Developmental Biology 360 (2011) 230–240

Contents lists available at SciVerse ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/developmentalbiology

Genomes and Developmental Control

# Gene length may contribute to graded transcriptional responses in the *Drosophila* embryo

Peter McHale [a,1,2], Claudia M. Mizutani [b,c,d,1,3], David Kosman [b], Danielle L. MacKay [c], Mirela Belu [c], Anita Hermann [b], William McGinnis [b], Ethan Bier [b,*], Terence Hwa [a,*]

[a] Center for Theoretical Biological Physics & Physics Department, University of California San Diego, La Jolla, CA 92037, USA
[b] Section of Cell and Developmental Biology, University of California San Diego, La Jolla, CA 92037, USA
[c] Department of Biology, Case Western Reserve University, Cleveland, OH 44106, USA
[d] Department of Genetics, Case Western Reserve University, Cleveland, OH 44106, USA

ARTICLE INFO

*Article history:*
Received for publication 4 August 2011
Accepted 28 August 2011
Available online 3 September 2011

*Keywords:*
Morphogen gradient
Transcriptional delay
Computational modeling
Prepattern
Multiplex in situ hybridization
*Drosophila*

ABSTRACT

An important question in developmental biology is how relatively shallow gradients of morphogens can reliably establish a series of distinct transcriptional readouts. Current models emphasize interactions between transcription factors binding in distinct modes to *cis*-acting sequences of target genes. Another recent idea is that the *cis*-acting interactions may amplify preexisting biases or prepatterns to establish robust transcriptional responses. In this study, we examine the possible contribution of one such source of prepattern, namely gene length. We developed quantitative imaging tools to measure gene expression levels for several loci at a time on a single-cell basis and applied these quantitative imaging tools to dissect the establishment of a gene expression border separating the mesoderm and neuroectoderm in the early *Drosophila* embryo. We first characterized the formation of a transient ventral-to-dorsal gradient of the Snail (Sna) repressor and then examined the relationship between this gradient and repression of neural target genes in the mesoderm. We found that neural genes are repressed in a nested pattern within a zone of the mesoderm abutting the neuroectoderm, where Sna levels are graded. While several factors may contribute to the transient graded response to the Sna gradient, our analysis suggests that gene length may play an important, albeit transient, role in establishing these distinct transcriptional responses. One prediction of the gene-length-dependent transcriptional patterning model is that the co-regulated genes *knirps* (a short gene) and *knirps-related* (a long gene) should be transiently expressed in domains of differing widths, which we confirmed experimentally. These findings suggest that gene length may contribute to establishing graded responses to morphogen gradients by providing transient prepatterns that are subsequently amplified and stabilized by traditional *cis*-regulatory interactions.



## Introduction

Pattern formation along the dorsal–ventral axis of the *Drosophila* embryo is initiated by a nuclear concentration gradient of the NFκ-B-related transcription factor Dorsal (Roth et al., 1989; Rushlow et al., 1989; Steward, 1989), wherein high levels of Dorsal in ventral regions specify expression of the key mesodermal determinants Twist (Twi) and Snail (Sna) (Ip et al., 1992b; Jiang et al., 1991; Kosman et al., 1991; Leptin, 1991; Rao et al., 1991; Ray et al., 1991; Thisse et al., 1991), while lower levels in lateral regions activate expression of neuroectodermal genes (Francois et al., 1994; Ip et al., 1992a; Jazwinska et al., 1999a, 1999b; Kosman et al., 1991; Leptin, 1991; McDonald et al., 1998; Mellerick and Nirenberg, 1995; Rao et al., 1991; Ray et al., 1991). Expression of neuroectodermal genes is excluded from the mesoderm by the Sna repressor, which binds to *cis*-regulatory enhancer sequences of these direct target genes (Francois et al., 1994; Ip et al., 1992a; Kosman et al., 1991; Leptin, 1991; Mellerick and Nirenberg, 1995; Rao et al., 1991). Sna expression is restricted to the most ventral nuclei of the embryo because the Dorsal binding sites in its *cis*-regulatory enhancer are low-affinity sites that can only be fully occupied by the high nuclear concentrations of Dorsal present ventrally (Ip et al., 1992b; Jiang et al., 1991; Kosman et al., 1991; Leptin, 1991; Rao et al., 1991; Ray et al., 1991). Full ventral activation of Sna also depends on the Twi transcription factor (Ip et al., 1992b; Thisse et al., 1988), which, like Sna, is activated by Dorsal. In contrast to the mesodermal targets of Dorsal, low levels of Dorsal can activate neuroectodermal genes because their *cis*-regulatory elements contain high-affinity Dorsal binding sites (Ip et al., 1992a; Mellerick and Nirenberg,

* Corresponding authors.
*E-mail addresses:* ebier@ucsd.edu (E. Bier), hwa@ucsd.edu (T. Hwa).
[1] Contributed equally to this work.
[2] Present address: Center for Complex Biological Systems & Department of Cell and Developmental Biology, University of California Irvine, Irvine, CA 92697, USA.
[3] Present address: Departments of Biology and Genetics, Case Western Reserve University, Cleveland, OH 44106, USA.

0012-1606/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2011.08.016

1995; Rusch and Levine, 1996) and optimal spacing between Dorsal and Twist binding sites (Crocker et al., 2008; Szymanski and Levine, 1995). These neuroectodermal genes are therefore expressed more laterally than Sna, but are repressed in ventral-most nuclei by Sna, which acts directly by binding to sites in neuroectodermal enhancers. Sna expression sharpens during early blastoderm stages culminating in the formation of an on/off border with the neuroectoderm. Expression of Sna in this sharply defined domain results in full repression of all neuroectodermal genes in the presumptive mesoderm.

Whereas the genetic mechanisms leading to establishment of mesodermal versus neuroectodermal domains have been well characterized, less is known regarding the temporal sequence by which the relevant gene expression patterns resolve into their final sharp domains. To analyze this highly dynamic process with greater precision, we developed quantitative imaging and data analysis tools to measure the expression levels of several genes at a time using multiplex in situ hybridization on a nucleus-by-nucleus basis (Kosman et al., 2004). Analysis of such multiplex data reveals a transient intermediate stage in the formation of the Mesoderm/Neuroectoderm Border (MNB) in which expression of neuroectodermal genes is differentially repressed by Sna in a nested pattern within the mesoderm. While several factors are likely to contribute to this graded response, quantitative modeling of this process suggests that one important input is the delay in transcriptional repression associated with the differing transcription unit lengths of neural genes. We confirm an expectation of this transcriptional-delay model, namely that the gap gene *knirps* (*kni*), a short gene, and its paralog *knirps-related* (*knrl*), a long gene, should transiently be expressed in domains of differing width. We propose that gene length may provide a transient prepattern bias that is then amplified and stabilized by well-characterized interactions between transcriptional regulators and *cis*-regulatory modules (CRMs).

## Materials and methods

### *In situ hybridization and immunofluorescence*

A detailed description of fluorescent methods used to detect multiple mRNA transcripts is given in (Kosman et al., 2004). Briefly, Bluescript (*sna*, *brk*, *vnd*, *rho*) or pNB40 (*sog*) vectors harboring full-length cDNAs containing the entire coding sequence (1.7 kb *sna*; 4.7 kb *sog*; 3 kb *brk*; 3 kb *vnd* transcript that corresponds to the shorter transcription unit expressed early in embryogenesis; 2.5 kb *rho*) were linearized using HindIII (*sna*, *sog*, *rho*) or Xba (*brk*, *vnd*) and transcribed with RNA polymerase T3 (*sna*, *rho*) or T7 (*sog*, *brk*, *vnd*) to generate RNA antisense probes labeled with digoxigenin (DIG), biotin (BIO), dinitrophenol (DNP) and fluorescein (FITC) haptens. Primary antibodies used in this study were: sheep anti-DIG (Roche, diluted at 1:1000), mouse anti-BIO (Roche, diluted at 1:1000), rabbit anti-DNP (Invitrogen, diluted at 1:2000), guinea-pig-anti-FITC (diluted at 1:1000), and guinea-pig-anti-Sna (kindly provided by Michael Levine, diluted at 1:300). Alexa-fluor-coupled secondary antibodies were obtained from Invitrogen and diluted at 1:500 (donkey anti-sheep, donkey anti-mouse, donkey anti-rabbit, donkey anti-guinea pig). To avoid an observed unspecific cross-reaction between the guinea pig anti-Sna and mouse anti-BIO, incubation and secondary detection of mouse anti-BIO was carried out before incubation with guinea pig-anti-Sna.

### *Drosophila stocks*

Oregon-R strain was used as wild type. Df(2L)Sco[rv25]/CyO (Bloomington Stock Center) contains a deficiency for the *sna* locus and was used to generate $sna^{-}/+$ embryos. A single *sna* NTD per nucleus distinguishes $sna^{-}/+$ embryos from wild-type and $sna-/-$ embryos.

### *Image acquisition*

All fluorescently stained embryos were counterstained with the fluorescent nuclear marker DAPI, flat-mounted, and imaged as optical sections on a Leica SP2 or a Zeiss LSM700 scanning confocal microscope with a 63X objective lens and digital zoom of either 1× or 2.5×. The gain and offset of the photomultiplier tubes were set so that all the pixels of interest fell within the 8-bit dynamic range.

### *Data analysis and modeling*

The Supporting text outlines the computational methods used to analyze the fluorescence images, together with a model of the dynamic *sna* cytoplasmic mRNA profile and a model relating the calculated *sna* profile to the neuroectodermal NTD profiles.

## Results and discussion

### *Transient formation of a sna gradient in the dorsal mesoderm*

It has previously been noted that *sna* expression is initiated in the 10–12 most ventral mesodermal cells and then expands to occupy the full mesoderm (18 cells wide) (Fowlkes et al., 2008; Ip et al., 1992b). During this period of early *sna* expression, neural genes are transiently expressed in the mesoderm before sufficient levels of Sna repressor accumulate to shut them off. Once *sna* expression occupies the full expanse of the mesoderm, its dorsal border abutting the neuroectoderm becomes very sharp (O'Neill and Bier, 1994), and shortly thereafter leads to the induction of a single row of *single minded* (*sim*) expressing mesectodermal cells along the ventral border of the neuroectoderm (Bardin and Schweisguth, 2006; Chitnis, 2006; Cowden and Levine, 2002; Morel et al., 2003; Morel and Schweisguth, 2000; Schweisguth, 2004). By this stage neuroectodermal gene expression in the mesoderm is fully repressed.

