

# Temporally Changing Roles of Morphogen Dorsal in the *Drosophila* Early Embryo

Thesis by  
Jihyun Irizarry

In Partial Fulfillment of the Requirements for  
the degree of  
Doctor of Philosophy

The logo for the California Institute of Technology (Caltech), featuring the word "Caltech" in a bold, orange, sans-serif font.

CALIFORNIA INSTITUTE OF TECHNOLOGY  
Pasadena, California

2020  
(Defended July 31, 2020)

© 2020

Jihyun Irizarry

ORCID: 0000-0003-3967-2288

## **Acknowledgement**

Many thanks to my advisor, Angela Stathopoulos. She is one of greatest mentors I have ever met. Without her support and guidance, this would not have been possible.

Special thanks to my committee, Marianne Bronner, Ellen Rothenberg, and Mitchell Guttman, for tremendous support and feedback.

And to my mother, Kumsun Choi, I am grateful for your love and support.

## Abstract

Morphogen gradients provide positional cues during development, with cell fate specification proceeding in a morphogen concentration-dependent manner during patterning. However, morphogens also are dynamic as their concentrations change not only in space but also in time, but how these dynamics are translated into cell fate specification over time is not well understood. To provide a better understanding of morphogens' temporal roles, we studied how *Drosophila* dorsal-ventral body patterning is controlled by the dynamic morphogen Dorsal (Dl). Dl is present in a nuclear-cytoplasmic gradient along the dorsal-ventral (DV) axis, but Dl levels also continuously increase between and within nuclear cell divisions associated with the early syncytial embryo. To experimentally manipulate Dl levels in time in order to determine whether these dynamics are important, we developed a light-activated degradation system. The blue light inducible degron domain, BLID, was fused to the C-terminus of Dl by genomic editing using CRISPR-Cas9. To assay effects on temporally manipulated Dl levels, we combined this light-inducible degradation system with the MS2-MCP.GFP nascent transcript imaging system, and used to monitor transcription changes *in vivo* at the *snail* (*sna*) locus, a gene requiring high Dl levels. We found that while high Dl levels are required for *sna* activation at early nuclear cycle 14, late expression can be supported even if Dl levels are extinguished. Twist, an early Dl target gene, is later auto-activating and can support the later *sna* expression without Dl. Surprisingly, we found that peak levels of Dl, present at late nuclear cycle 14, are required only to fine tune, in particular to decrease, *sna* levels. This differential action of Dl, first functioning as an activator and next as a damper of expression, is manifest by the coordinate action of two enhancers acting at the *sna* locus. Here, we highlight how morphogen roles change in time, and suggest that this may be a general characteristic of dynamic morphogens that allows them to control developmental patterning.

# Preface

## *Insight into the temporally changing role of the morphogen Dorsal*

Morphogen input is likely most critical at the earliest stage of the patterning process. Since morphogens can be transcription factors that act early in the *Drosophila* patterning process, we hypothesized that they open chromatin, essentially priming the genome for zygotic transcription, in order to initiate the transcriptional activation of target genes. Furthermore, positional cues based on morphogen concentration also dictate the spatial domain of early target gene expression patterns. However, these patterns are refined overtime due to a second wave of input from additional transcription factors. Might it be then that morphogens become dispensable after providing their initial input? Or are morphogens continuously required, possibly supporting multiple functions and changing roles, throughout development?

During my graduate study, I investigated the temporal action of one morphogen acting in early *Drosophila* embryos, the Dorsal (Dl) transcription factor. Dl levels continuously increase even after the expression domains of target genes are specified, and no longer change. These data suggest the possibility that morphogens support additional roles later in the patterning process that do not relate to spatial patterning. However, it is also known that another morphogen transcription factor, Bicoid (Bcd), is required continuously to support particular target genes. Is the action of morphogens gene specific and varied? Could Dl and Bcd exhibit different mechanisms of action? We set out to provide insight into Dorsal's action over time by studying multiple cis-regulatory modules associated with target genes, to determine if they each reach Dl-input and how they coordinate, as well as developing an optogenetic approach to control Dl activity over time.

