, Brody Y, Shenoy SM, Phair RD, Singer RH. 2007. In vivo dynamics of RNA polymerase II transcription. *Nat Struct Mol Biol* **14:** 796–806.

Gabay L, Scholz H, Golembo M, Klaes A, Shilo BZ, Klambt C. 1996. EGF receptor signaling induces pointed P1 transcription and inactivates Yan protein in the *Drosophila* embryonic ventral ectoderm. *Development* **122:** 3355–3362.

Gray S, Levine M. 1996. Short-range transcriptional repressors mediate both quenching and direct repression within complex loci in *Drosophila*. *Genes Dev* **10:** 700–710.

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. 2007. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130:** 77–88.

Hendrix DA, Hong JW, Zeitlinger J, Rokhsar DS, Levine MS. 2008. Promoter elements associated with RNA Pol II stalling in the *Drosophila* embryo. *Proc Natl Acad Sci* **105:** 7762–7767.

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* **105:** 20072–20076.

Ip YT, Kraut R, Levine M, Rushlow CA. 1991. The dorsal morphogen is a sequence-specific DNA-binding protein that interacts with a long-range repression element in *Drosophila*. *Cell* **64:** 439–446.

Jaźwińska A, Kirov N, Wieschaus E, Roth S, Rushlow C. 1999. The *Drosophila* gene *brinker* reveals a novel mechanism of Dpp target gene regulation. *Cell* **96:** 563–573.

Kasai Y, Nambu JR, Lieberman PM, Crews ST. 1992. Dorsal-ventral patterning in *Drosophila*: DNA binding of snail protein to the *single-minded* gene. *Proc Natl Acad Sci* **89:** 3414–3418.

Lis JT. 2007. Imaging *Drosophila* gene activation and polymerase pausing in vivo. *Nature* **450:** 198–202.

Lis J, Wu C. 1993. Protein traffic on the heat shock promoter: Parking, stalling, and trucking along. *Cell* **74:** 1–4.

Ludwig MZ. 2002. Functional evolution of noncoding DNA. *Curr Opin Genet Dev* **12:** 634–639.

Maxton-Küchenmeister J, Handel K, Schmidt-Ott U, Roth S, Jäckle H. 1999. Toll homologue expression in the beetle *Tribolium* suggests a different mode of dorsoventral patterning than in *Drosophila* embryos. *Mech Dev* **83:** 107–114.

Morel V, Le Borgne R, Schweisguth F. 2003. Snail is required for Delta endocytosis and Notch-dependent activation of *single-minded* expression. *Dev Genes Evol* **213:** 65–72.

Moses AM, Pollard DA, Nix DA, Iyer VN, Li XY, Biggin MD, Eisen MB. 2006. Large-scale turnover of functional transcription factor binding sites in *Drosophila*. *PLoS Comput Biol* **2:** e130.

Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, Zeitlinger J, Adelman K. 2007. RNA polymerase is poised for activation across the genome. *Nat Genet* **39:** 1507–1511.

Nunes da Fonseca R, von Levetzow C, Kalscheuer P, Basal A, van der Zee M, Roth S. 2008. Self-regulatory circuits in dorsoventral axis formation of the short-germ beetle *Tribolium castaneum*. *Dev Cell* **14:** 605–615.

Raj A, van Oudenaarden A. 2008. Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell* **135:** 216–226.

Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. 2006. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* **4:** e309.

Raser JM, O'Shea EK. 2004. Control of stochasticity in eukaryotic gene expression. *Science* **304:** 1811–1814.

Raser JM, O'Shea EK. 2005. Noise in gene expression: Origins, consequences, and control. *Science* **309:** 2010–2013.

Roth S, Stein D, Nüsslein-Volhard C. 1989. A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the *Drosophila* embryo. *Cell* **59:** 1189–1202.

Roth S, Hiromi Y, Godt D, Nüsslein-Volhard C. 1991. *cactus*, a maternal gene required for proper formation of the dorsoventral morphogen gradient in *Drosophila* embryos. *Development* **112:** 371–388.

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.

Saunders A, Core LJ, Lis JT. 2006. Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol* **7:** 557–567.

Shao X, Koizumi K, Nosworthy N, Tan DP, Odenwald W, Nirenberg M. 2002. Regulatory DNA required for vnd/NK-2 homeobox gene expression pattern in neuroblasts. *Proc Natl Acad Sci* **99:** 113–137.

Stathopoulos A, Van Drenth M, Erives A, Markstein M, Levine M. 2002. Whole-genome analysis of dorsal-ventral patterning in the *Drosophila* embryo. *Cell* **111:** 687–701.

Stein D, Nüsslein-Volhard C. 1992. Multiple extracellular activities in *Drosophila* egg perivitelline fluid are required for establishment of embryonic dorsal-ventral polarity. *Cell* **68:** 429–440.

Steward R. 1989. Relocalization of the dorsal protein from the cytoplasm to the nucleus correlates with its function. *Cell* **59:** 1179–1188.

Weiss JB, Von Ohlen T, Mellerick DM, Dressler G, Doe CQ, Scott MP. 1998. Dorsoventral patterning in the *Drosophila* central nervous system: The intermediate neuroblasts defective homeobox gene specifies intermediate column identity. *Genes Dev* **12:** 3591–3602.

Zeitlinger J, Zinzen RP, Stark A, Kellis M, Zhang H, Young RA, Levine M. 2007a. 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.

Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, Adelman K, Levine M, Young RA. 2007b. RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat Genet* **39:** 1512–1516.

Zinzen R, Cande JD, Papatsenko D, Levine M. 2006. Evolution of the ventral midline in insect embryos. *Dev Cell* **11:** 895–902.