To examine the refinement of the MNB with greater precision, we coupled multiplex FISH (Kosman et al., 2004) with image segmentation techniques developed in this work (see Supporting text) to analyze the simultaneous expression of several gene markers within the mesoderm on a nucleus-by-nucleus basis at various times during nuclear cycle 14 of the blastoderm embryo. This strategy allowed us to create detailed snap-shots of Sna and neuroectodermal gene expression in the mesodermal domain. We used DAPI staining to create 3D segmentations of nuclei thereby resolving cytoplasmic and nuclear signals. Once the nuclei were segmented, we extracted robust signals from sites of nascent transcription located within nuclei, also referred to as nuclear transcription dots or NTDs, by thresholding the nuclear mRNA fluorescence intensity. NTDs arise from newly synthesized mRNAs clustered at the site of genes encoding those nascent transcripts (Wilkie et al., 1999), and are ideally suited to reveal rapid changes in gene expression patterns since they serve as a direct read-out of transcriptional regulation.

Because the Sna repressor is known to bind directly to the *cis*-regulatory enhancer sequences of neuroectodermal genes (Gray et al., 1994; Ip et al., 1992a; Jiang and Levine, 1993; Zeitlinger et al., 2007b), we first quantified the establishment of the *sna* expression profile within the mesodermal domain. Co-staining early cycle-14 blastoderm embryos with an antisense mRNA probe to detect *sna* transcripts and an anti-Sna antibody to detect Sna protein revealed gradients of both *sna* mRNA (Fig. 1A) and Sna nuclear protein (Fig. 1B), consistent with prior reports of a ventral-to-dorsal gradient of sna cytoplasmic mRNA expression (Fowlkes et al, 2008) (Fig. S3A). Quantification revealed a linear correlation between cytoplasmic *sna* mRNA and nuclear Sna protein over a broad range of individual nuclei (Fig. 1C), validating the use of *sna* cytoplasmic mRNA

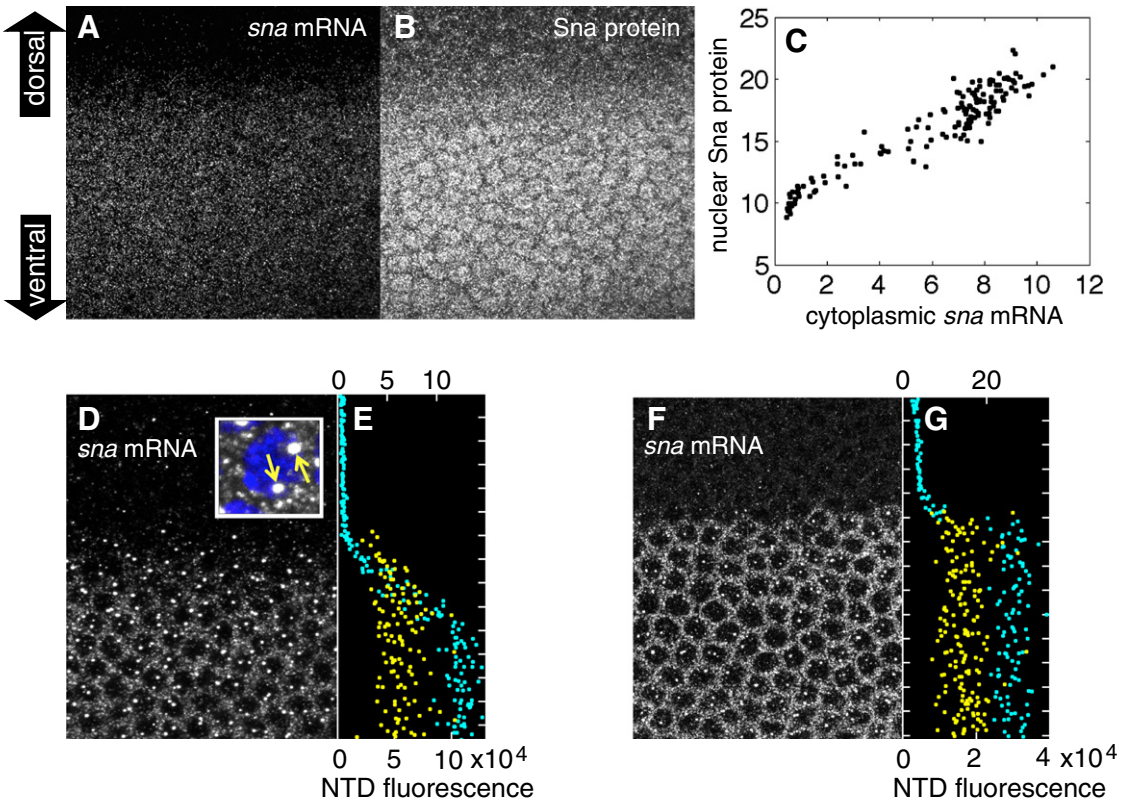


**Fig. 1.** Quantification of a transient *sna* gradient. (A) Early nuclear cycle 14 embryo stained for *sna* mRNA via fluorescent in situ hybridization. The *sna* mRNA fluorescence shown is a maximum projection of a confocal stack spanning the apical cytoplasm where *sna* cytoplasmic mRNA is localized. The embryo is oriented with ventral side down. Anterior is to the left and posterior is to the right. (B) The same embryo as in (A) co-stained for Sna protein. The image is a maximum projection of a confocal stack spanning the apical-basal extent of the nuclei. The fluorescence has been amplified by a factor of 4 to better visualize the nuclear localization of Sna protein, but the original unamplified data was used in our analysis. (C) Sna protein in individual nuclei is linearly correlated with *sna* cytoplasmic mRNA. The signals represent voxel intensities averaged over nuclear segmentations and cytoplasmic Voronoi cellular reconstructions. Data were taken from the individual embryo shown in panels A, B. (D) Another early cycle 14 embryo stained for *sna* mRNA representing a confocal stack spanning the nuclei. Cytoplasmic mRNA stains in a dorsal–ventral gradient (raw data amplified by factor 2.5 to make gradient more apparent in this panel but unamplified data used in subsequent analysis). Two strong foci of *sna* mRNA fluorescence intensity (referred to as NTDs in the main text) are visible in most nuclei in the presumptive mesoderm (lower half of panel). Inset shows a high-magnification image of two NTDs (yellow arrows) in a single nucleus (indicated by DAPI staining in blue). (E) *sna* NTD dorsal–ventral profile, corresponding to the single embryo shown in (D), is plotted in yellow with NTD intensities corresponding to the bottom axis. Dorsal–ventral position of the center of mass of each nucleus is plotted in cyan against the mean fluorescence intensity per voxel of *sna* mRNA in the associated cytoplasm (quantification of fluorescence intensity is indicated on the top axis). (F, G) Same as (D, E) except that data was acquired from a single embryo in the late cellularization stage of nuclear cycle 14. The *sna* mRNA signal is amplified (by a factor of 2) in (F) but not in (G). Notice the much sharper profile of *sna* mRNA intensity compared to that in (D, E).

as a surrogate for Sna protein levels in the nucleus. In subsequent experiments we made use of this correlation between transcript and protein levels to estimate Sna input protein levels while simultaneously detecting multiple target gene transcripts by multiplex FISH.

We next analyzed *sna* cytoplasmic mRNA levels as a function of dorsal–ventral position of individual nuclei (Fig. 1D). A nucleus-by-nucleus quantification revealed a *sna* concentration gradient with highest levels surrounding the 12 most ventral nuclei of the mesoderm, and progressively lower levels more dorsally in the border zone, comprising approximately 3 mesodermal nuclei on each side of the MNB (cyan profile in Fig. 1E, corresponding to the field of view shown in panel D). It is noteworthy that in contrast to the graded profile of cytoplasmic *sna* mRNA (and presumably also Sna protein since these two variables are linearly correlated), the distribution of *sna* NTDs exhibited a much sharper border along the dorsal–ventral axis (yellow profile in Fig. 1E). An informative detail of this analysis (see Supporting text) is that the graded *sna* cytoplasmic mRNA profile (cyan) *trails* the frontier of nascent *sna* expression (yellow) defined by the NTDs (Fig. 1E). By the end of cellularization stage of cycle 14, the *sna* cytoplasmic mRNA gradient sharpens to form a one-nucleus-sharp profile (Fig. 1F with quantification indicated by cyan profile in G).

### *An expansion-accumulation model accounts for the transient sna gradient*

We analyzed the graded profile of *sna* cytoplasmic mRNA near the MNB in early cycle 14 embryos and found it unlikely to be caused by diffusion (Supporting text). This result prompted us to search for an alternative explanation for the observed sharpening of the *sna* gradient. The previously noted dorsally expanding domain of *sna* expression (Fowlkes et al., 2008; Ip et al., 1992b) suggested a potential mechanism for generating a transient *sna* gradient abutting the MNB, in which an on–off profile of *sna* transcriptional activity encompassing the 10 ventral-most nuclei at t = 0 min (onset of cycle 14), expands dorsally at a steady rate to encompass the 18 ventral-most nuclei at t = 20 min; see Fig. S3B. Using this advancing front of *sna* transcriptional activity as input to a mathematical model that invokes only synthesis and turnover of *sna* mRNA (Eq. (S2) in Supporting text), we obtained a dynamic *sna* profile which is graded at times shortly after the onset of expansion (Fig. S3E; $t = 20$ min). This profile is similar to that observed in early cycle 14 (Fig. 1E) and results from the differential accumulation of *sna* mRNA at different locations as the front expands. The profile then gradually sharpens as the border of *sna* transcription advances dorsally to its final location at the MNB. Finally, attenuation of *sna* levels (due to turnover of *sna* cytoplasmic mRNA) generates an abrupt expression

border (Fig. S3E; $t \approx 50$ min), similar to what is experimentally observed in late cycle 14 (Fig. 1G). A detailed description of this mechanism can be found in the Supporting text.