This thesis is composed of 4 chapters. In the first, we compare and contrast how Dl and Bcd transcription factors regulate patterning in the early embryo, taking into consideration both levels of these morphogens in space and time in the support of target gene expression. As Dl gradients are dynamic, this raised the possibilities that Dl may have temporally changing roles

throughout the blastoderm embryo. In chapter 2, to dissect temporal roles of Dl during the DV patterning, we discuss our deployment of a blue light inducible degron system (BLID) to control Dl activity in time through fusion of the BLID domain to the C-terminus of Dl protein using recently developed CRISPR-Cas9 genome editing methods. This allowed us to temporally manipulate Dl levels. We demonstrate endogenous Dl levels are required to support presumptive mesodermal genes at the early stage of the DV patterning process. In contrast, later, the presumptive mesodermal genes become independent of high Dl levels with support from *Twi*, which functions as a molecular ratchet.

Thereby, Dl essentially stops functioning as a morphogen, because it no longer provides positional cues at the late stage of the DV patterning process. However, Dl levels continue to increase, even at the late stage. In chapter 3, we identify a previously unappreciated function of Dl during this late DV patterning stage that does not relate to regulation of spatial pattern. At the late stage, Dl functions, through one particular enhancer, the *sna* proximal enhancer, to decrease levels of *sna* total expression output. Through this analysis, we elucidate that Dl's role changes over time.

Finally, in chapter 4, we discuss the implication of our studies and future directions.

## PUBLISHED CONTENT AND CONTRIBUTIONS

Irizarry, J., McGehee, J., Kim, G., Stein, D. and Stathopoulos, A. (2020). Twist-dependent ratchet functioning downstream from Dorsal revealed using a light-inducible degenon. *Genes Dev.* 34, 965–972. doi: 10.1101/gad.338194.120.

J.I., A.S., D.S., and J.M. conceived the project and planned the experimental approach. J.I. performed all imaging. J.M. performed all quantitative analysis of the imaging data, Crispr/Cas9 genomic engineering, Western analysis, and viability studies. J.I. and J.M. performed staining. D.S. and G.K. validated the use of BLID for studies in *Drosophila* and helped to develop protocols for its use. J.I., J.M., and D.S. developed the protocol for fixed embryo analysis, whereas J.I. and J.M. developed protocols for BLID live imaging. Data were analyzed by J.I., J.M., and A.S. The manuscript was written by J.I., J.M., and A.S. with edits provided by D.S.

Irizarry, J. and Stathopoulos A. (2020) “Dynamic Patterning by Morphogens Illuminated by Cis-Regulatory Studies”. In submission.

J.I., and A.S. wrote the manuscript.

Irizarry, J. and Stathopoulos, A. (2015). FGF signaling supports *Drosophila* fertility by regulating development of ovarian muscle tissues. *Dev. Biol.* 404, 1–13. doi: 10.1016/j.ydbio.2015.04.023.

J.I. designed and carried out experiments and analyzed all data. J. I. and A. S wrote the manuscript.

Sandler, J. E., Irizarry, J., Stepanik, V., Dunipace, L., Amrhein, H. and Stathopoulos, A. (2018). A Developmental Program Truncates Long Transcripts to Temporally Regulate Cell Signaling. *Dev. Cell* 47, 773–784.e6. doi: 10.1016/j.devcel.2018.11.019.

J. E. S., J. I., V. S., and L.D. carried out the experiments. J.E.S. analyzed data with help from H. A. and wrote the manuscript with A. S. and input from others.

# Table of Contents

<b>Chapter 1. Introduction: Dynamic Patterning by Morphogens Illuminated by Cis-Regulatory Studies</b>	<b>1</b>
Summary	2
Introduction	2
Discussion	4
Conclusion	17
References	20
Figures and Legends	27
<b>Chapter 2. Twist-dependent ratchet functioning downstream of Dorsal revealed using a light-inducible degron</b>	<b>33</b>
Abstract	34
Introduction	34
Results	35
Discussion	44
Materials & Methods	46
References	54
Figures & Legends	57
Supplementary Figures & Legends	65
<b>Chapter 3. The highest levels of transcription factor Dorsal dampen, not promote, gene expression by regulating enhancer action</b>	<b>79</b>
Abstract	80
Results & Discussion	80
Materials & Methods	87
References	93
Figures & Legends	95
Supplementary figures & Legends	102
<b>Chapter 4. Discussion</b>	<b>105</b>
Discussion	106
References	112

**Appendices**

- A. FGF signaling regulates cell proliferation and differentiation during morphogenesis of the *Drosophila* ovary to support fertility** 113
- B. A developmental program truncates long transcripts to temporally regulate cell signaling** 155

# Chapter 1

## Introduction

### Dynamic Patterning by Morphogens Illuminated by Cis-Regulatory Studies

This chapter was written by Jihyun Irizarry and Angelike Stathopoulos and is under review.