*Neural genes are expressed in a stratified pattern in the mesoderm of early embryos*

As mentioned above, many neuroectodermal genes are transiently expressed in the mesoderm prior to sufficient levels of Sna accumulating to repress them ventrally. To test whether the degree of neural gene expression was related to the dynamic pattern of Sna expression, we examined the relative pattern of neural gene expression during the expansion phase of Sna expression.

We quantified the extent and overlap of neural transcription patterns at different stages in the mesoderm by creating nuclear masks for fields of nuclei straddling the MNB. In these reconstructed fields of nuclei, each nucleus is represented by a circle whose color indicates the combination of neuroectodermal genes active in that nucleus (Fig. 2A). In early blastoderm-stage embryos (20–25 min into cycle 14), these on/off coded gene expression maps accurately reproduce the known relative positions of dorsal borders of neuroectodermal gene expression (Francois et al., 1994; Jazwinska et al., 1999a; Kosman et al., 2004; Mellerick and Nirenberg, 1995). For instance, *brinker* (*brk*) expression shown in red extends further dorsally than *rhomboid* (*rho*), shown in magenta, followed by *ventral nervous system defective* (*vnd*), shown in white. In addition, this analysis revealed a transient nested series of distinct ventral borders for these genes within the presumptive mesoderm. In these embryos, the ventral limits of *rho* (indicated by cyan) and *vnd* (indicated by green) extended approximately 1–2 nuclear diameters into the presumptive mesoderm (as defined by the ventral *brk* border). *vnd* has previously been shown to be expressed approximately one nuclear diameter more ventrally than *rho* (Zinzen et al., 2006), and our data are consistent with this assessment. Finally, expression of the *short gastrulation* (*sog*) gene is transcribed yet further into the mesoderm (shown in white in Fig. 2B, C, E, F; see below).

At the low magnification used for imaging the embryo in Fig. 2A, signals from NTDs are difficult to quantify since they consist of only a small number of voxels. To study the pattern of NTDs more quantitatively, we imaged expression of *sna* and various combinations of neuroectodermal genes at higher magnification in another embryo and reconstructed a field of nuclei at a higher voxel density (Fig. 2B; same embryo as in Figs. 1D and E). In the coded gene expression map representing this high-magnification reconstruction, the degree of white-to-gray in a nuclear segmentation represents the level of Sna mRNA staining with white representing the highest levels. As in the case of the lower-magnification field (Fig. 2A), overlapping patterns of neuroectodermal gene expression were clearly resolved within the presumptive mesoderm (Fig. 2B). By comparing gene expression maps obtained with various combinations of the four probes at this early stage, we confirmed our conclusions based on lower-magnification reconstructions that *sog* exhibits the broadest degree of mesodermal expression of the neuroectodermal genes analyzed (Fig. 2B), followed by *vnd* (Figs. 2A, B) and *rho* (Fig. 2A), and that *brk* is strictly excluded from the mesoderm (Figs. 2A, B).

It is also possible to extract relative intensities of the spectral outputs of NTDs when imaged at high magnification by integrating voxel fluorescence intensity over each NTD (Supporting text). We plotted the integrated spectral outputs (total intensity) from each NTD as a

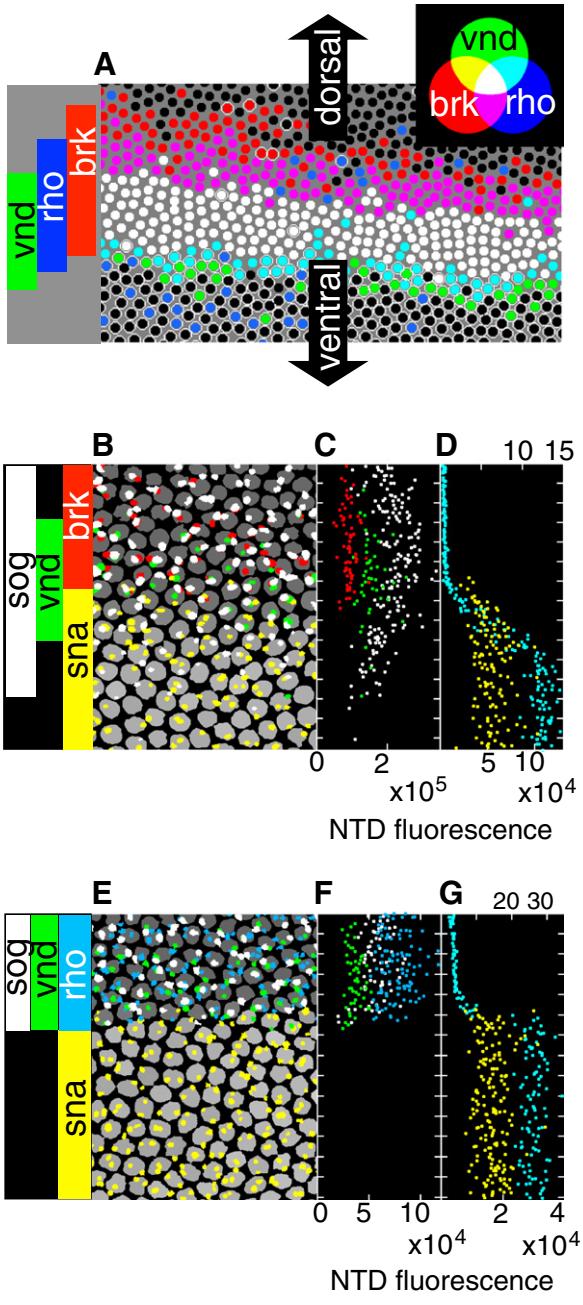


**Fig. 2.** Neuroectodermal genes are differentially repressed in the mesoderm. (A) Reconstruction of a low-magnification in situ hybridization confocal stack taken from an individual embryo, imaged at early-mid cycle 14, showing expression of three neuroectodermal genes (*brk*, *vnd*, and *rho*) and the mesodermal transcriptional repressor *sna*. The circles represent nuclei and the colors indicate the combination of Nuclear Transcription Dots (NTDs) detected for the various neuroectodermal genes. Black indicates that no neuroectodermal NTDs were detected. A circle circumscribed by a white outline indicates a nucleus in which a *sna* NTD was detected. (B) Reconstruction of the individual embryo shown in Fig. 1D with the same dorsal–ventral and anterior–posterior orientation. The area of each NTD is proportional to the projected area of the 3D NTD from which it is reconstructed, but the projected areas are magnified by a fixed factor to aid visualization. Also shown is a single optical section through the set of 3D nuclear segmentations, where the shading of the nuclear masks is proportional to the mean voxel intensity of *sna* cytoplasmic mRNA in the corresponding Voronoi cells surrounding each nucleus (see Supporting text). Notice the jagged sna NTD border. (C) The dorsal–ventral position of each NTD shown in (B) is plotted against the integrated fluorescence intensity of the NTD. Horizontal tick marks represent average nuclear spacing. Raw spatial profiles are uncorrected for background, variations in number of fluorophores associated with a target mRNA, or fluorophore efficiency. (D) *sna* NTD and cytoplasmic mRNA dorsal–ventral profile are reproduced from Fig. 1E for direct comparison with neuroectodermal gene expression. (E) Same format as (B), but for an individual embryo in late cellularization stage of cycle 14 stained for *rho* instead of *brk*. Notice that an on/off border separates the *sna* and neuroectodermal gene expression domains, which are virtually mutually exclusive at this stage and that the *sna* expression border defined by *sna* NTDs is straight in contrast to the jagged, but equally sharp, NTD border present in earlier embryos (compare the straight yellow border of reconstructed *sna* NTDs in this panel to the meandering border evident in (B)). (F, G) Same format as (C, D).

function of dorsal-ventral position in the embryo (Fig. 2C). In the case of *sog*, which is expressed furthest into the mesoderm, we found that there is a graded ventral decline in the total intensity of NTDs (Fig. 2C; see also Fig. 2B). The nested expression patterns of neuroectodermal genes within the dorsal region of the mesoderm and the graded size of NTDs in this same zone suggest that the expression of these genes is regulated in response to a gradient.

### *Gene-length correlates with the observed patterns of nested neural gene expression in the mesoderm*

The dorsal-most portion of the mesoderm over which *sna* levels are graded coincides with the zone of differential neuroectodermal gene expression (compare Fig. 2D with B or C), suggesting that Sna may differentially repress neuroectodermal genes within this region. Previous detailed studies have established that neural gene repression in the mesoderm is accomplished by direct binding of Sna to enhancers of these genes (Gray et al., 1994; Ip et al., 1992a; Jiang and Levine, 1993; Zeitlinger et al., 2007b), and differential DNA occupancies of Sna protein on these CRMs has been determined using whole-genome tiling arrays (MacArthur et al., 2009). One obvious possibility is that the CRMs of the various neural genes have differing affinities or numbers of Sna binding sites and that genes transiently expressed more ventrally into the mesoderm (e.g., *sog*) have fewer and/or weaker Sna sites than those whose expression is strictly limited to the neuroectoderm (e.g., *brk*). However, the predicted order of nested gene expression based on relative DNA occupancy of Sna protein differs from that observed. The relative CRM occupancy order is *sog*, *rho*, *brk*, *vnd*, in which *sog*/*rho* exhibit weak Sna affinity whereas *brk*/*vnd* exhibit high Sna affinity (Fig. S6A), whereas the observed expression order is *sog*, *vnd*, *rho*, *brk*. Thus, differing sensitivities of neural CRMs to Sna do not seem to play the primary role in this patterning process, although they are likely to contribute to it (see analysis of *sna*+/− heterozygote data below).

Since differential *cis*-mediated regulation of neural target genes by Sna seemed unable to explain the nested pattern of neuroectodermal gene expression in the mesoderm, we considered alternative models. One correlation we observed is that the order of neural gene expression in the mesoderm corresponds with that of transcription unit length. We reasoned that if Sna acted by repressing initiation of gene expression, then longer genes would take more time to be repressed than shorter ones since it would take more time for previously loaded RNA Polymerase (RNAP) molecules to traverse longer genes. For example, a long transcription unit such as *sog* (~22 kb in length) would require ≈20 min to be fully transcribed, whereas a short gene such as *brk* (~3.5 kb) would take only about 3.5 min, assuming that RNAP progresses at its typical rate of ≈1 kb per minute (O'Brien and Lis, 1993; O'Farrell, 1992). Thus, completion of transcription and hence onset of repression for such a pair of long versus short genes could amount to a differential delay of ≈15 min, which is a significant fraction of the cellularization stage of nuclear cycle 14. To examine this idea more fully, we developed a "delayed-repression model" based on the transient Sna gradient acting on target genes of differing gene lengths. As described below, this simple model can indeed account for much of the nested pattern of nascent neuroectodermal gene expression observed in vivo in wild-type embryos (Fig. 2C).

### *A simple gene-length model predicts the observed expression data in wild-type embryos*

The integrated fluorescence intensity level of an NTD is expected to be proportional to the number of nascent transcripts on the gene when plotted in individual cells against their positions relative to the MNB. We therefore determined the expected space- and time-dependent intensity profiles of *sog* NTDs and developed a mathematical model (Eq. (S3) in Supporting text) that incorporates the kinetics of forming the Sna gradient with the delayed repression effect described above. Fig. 3A depicts a typical *sna* cytoplasmic mRNA time course (blue) as obtained from the expansion-accumulation model, and a threshold concentration (horizontal dashed line) corresponding to a hypothetical Sna level (Eq. (S5)) that completely blocks transcriptional initiation. The black solid curve in Fig. 3A is the resulting time course of the *sog* NTD intensity (Eq. (S3); see discussion in the Supporting text.) Aligning the time courses of *sog* NTD intensity for all nuclei along the dorsal–ventral axis yields the surface depicted in Fig. 3B. In Fig. 3C, we extracted two time slices from this calculated surface, yielding a transient gradient of *sog* NTD intensity (t = 25 min) that sharpens to form an on–off profile (t = 50 min.). These calculated profiles can account for the experimentally observed *sog* patterns at early and late cellularization stage of cycle 14 (the white scatter plots in Figs. 2C and F respectively).

The respective transcription-unit lengths of the *sog*, *vnd* and *brk* genes during cycle 14 are: *sog* ≈ 22 kb, *vnd* ≈ 7 kb, and *brk* ≈ 3.5 kb. The delayed-repression model predicts the transient formation of a gradient of neuroectodermal nascent gene expression, which has a width (the dorsal–ventral distance over which neuroectodermal gene expression changes) that varies in proportion to gene length (explained in Supporting text) as depicted in Fig. 3D. These predicted early gene expression profiles are similar to those observed experimentally (Fig. 2C).

Thus, the simple model presented here, based on a ventral-to-dorsal expansion of *sna* expression and the delayed repression of neuroectodermal genes of differing lengths, accounts for not only the observed gradient in *sna* mRNA concentration near the MNB, but also much of the nested pattern of neuroectodermal nascent gene expression within the mesoderm of wild-type embryos. We emphasize that according to this highly simplified model, the *cis*-regulatory regions of *brk*, *vnd* and *sog* need not be differentially sensitive to Sna repression to generate the transient, distinct spatial expression profiles shown in Fig. 3D. Of course, different sensitivities to Sna repression for each of the neuroectodermal enhancers are also likely to contribute to generating such patterns (as detailed below).

### *Widespread ectopic neural gene expression in the mesoderm of sna+/− embryos*

In addition to examining the pattern of neural gene expression in the mesoderm of wild-type embryos, we also examined their expression in *sna*−/+ heterozygous embryos. This genetic background allowed us to test the *cis*-regulatory and gene-length models by comparing observed expression patterns with those predicted by a 50% reduction of Sna levels (see Supporting text). As expected, there is a graded profile of *sna* (Fig. 4D) of the same width (~4 nuclei), but with half the maximal level of that measured in wild-type embryos (Figs. 1E and 2D).

In *sna*−/+ embryos of early-mid cellularization stage of cycle 14, on/off plots of NTD expression for *brk*, *vnd* and *rho* revealed that *vnd* (green) and *rho* (blue) become distributed broadly and chaotically throughout the mesoderm, whereas expression of *brk* (red) retains a sharp ventral border (boundary between cyan and white in Fig. 4A). Quantitative analysis of NTD intensities confirmed the conclusions from the on/off analysis (Figs. 4B–D). In contrast to the nested gradients of *sog* and *vnd* NTD intensity previously observed in wild-type embryos (Fig. 2C), the mesodermal intensities of NTDs for these genes in early *sna*−/+ heterozygous embryos are nearly equal to their respective values in the neuroectoderm (Fig. 4C), indicating a nearly complete de-repression of these genes ventrally. Ventral de-repression of neuroectodermal genes is still apparent in the mid-stage embryo shown in Figs. 4E and F, although a fraction of nuclei have lost *sog* and *vnd* NTDs by this stage. Despite the prolonged persistence of high-level neuroectodermal gene expression in the

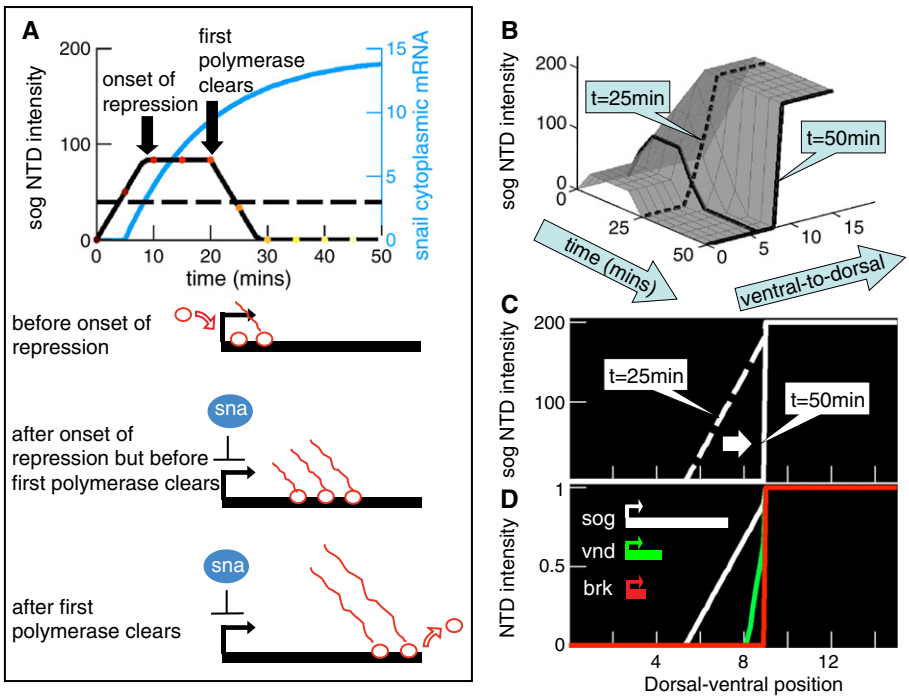


**Fig. 3.** Differences in neuroectodermal gene length yield a stratified pattern of neuroectodermal nascent gene expression in a delayed-repression model. (A) Time course of *sog* NTD intensity (black line) in a mesodermal nucleus as calculated in the delayed-repression model (Eq. (S3)), using as input the *sna* cytoplasmic mRNA levels (blue line) calculated in the expansion-accumulation model (Eq. (S2)). The horizontal dashed line indicates the threshold concentration for *sna* repression ($K=3$; see Eq. (S5)). The delay in repression arising from the length of the *sog* gene is assumed to be $\tau=20$ min. (Eq. (S4)). (B) *sog* NTD time courses for various nuclei can be combined to construct a surface describing the spatiotemporal dependence of *sog* NTD intensity. The black line corresponds to the temporal profile analyzed in (A). Spatial profiles at early (t = 25 min.) and late (t = 50 min.) cellular blastoderm are indicated. (C) Slices taken from the surface in (B) at early (t = 25 min.) and late (t = 50 min.) stages reveal a transient spatial gradient of *sog* NTD intensity. (D) The delayed-repression model (Eq. S3) predicts that transient nested gradients of *sog*, *vnd* and *brk* NTD intensity, similar to those observed experimentally (Fig. 2C), occur in early cycle 14 (t = 25 min.) due solely to differences in neuroectodermal gene length. Transcriptional delays are assumed to be $\tau=20$ min. (*sog*), 6 min. (*vnd*) and 2 min. (*brk*), corresponding respectively to gene lengths of 22 kb (*sog*), 7 kb (*vnd*), and 3 kb (*brk*). These estimated gene lengths are based on RNA-seq data available through the modEncode *Drosophila* genome browser located at http://modencode.oicr.on.ca/fgb2/gbrowse/fly/ and on ChIP-chip polymerase-II binding data in early embryos (Zeitlinger et al., 2007a); see also Fig. S2 of (Boettiger and Levine, 2009). Note that *vnd* has an alternate remote upstream promoter but the modEncode RNA-seq data indicate that this is not used until later in embryogenesis. The transcription initiation rate was chosen to be $\alpha_{on}=10$/min (same for all neuroectodermal genes), though its actual value does not affect our conclusions ($\alpha_{on}$ is defined in Eq. (S5)).

mesoderm of *sna*+/− embryos relative to wild-type embryos, by late cellularization stage the typical one-nucleus sharp *sna* gradient is eventually established (Figs. 4H–J).

### *A role for cis-acting inputs in Sna-mediated repression of neural genes*

The observed relative expression patterns of neuroectodermal genes in *sna*−/+ heterozygotes would not be predicted by models based solely on CRM sensitivity as the response of the *vnd* CRM should be weak like that of *sog*/*rho* (all of which are ventrally de-repressed) whereas Sna ChIP data (Fig. S6A) indicate that the *vnd* CRM is actually more fully occupied than that of *brk* (which is never ventrally de-repressed). Widespread and prolonged mis-expression of *sog* (a long gene) in the mesoderm of *sna*−/+ embryos during early-mid cycle 14 (Figs. 4B–G) is predicted, however, by the delayed-repression model developed above (see Fig. S5B, C and Supporting text). In addition, this simple model correctly predicts (Fig. S5C) the observed repression of *sog* from the mesoderm by the end of cellular blastoderm stage (Figs. 4H–J), giving way to a common sharp border of nascent expression. Another expectation of the model is that expression of the short gene *brk* would remain restricted in *sna*−/+ heterozygotes as it is in wild-type embryos.