## SUMMARY

Morphogens are dynamic as their concentrations change not only in space but also in time during development. However, how morphogen dynamics are translated over time into cell fate specification is not well understood. In this review, we focus on two morphogens, the maternal transcription factors Bicoid and Dorsal that control patterning of *Drosophila* embryos. We discuss how gradients of these morphogens are formed, and how the cis-regulatory logic inherent in target gene enhancer sequences interprets their input to support concentration-dependent outputs. Lastly, we discuss how cis-regulatory mechanisms ensure precise expression outputs and how recent live imaging studies have provided insight.

## INTRODUCTION

Morphogens are proteins that exist in a graded manner, and the gradients provide positional cues for cell fate specification during development (Rogers and Schier, 2011; Wolpert, 1996). In general, morphogens are produced from localized sources and diffuse in the extracellular space toward neighboring cells. Once the gradients are formed in this way, the receiving cells interpret their respective positional information by sensing the concentration of morphogens. In response, cells activate different sets of target genes, initiating the patterning processes and ultimately defining cell fate. This phenomenon is observed for a number of morphogens acting in various organisms, such as Sonic hedgehog (Shh) in chick neural tube patterning (Briscoe et al., 2001), Wingless (Wg) in *Drosophila* appendage development (Neumann and Cohen, 1997), Squint in zebrafish germ layer patterning (Chen and Schier, 2001), and Activin in *Xenopus* mesoderm and ectoderm induction (McDowell and Gurdon, 1999). Due to the importance of morphogens for patterning, dissecting the molecular mechanisms by which they control gene activation has been an intensively studied field for decades.

One of the core questions investigated in regards to morphogens is the mechanism by which their concentrations are translated into gene expression outputs. One of the earliest models of patterning proposed that differences in target gene spatial domains relates directly to morphogen concentration (Wolpert, 1969) (Fig. 1A). In this threshold-dependent model, target gene transcription is only initiated when the morphogen concentration is present above a certain level, whereas transcription is not initiated in domains where the morphogen concentration is present below this level. In this manner, gradients of morphogens regulate patterning by differentially controlling gene expression in response to the concentration change present in space.

However, recent studies have shed light on the importance of time, both duration and timing of exposure to morphogens, on target gene expression (rev. in Sagner and Briscoe, 2017). Target gene expression is influenced by morphogen input as well as the current state of the receiving cells in terms of their gene expression program and, because the morphogen can cause a change in gene expression, this then influences the state of cells so that further morphogen-dependent responses are different. For example, during dorsal-ventral (DV) patterning in zebrafish embryos during gastrulation, BMP signaling controls cell fate specification in a temporally progressive manner. BMP signaling at an early stage during gastrulation supports patterning of anterior ventrolateral domains; whereas signaling at a later stage promotes patterning of posterior ventrolateral domains (Tucker et al., 2008). The highly context-specific response highlights the importance of understanding morphogens' changing roles over time. Despite accumulating data supporting temporal roles of morphogens in development, it is unclear whether morphogen dynamics relate to their direct action or are a function (indirect response) of the underlying gene regulatory networks.

In the following sections, we discuss how gradients of the transcription factors (TFs) Bicoid (Bcd) and Dorsal (Dl) are formed, and how these gradients control differential target gene

expression along the AP and DV axes of *Drosophila* embryos, respectively. In this *Drosophila* system, the input from morphogens to target gene expression has been dissected at the cis-regulatory level, more so perhaps than any other developmental system, and provides a rich platform for understanding how these morphogen inputs are translated into spatially distinct target gene expression patterns.