Although the simple gene length model captures many features of the data presented thus far, one observation that is not accounted for is that the expression patterns of *rho*/*vnd* in *sna*−/+ embryos expand chaotically within the mesoderm, similar to that of *sog* (Fig. 4). The gene-length model predicts that *rho*/*vnd* should behave more like *brk* (which is always ventrally repressed), however, since all three genes are relatively short. The discrepancy can be resolved by incorporating differential CRM sensitivity to Sna repression into the gene-length model. We therefore scaled the threshold *sna* mRNA concentration for the onset of neuroectodermal repression ($K$ in Eq. (S5)) with the inverse of the Sna sensitivity score (Fig. S6B) in a revised model, which alters its predictions (compare Fig. S6C and D) and more accurately models the observed data (see Supplementary text for additional details).

### *A natural test of the gene-length hypothesis*

As described above, the modified gene-length model accounts for key features of the observed transient patterns of neural gene expression within the mesoderm. A natural test of this model is fortuitously provided by a pair of genes, *knirps* (*kni*, a short gene—3 kb) and *knirps-related* (*knrl*, a long gene—23 kb), whose expression along the anterior–posterior axis is thought to be controlled by a common CRM (Rothe et al., 1989, 1992). This identified CRM is located just upstream of the *kni* gene and drives reporter gene expression in a pattern coinciding with that of the endogenous central stripes of both *kni* and *knrl* at blastoderm stage (Rothe et al., 1992). Analysis of this phase of *kni* and *knrl* expression has implicated two repressor gradients, Hunchback and Giant, that respectively contribute to setting the anterior and posterior limits of *kni*/*knrl* expression (Capovilla et al., 1992; Eldon and Pirrotta, 1991; Hewitt et al., 1999; Hulskamp et al., 1990; Hulskamp et al., 1994; Rothe et al., 1994; Zinzen and Papatsenko, 2007).

On the lateral surface of the embryo, the *kni* stripe undergoes substantial shifts along the anterior–posterior axis during mid to late blastoderm stage of cycle 14 (Jaeger et al., 2004). In particular, the

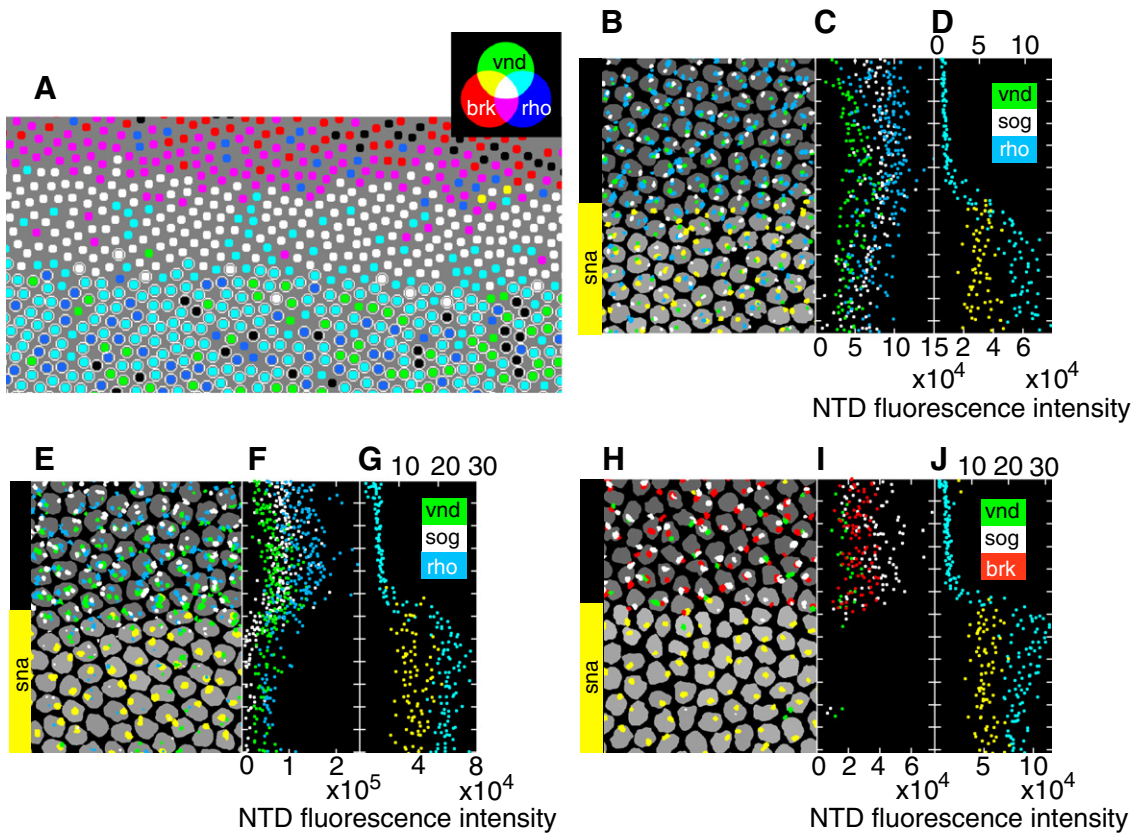


**Fig. 4.** Widespread ectopic neuroectodermal gene expression in *sna*−/+ heterozygous embryos illustrates the haploinsufficiency of *sna*. (A) Reconstruction of an individual *sna*−/+ embryo in early-mid cellularization stage of cycle 14 (same format as Fig. 2A). There is widespread mis-expression of *vnd* and *rho* in the mesoderm; in contrast, *brk* is completely excluded from the mesoderm as in wild-type embryos. (B–D) Reconstructed image of a single early cycle 14 embryo, similar to Figs. 2B–D. Note that mesodermal nuclei contain only one NTD of *sna* each, confirming that the embryo is a *sna*−/+ heterozygote. (E–G) Mid-cellularization-stage cycle 14 embryo. (H–J) Late cellularization-stage cycle-14 embryo.

*kni* posterior border moves toward the anterior during the time interval defined by 0–50% membrane invagination (Fig. S4 of Ref. (Fowlkes et al., 2008)), presumably as a result of transcriptional repression by the Giant protein. The gene-length model predicts that the observed anteriorly-advancing front of *kni* transcriptional repression should leave behind a wake of NTDs of the longer *knrl* gene, mainly in nuclei located posterior of the *kni* stripe.

We tested the prediction that the *knrl* expression domain is broader than that of *kni* by comparing the expression domains of these sibling genes along the lateral surface of early (Figs. 5A–C) and mid-late (Figs. 5D–I) cellularization-stage embryos. Reconstruction of the *kni* and *knrl* NTDs (Figs. 5C and F) and quantitations of their anterior–posterior profiles (Figs. 5A–B and G–I) indicated that nascent *knrl* transcripts (red) coincided with *kni* NTDs (green) at early stages (Figs. 5A–C), but were more broadly distributed than those of *kni* at later stages (Figs. 5F—H). The expanded expression of *knrl* relative to *kni* was most notable along the posterior border of these gene expression domains (Fig. 5I). The difference between *knrl* and *kni* expression observed in this embryo was also observed in other embryos (Figs. S8, 9). We conclude that the gene-length hypothesis is likely to account for the observed differences in *kni* versus *knrl* expression.

### *Timing as a prepatterning bias to enhance morphogen patterning*

The studies reported here suggest that variations in gene length of target genes may contribute to robust interpretation of crude positional information provided by morphogen gradients. As a generalization of the trend we observed in the mesoderm, we propose that one can define an activation-repression axis for many well characterized morphogen gradients. Signal transduction pathways can activate gene expression either by directly promoting transcription and/or by relieving constitutive repression of target genes (Barolo and Posakony, 2002). In cases where morphogens also act by a double-negative mechanism, this repressor transcriptional delay mechanism could act as a timing mechanism to generate spatial pattern. Direct activation of gene expression can also, in principle, provide temporal/spatial information based on gene length whereby shorter genes will be transcribed and translated sooner than longer genes. Both delayed repression and precocious activation would act in a coupled fashion by favoring short genes to be expressed near the activation (A) pole and at the same time favoring longer genes to be expressed near the repressor (R) pole of a gradient.

We surveyed several well-studied morphogen gradients and their known direct target genes to ask whether there is any correlation between gene length and position on the A/R axis (Fig. 6). This analysis suggests that there is indeed a general trend to obey the A/R rule for target genes of the Dorsal, Sna (this study), and Dpp gradients in the embryo, as well as the Hh and Dpp gradients in the wing imaginal disc. Noteworthy is a recent analysis of Hh signaling in the wing disc that suggests that an early transient pattern of Hh activation is consolidated by stabilizing feedback mechanisms (Nahmad and Stathopoulos, 2009).

The general idea that timing plays an important role in developmental patterning is not new. In the example of stabilization of Hh response in the wing disc, differential rates of protein degradation was proposed as an important factor (Nahmad and Stathopoulos, 2009). Another example of this patterning principle occurs in somitogenesis, where a moving maturation wave front generates a periodic array of somite boundaries. As the wave moves over the tissue,

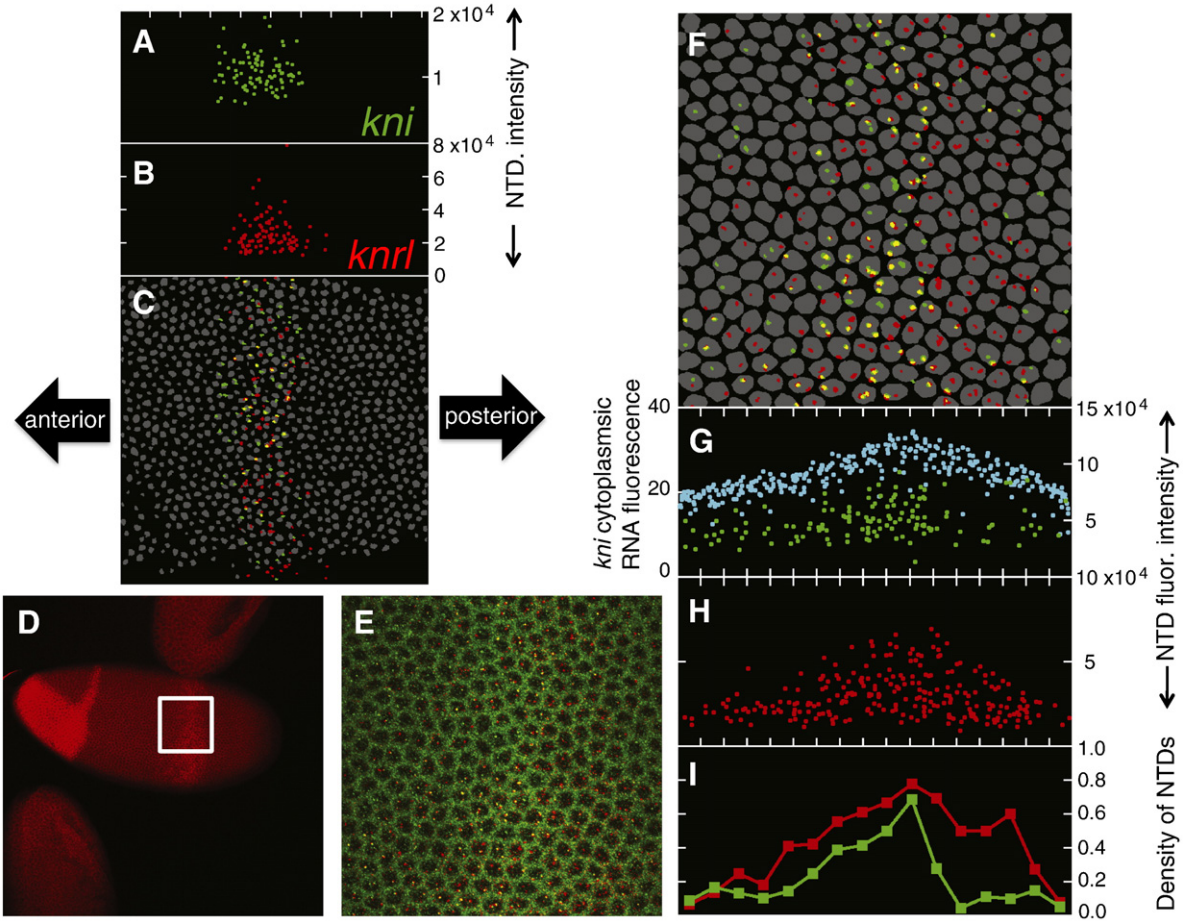


**Fig. 5.** *Kni* and *knrl* constitute a natural test of the gene-length hypothesis. (A–C) Reconstruction (similar to Figs. 2 and 4) of an early-blastoderm-stage embryo with the field of view spanning the central *kni*/*knrl* stripe. Anterior is to the left, posterior to the right, ventral is down and dorsal is up. (D) Mid-blastoderm stage embryo stained for *knrl* mRNA (red) showing expression in an anterior patch and a central stripe. Signal amplified by factor of 2 to aid visualization but not for subsequent analysis. A higher magnification confocal stack (white square) spanning the central stripe was acquired and analyzed in (E–I). (E) Maximum projection of the central *kni*/*knrl* stripe stained for *kni* mRNA (green) and *knrl* mRNA (red). Image corresponds to the white square shown in (D). Signal has been amplified by factor of 2 (*kni*) and 2.5 (*knrl*) to aid visualization but unamplified data were used in subsequent analyses. (F–H) Reconstruction of the embryo in (D) and (E). In (G), *kni* cytoplasmic mRNA fluorescence (cyan) is plotted on the left axis whereas *kni* NTD fluorescence intensity (green) is plotted on the right axis. (I) Anterior–posterior profile of NTD density. For each gene, we quantitated the spatial density of NTDs along the dorsal–ventral axis by binning the anterior–posterior axis, calculating the number of NTDs in each bin, and then normalizing by twice the number of nuclei in the corresponding bin (the factor of two arises because each gene exists in two copies in each nucleus).

cells at the frontier of the wave become "locked-in" to different states according to their phase in a gene expression cycle (Dequeant and Pourquie, 2008; Ozbudak and Pourquie, 2008). Indeed, transcriptional delay is partly responsible for the oscillating gene expression underlying the moving maturation front (Takashima et al., 2011). Another example is the temporally ordered sequence of gene activation in distinct domains of the mesoderm and neuroectoderm in the early *Drosophila* embryo. In this case, more ventrally expressed genes are activated earlier (e.g., *sna->vnd->ind->msh*). Interestingly, their order of activation corresponds to the sequence of cross-repressive interactions observed among transcription factors encoded by these genes (i.e., Sna represses all neuroectodermal genes, Vnd represses expression of *ind* and *msh*, and Ind represses expression of *msh*), a chain of interactions that is referred to as "ventral dominance" (Cowden and Levine, 2003). With regard to the neuroectoderm, prior expression of Vnd should reinforce its ability to effectively repress *ind* expression in the most ventral cells of the neuroectoderm, and for Ind to gain dominance over *msh* in more lateral cells. Patterning of the vertebrate spinal cord has also been proposed to rely on similar timing mechanisms (Kutejova et al., 2009; Lek et al., 2010; Liem et al., 1997).

As mentioned above, there appears to be a general gene-length versus A/R axis trend extending to a variety of patterning systems, including those acting in the wing disc over much longer developmental periods (days) than establishment of the embryonic dorsal-ventral axis (≈1 h). Thus, it may be that even though typical gene-length variation (1 kb–100 kb) can only act over a period of an hour or so, it may nonetheless seed a refinement process resulting in stable robustness that is gained from the early prepattern. For example, the gene-length mechanism may contribute to reliably setting closely spaced borders in cases where noise makes it difficult to initially distinguish corresponding thresholds in a gradient, as in situations where closely spaced borders are set by morphogen gradients (Yu and Small, 2008). Yet another possibility is that transient patterns function to bias cell-fate decisions controlled by commonly occurring cross-repressive interactions and feedback loops. In such cases, bistability may arise in which pre-steady-state dynamics can affect the outcome of which genes are eventually expressed in different domains.

In all of the cases described above, transient biases in gene expression acting in conjunction with a morphogen gradient can help establish discrete gene expression patterns that are then amplified and stabilized by other mechanisms including autoactivating positive feedback loops and cross-repressive interactions. It will be interesting in future studies to test whether transient spatial patterns generated by transcriptional/translational delay mechanisms such as those

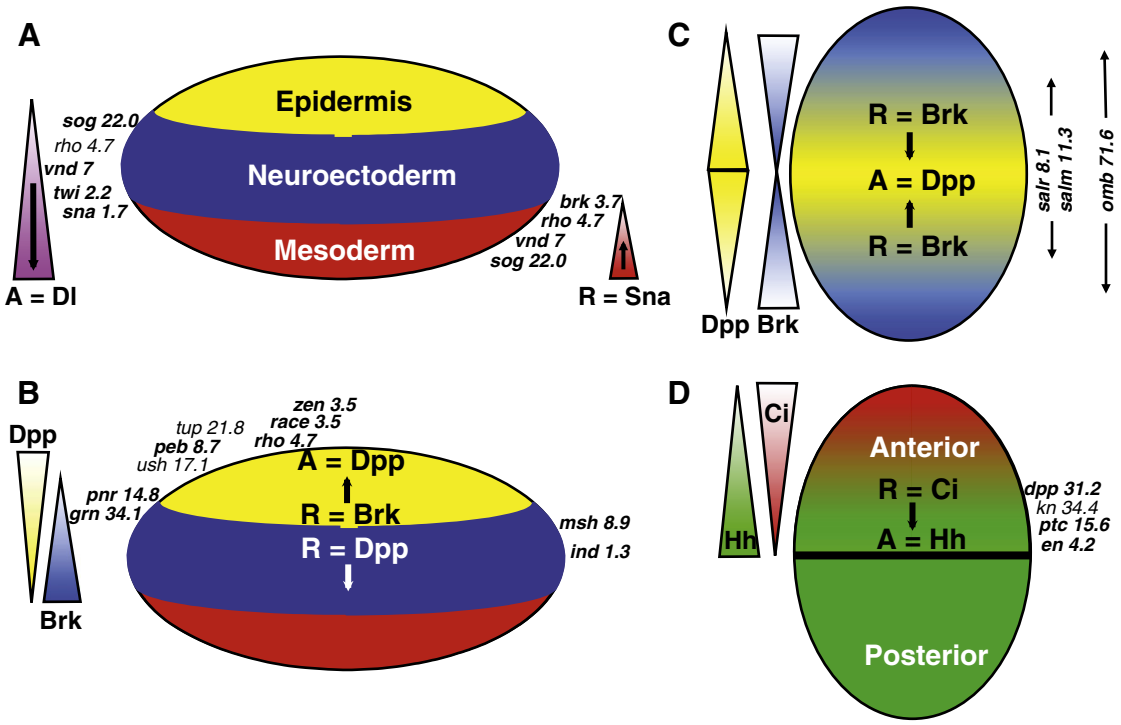


**Fig. 6.** Target genes of the Dl, BMP, and Hh morphogens generally follow the A/R rule. (A) Relative positions of dorsal-most expression borders of *dorsal* target genes, together with their genomic length (in kb), are shown schematically on the left of the embryo. Relative positions of ventral-most expression borders of *sna* target genes, together with their genomic length (in kb), are shown on the right of the embryo. The arrows on the morphogen gradients indicate the hypothesized direction of gradient movement in the region of the embryo where the target genes are expressed. (B) Relative border positions of *dpp* target genes in the dorsal epidermis (left) and relative dorsal-most borders of *dpp* targets in the neuroectoderm (right) of the embryo. Also shown are the oppositely directed gradients of *dpp* and *brk*. (C) Relative extents of gene expression of *Dpp* target genes in the wing imaginal disc with opposing gradients of *Dpp* and *brk*. (D) Relative border positions of *Hh* target genes in the anterior half of the imaginal wing disc along with the reciprocal gradients of *Hh* and *Ci*. In all four panels, there is a general trend, represented by target-gene names in bold type, favoring borders of short genes near the activation (A) pole and borders of longer genes near the repressor (R) pole of the corresponding morphogen gradient.

described here may also contribute to patterning effected by other morphogen gradients.