## **Discussion**

### **I. *Drosophila* morphogens Bicoid and Dorsal: gradient formation and dynamics**

Upon fertilization, *Drosophila* embryos undergo thirteen rounds of rapid cell divisions, resulting in 14 nuclear cycles. These divisions occur in a syncytium in which there are no cellular membranes. During the first nine nuclear cycles, each cycle lasts about eight minutes. Starting at nuclear cycle (nc) 10, the duration of nuclear cycles lengthens, and widespread zygotic gene expression is initiated. In this earliest developmental process, body patterning along the anteroposterior (AP) and dorsal-ventral (DV) axes occurs (rev. in Stathopoulos and Newcomb, 2020). Two maternally deposited morphogens, Bicoid (Bcd) and Dorsal (Dl), are key factors that orchestrate this patterning process.

*Bicoid: a morphogen that controls patterning along the AP axis*

*bcd* mRNA contains localization sequences that result in concentration of its transcripts at the anterior pole of *Drosophila* embryos, which upon fertilization are translated resulting in a corresponding gradient of protein along the AP axis (Fig. 1B) (Driever and Nüsslein-Volhard, 1988). There is some debate in the field regarding how this gradient is established, whether it results from local translation of a mRNA gradient or diffusion of protein from a tightly localized mRNA source (e.g. Little et al., 2011). In either case, once the gradient forms, Bcd, a homeodomain (HD) transcription factor, binds to specific DNA sequence motifs, and differentially

activates genes along the AP axis. Bcd activates more than 40 target genes to initiate AP patterning in an apparently concentration-dependent manner but is also known to coordinate with other factors in supporting target gene expression (rev. in Briscoe and Small, 2015) .

While most studies of Bcd target gene expression are focused on nc14, nuclear Bcd is observed as early as nc 6 and exhibits dynamics (rev. in Huang and Saunders, 2020; Little et al., 2011). Bcd levels build fast within nuclei at the onset of each nuclear cycle (Fig. 1B). Furthermore, peak levels slightly increase from one nuclear cycle to the next nuclear cycle but, compared to the Dl dynamics discussed below, Bcd changes are relatively small when measured from the perspective of a single nucleus present at the anterior of embryos (Fig. 1B) (Gregor et al., 2007; rev. in Sandler et al., 2018). In addition, studies of Bcd input to particular target genes such as *hunchback* (*hb*) have suggested that Bcd becomes dispensable by late nc14. At this point, the expression boundaries of many targets, including *hb*, *knirps* (*kni*), and *Krüppel* (*Kr*), are supported by cross-regulation of gap genes that function as repressors (Jaeger et al., 2004; Liu et al., 2013; Manu et al., 2009). These particular studies suggested that early nc14 is the pivotal time point for Bcd to provide positional information towards patterning (Liu et al., 2013). However, another recent study showed that removal of Bcd at any developmental time point between nc10 and nc14 causes general developmental defects suggesting other Bcd target genes require input earlier (Huang et al., 2017). It is likely that the role of Bcd in supporting gene expression is target gene-specific and that there is not one particular timepoint in which Bcd acts but rather it is required at different times for distinct target genes (rev. in Huang and Saunders, 2020).

*Dorsal: a morphogen that controls patterning along the dorsal-ventral (DV) axis*

*dl* mRNA, like *bcd*, is maternally deposited but, unlike *bcd*, *dl* mRNA and protein are ubiquitously distributed in embryos (Anderson and Nüsslein-Volhard, 1984). Dl cannot enter the nucleus constitutively, however, and the controlled nuclear-localization of Dl protein forms a

gradient that functions as a morphogen to support DV axis patterning (Roth et al., 1989; Rushlow et al., 1989; Steward, 1989). Toll signaling pathway activation regulates nuclear translocation of Dl protein. In the absence of Toll signaling pathway activation, Dl is bound and sequestered in the cytoplasm by the I $\kappa$ B homolog, Cactus (Belvin and Anderson, 1996). Upon activation of the Toll receptor, a downstream intracellular cascade occurs that ultimately leads to phosphorylation of Cactus, which targets this protein for degradation. Concomitantly, Dl is freed and is imported into nuclei, where it activates over 50 genes (Belvin et al., 1995; Biemar et al., 2006; Stathopoulos et al., 2002). Toll receptor is ubiquitously expressed throughout embryos; however, its ligand Spätzle is activated in a graded manner along the DV axis within the extracellular space, such that highest activity Toll activation occurs in ventral regions (rev. in Reeves and Stathopoulos, 2009). The result of Toll signaling is a nuclear-cytoplasmic Dl gradient exhibiting the highest levels of nuclear Dl in ventral regions and progressively lower levels dorsally (Fig. 1B).