*Is the early Sna gradient and graded repression of neural genes biologically relevant?*

A fair question to ask regarding the transient formation of a Sna gradient described here is whether it plays any biological role or is just an epiphenomenon associated with the mechanics of Sna activation by the dorsal gradient. Thus, we tested whether the relative sensitivity of the two genes on the extreme end of the spectrum, *sog* versus *brk*, might be biologically relevant. These two genes make for an informative comparison since they both act to inhibit the BMP signaling pathway. BMP signaling plays an essential role in the formation of somatic, visceral, and cardiac components during early gastrulation and blocking BMP signaling during this phase leads to severe defects in mesodermal patterning. Sog is an extracellular inhibitor, which prevents BMP ligands from binding the receptor, whereas Brk acts within the nucleus as a transcriptional repressor. Because Sog protein can readily diffuse, the protein diffusing from the neuroectoderm accumulates in a graded pattern within the mesoderm (Srinivasan et al., 2002), even after its expression is extinguished in the mesoderm by Sna. However, Sog would not be expected to affect BMP signaling during gastrulation because it acts by inhibiting signaling by the heterodimeric Dpp:Scw ligand and Scw is only produced during blastoderm stages. In contrast, Brk acts cell-autonomously and its expression is absent from the mesoderm at both the mRNA and protein levels. Hence the ectopic presence of Brk in the mesoderm should produce defects similar to those resulting from global loss of BMP signaling. Thus, we reasoned that Sog misexpression in the mesoderm of gastrulating embryos should have little effect on patterning in that tissue in contrast to Brk misexpression, which should lead to gross defects. We tested these predictions by expressing *sog* versus *brk* in the mesoderm using a *twist*-GAL4 driver (Fig. 7). As expected, *sog* misexpression had little effect on specification of *eve*-positive cells of the visceral mesoderm, while comparable expression of *brk* severely reduced their numbers (Fig. 7). Thus, evolutionary selection on transient *sog* mRNA expression in the mesoderm might have been more relaxed than for *brk*. Alternatively, there may have been positive selection for Sog gene length relevant to its known patterning effects in the dorsal epidermis or to yet undiscovered functions it may play in mesodermal patterning.

Additional circumstantial evidence indicates that the Sna gradient might also be relevant to mesoderm invagination. For instance, expression patterns of genes involved in mesoderm invagination, such as *folded gastrulation*, are excluded from the region occupied by the graded portion of the Sna profile, whereas cells within the graded region undergo distinct morphological changes. Those changes in cell shape create an inflection in apical-basal versus basal-apical constriction, creating the Omega shape of invaginating epithelium (Martin et al., 2010). Whether additional components of this morphogenic system are expressed in patterns controlled directly or indirectly by the Sna gradient remains to be determined as this process is further dissected.

*Possible evolutionary considerations*

Non-coding DNA such as intergenic or intronic sequences tend to evolve more rapidly than coding sequences (Petrov and Hartl, 1998). Thus, altering gene length by lengthening or shortening intronic sequences should occur often during evolution (Petrov, 2002). If the type of timing bias we report is broadly relevant to other morphogen

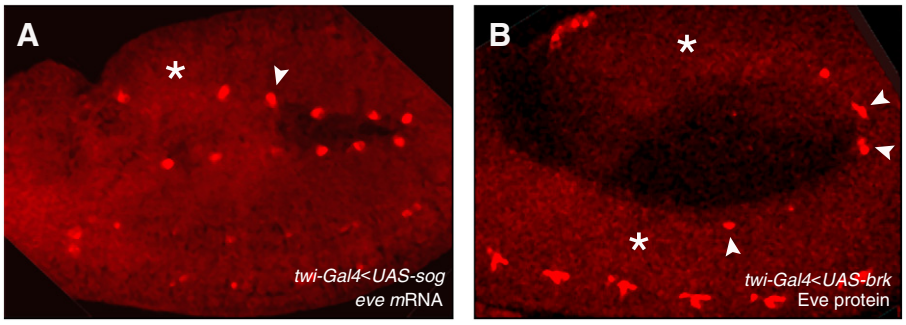


**Fig. 7.** Mis-expression of Sog versus Brk in the developing mesoderm has divergent effects on developing visceral cells. (A) in situ hybridization for *eve*, which labels precursor cells of the visceral mesoderm, in a *twi-Gal4<UAS-sog* E11 embryo. The majority of the visceral mesoderm precursor cells are present (arrowhead), with only a few exceptions (asterisk). (B) anti-Eve protein staining in a *twi-Gal4<UAS-brk* E11 embryo. Note that most *eve+* cells are missing (asterisks). In both cases, expression of *eve* in the ventral nerve cord is normal. Embryos are oriented with anterior to the left. Both images have been smoothened and amplified by a factor of 4 (eve mRNA) and 3 (eve protein).

patterning systems, one could imagine that gene-length would be a readily evolvable means for creating a prepattern. Thus, "one-step" insertions or deletions that change gene length may result in dramatic and immediate changes in transient spatial patterns of gene expression. Gene-length mechanisms may also fine-tune the enhancer sensitivities of target genes to morphogen gradients during evolution. An interesting evolutionary question to consider as additional genomes are sequenced and precisely annotated in various organisms will be whether there are correlations in gene length along A–R axes of various morphogen gradients in other species, as there tend to be in *Drosophila*.

## Author contributions

P.M., C.M.M., D.K., W.M., E.B., T.H. designed research; P.M., C.M.M., D.K., D.L.M, M.B., A.H., W.M., E.B., T.H. performed the research; P.M. contributed new analytic tools; P.M., C.M.M., D.K., W.M., E.B., T.H. analyzed the data; P.M., C.M.M., W.M., E.B., T.H. wrote the paper.

## Acknowledgments

We thank William Beaver, Ella Tour, Yoav Freund, and Michael Levine for valuable discussions and comments. Michael Levine provided the anti-Sna antibody, and fly stocks were obtained from the Bloomington Stock Center. P.M. thanks Arthur Lander for critical comments and suggestions, and the Center for Complex Biological Systems (UCI) where this work was completed. C.M.M. thanks Cheryl Hsia and William Beaver for advice with collection of confocal images. D.L.M. and M.B. were supported by start-up funds provided by the CWRU to C.M.M. This work was supported by grants from E.B. (NSF IBN 0120728 and IOS-0920785, NIH R01 NS29870) and W.M. (NIH grant R37HD28315). P.M. and T.H. acknowledge the support of the NSF through the Center for Theoretical Biological Physics (Grant PHY-0822283).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.ydbio.2011.08.016.

## Note added in proof

An assumption of our gene-length-dependent patterning model is that transcriptional repression in embryos occurs by preventing RNA polymerase from initiating new transcription, while most polymerases that are already elongating continue to the 3' end of the transcription unit. This assumption has been explicitly tested and found to be true for the *sog* gene by Bothma et al, "The Snail Repressor Inhibits Release, Not Elongation, of Paused Pol II in the Drosophila Embryo", Current Biology, (2011), doi:10.1016/j.cub.2011.08.019.

## References

Bardin, A.J., Schweisguth, F., 2006. Bearded family members inhibit neuralized-mediated endocytosis and signaling activity of delta in *Drosophila*. Dev. Cell 10, 245–255.
Barolo, S., Posakony, J.W., 2002. Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. Genes Dev. 16, 1167–1181.
Boettiger, A.N., Levine, M., 2009. Synchronous and stochastic patterns of gene activation in the *Drosophila* embryo. Science 325, 471–473.
Capovilla, M., Eldon, E., Pirrotta, V., 1992. The giant gene of *Drosophila* encodes a B-ZIP DNA-binding protein that regulates the expression of other segmentation gap genes. Development 114, 99–112.
Chitnis, A.B., 2006. Keeping single minded expression on the straight and narrow. Mol. Cell 21, 450–452.
Cowden, J., Levine, M., 2002. The snail repressor positions notch signaling in the *Drosophila* embryo. Development 129, 1785–1793.
Cowden, J., Levine, M., 2003. Ventral dominance governs sequential patterns of gene expression across the dorsal–ventral axis of the neuroectoderm in the *Drosophila* embryo. Dev. Biol. 262, 335–349.
Crocker, J., Tamori, Y., Erives, A., 2008. Evolution acts on enhancer organization to fine-tune gradient threshold readouts. PLoS Biol. 6, e263.
Dequeant, M.L., Pourquie, O., 2008. Segmental patterning of the vertebrate embryonic axis. Nat. Rev. Genet. 9, 370–382.
Eldon, E.D., Pirrotta, V., 1991. Interactions of the *Drosophila* gap gene giant with maternal and zygotic pattern-forming genes. Development 111, 367–378.
Fowlkes, C.C., Hendriks, C.L., Keranen, S.V., Weber, G.H., Rubel, O., Huang, M.Y., Chatoor, S., DePace, A.H., Simirenko, L., Henriquez, C., Beaton, A., Weiszmann, R., Celniker, S., Hamann, B., Knowles, D.W., Biggin, M.D., Eisen, M.B., Malik, J., 2008. A quantitative spatiotemporal atlas of gene expression in the *Drosophila* blastoderm. Cell 133, 364–374.
Francois, V., Solloway, M., O'Neill, J.W., Emery, J., 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.
Gray, S., Szymanski, P., Levine, M., 1994. Short-range repression permits multiple enhancers to function autonomously within a complex promoter. Genes Dev. 8, 1829–1838.
Hewitt, G.F., Strunk, B.S., Margulies, C., Priputin, T., Wang, X.D., Amey, R., Pabst, B.A., Kosman, D., Reinitz, J., Arnosti, D.N., 1999. Transcriptional repression by the *Drosophila* giant protein: cis element positioning provides an alternative means of interpreting an effector gradient. Development 126, 1201–1210.
Hulskamp, M., Pfeifle, C., Tautz, D., 1990. A morphogenetic gradient of hunchback protein organizes the expression of the gap genes Kruppel and knirps in the early *Drosophila* embryo. Nature 346, 577–580.
Hulskamp, M., Lukowitz, W., Beermann, A., Glaser, G., Tautz, D., 1994. Differential regulation of target genes by different alleles of the segmentation gene hunchback in *Drosophila*. Genetics 138, 125–134.
Ip, Y.T., Park, R.E., Kosman, D., Bier, E., Levine, M., 1992a. The dorsal gradient morphogen regulates stripes of rhomboid expression in the presumptive neuroectoderm of the *Drosophila* embryo. Genes Dev. 6, 1728–1739.
Ip, Y.T., Park, R.E., Kosman, D., Yazdanbakhsh, K., Levine, M., 1992b. Dorsal-twist interactions establish snail expression in the presumptive mesoderm of the *Drosophila* embryo. Genes Dev. 6, 1518–1530.
Jaeger, J., Surkova, S., Blagov, M., Janssens, H., Kosman, D., Kozlov, K.N., Myasnikova, E., Manu, Vanario-Alonso, C.E., Samsonova, M., Sharp, D.H., Reinitz, J., 2004. Dynamic control of positional information in the early *Drosophila* embryo. Nature 430, 368–371.
Jazwinska, A., Kirov, N., Wieschaus, E., Roth, S., Rushlow, C., 1999a. The *Drosophila* gene brinker reveals a novel mechanism of Dpp target gene regulation. Cell 96, 563–573.
Jazwinska, A., Rushlow, C., Roth, S., 1999b. The role of brinker in mediating the graded response to Dpp in early *Drosophila* embryos. Development 126, 3323–3334.
Jiang, J., Levine, M., 1993. Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. Cell 72, 741–752.

Jiang, J., Kosman, D., Ip, Y.T., Levine, M., 1991. The dorsal morphogen gradient regulates the mesoderm determinant twist in early *Drosophila* embryos. Genes Dev. 5, 1881–1891.
Kosman, D., Ip, Y.T., Levine, M., Arora, K., 1991. Establishment of the mesoderm-neuroectoderm boundary in the *Drosophila* embryo. Science 254, 118–122.
Kosman, D., Mizutani, C.M., Lemons, D., Cox, W.G., McGinnis, W., Bier, E., 2004. Multiplex detection of RNA expression in *Drosophila* embryos. Science 305, 846.
Kutejova, E., Briscoe, J., Kicheva, A., 2009. Temporal dynamics of patterning by morphogen gradients. Curr. Opin. Genet. Dev. 19, 315–322.
Lek, M., Dias, J.M., Marklund, U., Uhde, C.W., Kurdija, S., Lei, Q., Sussel, L., Rubenstein, J.L., Matise, M.P., Arnold, H.H., Jessell, T.M., Ericson, J., 2010. A homeodomain feedback circuit underlies step-function interpretation of a Shh morphogen gradient during ventral neural patterning. Development 137, 4051–4060.
Leptin, M., 1991. twist and snail as positive and negative regulators during *Drosophila* mesoderm development. Genes Dev. 5, 1568–1576.
Liem Jr., K.F., Tremml, G., Jessell, T.M., 1997. A role for the roof plate and its resident TGFbeta-related proteins in neuronal patterning in the dorsal spinal cord. Cell 91, 127–138.
MacArthur, S., Li, X.Y., Li, J., Brown, J.B., Chu, H.C., Zeng, L., Grondona, B.P., Hechmer, A., Simirenko, L., Keranen, S.V., Knowles, D.W., Stapleton, M., Bickel, P., Biggin, M.D., Eisen, M.B., 2009. Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. Genome Biol. 10, R80.
Martin, A.C., Gelbart, M., Fernandez-Gonzalez, R., Kaschube, M., Wieschaus, E.F., 2010. Integration of contractile forces during tissue invagination. J. Cell Biol. 188, 735–749.
McDonald, J.A., Holbrook, S., Isshiki, T., Weiss, J., Doe, C.Q., Mellerick, D.M., 1998. Dorsoventral patterning in the *Drosophila* central nervous system: the vnd homeobox gene specifies ventral column identity. Genes Dev. 12, 3603–3612.
Mellerick, D.M., Nirenberg, M., 1995. Dorsal–ventral patterning genes restrict NK-2 homeobox gene expression to the ventral half of the central nervous system of *Drosophila* embryos. Dev. Biol. 171, 306–316.
Morel, V., Schweisguth, F., 2000. Repression by suppressor of hairless and activation by Notch are required to define a single row of single-minded expressing cells in the *Drosophila* embryo. Genes Dev. 14, 377–388.
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.
Nahmad, M., Stathopoulos, A., 2009. Dynamic interpretation of hedgehog signaling in the *Drosophila* wing disc. PLoS Biol. 7, e1000202.
O'Brien, T., Lis, J.T., 1993. Rapid changes in *Drosophila* transcription after an instantaneous heat shock. Mol. Cell. Biol. 13, 3456–3463.
O'Farrell, P.H., 1992. Developmental biology. Big genes and little genes and deadlines for transcription. Nature 359, 366–367.
O'Neill, J.W., Bier, E., 1994. Double-label in situ hybridization using biotin and digoxigenin-tagged RNA probes. Biotechniques 17 (870), 874–875.
Ozbudak, E.M., Pourquie, O., 2008. The vertebrate segmentation clock: the tip of the iceberg. Curr. Opin. Genet. Dev. 18, 317–323.
Petrov, D.A., 2002. Mutational equilibrium model of genome size evolution. Theor. Popul. Biol. 61, 531–544.
Petrov, D.A., Hartl, D.L., 1998. High rate of DNA loss in the *Drosophila* melanogaster and *Drosophila* virilis species groups. Mol. Biol. Evol. 15, 293–302.
Rao, Y., Vaessin, H., Jan, L.Y., Jan, Y.N., 1991. Neuroectoderm in *Drosophila* embryos is dependent on the mesoderm for positioning but not for formation. Genes Dev. 5, 1577–1588.
Ray, R.P., Arora, K., Nusslein-Volhard, C., Gelbart, W.M., 1991. The control of cell fate along the dorsal–ventral axis of the *Drosophila* embryo. Development 113, 35–54.
Roth, S., Stein, D., Nusslein-Volhard, C., 1989. A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the *Drosophila* embryo. Cell 59, 1189–1202.
Rothe, M., Nauber, U., Jackle, H., 1989. Three hormone receptor-like *Drosophila* genes encode an identical DNA-binding finger. EMBO J. 8, 3087–3094.
Rothe, M., Pehl, M., Taubert, H., Jackle, H., 1992. Loss of gene function through rapid mitotic cycles in the *Drosophila* embryo. Nature 359, 156–159.
Rothe, M., Wimmer, E.A., Pankratz, M.J., Gonzalez-Gaitan, M., Jackle, H., 1994. Identical transacting factor requirement for knirps and knirps-related Gene expression in the anterior but not in the posterior region of the *Drosophila* embryo. Mech. Dev. 46, 169–181.
Rusch, J., Levine, M., 1996. Threshold responses to the dorsal regulatory gradient and the subdivision of primary tissue territories in the *Drosophila* embryo. Curr. Opin. Genet. Dev. 6, 416–423.
Rushlow, C.A., Han, K., Manley, J.L., Levine, M., 1989. The graded distribution of the Dorsal morphogen is initiated by selective nuclear transport in *Drosophila*. Cell 59, 1165–1177.
Schweisguth, F., 2004. Regulation of notch signaling activity. Curr. Biol. 14, R129–138.
Srinivasan, S., Rashka, K.E., Bier, E., 2002. Creation of a Sog morphogen gradient in the *Drosophila* embryo. Dev. Cell 2, 91–101.
Steward, R., 1989. Relocalization of the dorsal protein from the cytoplasm to the nucleus correlates with its function. Cell 59, 1179–1188.
Szymanski, P., Levine, M., 1995. Multiple modes of dorsal-bHLH transcriptional synergy in the *Drosophila* embryo. EMBO J. 14, 2229–2238.
Takashima, Y., Ohtsuka, T., Gonzalez, A., Miyachi, H., Kageyama, R., 2011. Intronic delay is essential for oscillatory expression in the segmentation clock. Proc. Natl. Acad. Sci. U. S. A. 108, 3300–3305.
Thisse, B., Stoetzel, C., Gorostiza-Thisse, C., Perrin-Schmitt, F., 1988. Sequence of the twist gene and nuclear localization of its protein in endomesodermal cells of early *Drosophila* embryos. EMBO J. 7, 2175–2183.
Thisse, C., Perrin-Schmitt, F., Stoetzel, C., Thisse, B., 1991. Sequence-specific transactivation of the *Drosophila* twist gene by the dorsal gene product. Cell 65, 1191–1201.
Wilkie, G.S., Shermoen, A.W., O'Farrell, P.H., Davis, I., 1999. Transcribed genes are localized according to chromosomal position within polarized *Drosophila* embryonic nuclei. Curr. Biol. 9, 1263–1266.
Yu, D., Small, S., 2008. Precise registration of gene expression boundaries by a repressive morphogen in *Drosophila*. Curr. Biol. 18, 868–876.
Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., Young, R.A., 2007a. RNA polymerase stalling at developmental control genes in the *Drosophila* melanogaster embryo. Nat. Genet. 39, 1512–1516.
Zeitlinger, J., Zinzen, R.P., Stark, A., Kellis, M., Zhang, H., Young, R.A., Levine, M., 2007b. 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.
Zinzen, R.P., Papatsenko, D., 2007. Enhancer responses to similarly distributed antagonistic gradients in development. PLoS Comput. Biol. 3, e84.
Zinzen, R.P., Senger, K., Levine, M., Papatsenko, D., 2006. Computational models for neurogenic gene expression in the *Drosophila* embryo. Curr. Biol. 16, 1358–1365.