In contrast to the fast nuclear import displayed by Bcd at the onset of each nuclear cycle, levels of nuclear Dl build slowly throughout the nuclear cycle. This difference likely relates to Toll-dependent signaling being required for Dl to be imported into nuclei, whereas alternatively Bcd is free to enter nuclei once they reform after division. Furthermore, nuclear Dl levels build up within each nuclear cycle, plummet upon nuclear envelope breakdown at division, and slowly are reestablished upon nuclei reformation following division (Reeves et al., 2012). Furthermore, Dl levels build between nuclear cycles, never reaching a steady state; such that from the perspective of a nucleus located in ventral regions input exhibits a saw-tooth trend (Fig. 1B).

## **II. *Cis*-regulatory interpretation of morphogen concentration and importance of additional inputs**

The dynamics associated with these morphogens, Bcd and Dl, suggest that both provide not only positional cues but also impart temporal information toward target gene expression during

the *Drosophila* body patterning processes. However, the differences in their dynamics suggest that these morphogens may use different mechanisms to regulate target gene expression. Enhancer sequences provide information about the cis-regulatory logic of inputs that function to support gene expression outputs, and may provide insight into these morphogens' mechanisms of action.

*Interactions between morphogen transcription factors and pioneer factors initiate gene activation during embryonic patterning*

Enhancers are short regions of DNA that control transcription of genes, presumably, by direct contact with the gene promoter (rev. in Furlong and Levine, 2018). In general, enhancers are cis-regulatory DNA sequences of a few hundred base pairs long (average ~500 bp), and include short specific recognition sites that mediate binding of TFs. Enhancer activity is context-dependent, and perturbing their action can lead to developmental defects and disease (Bhatia et al., 2013; Lupiáñez et al., 2015; Uslu et al., 2014).

During development of early *Drosophila* embryos, graded TFs serve as morphogens to support the cell fate specification process by controlling enhancer activity. As an initial step towards the activation of gene expression, enhancer regions need to become accessible, free of nucleosomes, to enable binding of TFs. During early embryonic development, this process is supported by a particular class of TFs, pioneer factors, which are the first DNA-binding factors to engage target sites within chromatin. A pioneer factor can bind to target enhancers in closed chromatin and facilitate chromatin remodeling processes, which allows other TFs also to bind to enhancers to control target gene activities (Zaret and Carroll, 2011). At the early blastoderm stage, maternally deposited Zelda (Zld) plays a crucial role as a pioneer factor to support local depletion of nucleosomes with the effect being dependent on the number and position of Zld motifs (Li et al., 2014; Schulz et al., 2015; Sun et al., 2015). In this manner, Zld increases chromatin accessibilities at its target gene enhancers. Furthermore, Zld functions as a global activator to

initiate zygotic gene expression, including early body axis patterning genes in embryos (Harrison et al., 2011; Nien et al., 2011).

Zld associates with enhancer sequences for patterning genes that, subsequently, also are bound by Bcd and Dl transcription factors. A recent study examined occupancy of Zld and Dl to DNA throughout the entire genome with fine temporal resolution, and revealed that Zld binding can be detected as early as nc8, at which point little to no Dl binding is observed (Li and Eisen). These results suggest that Zld input precedes that of Dl and possibly potentiates Dl binding at target gene enhancer regions. In *zld* mutants, chromatin accessibility at enhancers decreases, and, subsequently, Bcd or Dl binding is significantly reduced (Li and Eisen; Xu et al., 2014). Thus, Bcd or Dl dependent enhancers lose their activity when *zld* is reduced. Similarly, when the number of Zld binding sites in reporter constructs of enhancer regions of Bcd or Dl target genes is reduced, the timing of enhancer activation is delayed and the enhancer expression domain shifts. In contrast, addition of more Zld binding sites to enhancers can result in precocious expression (Bosch and t. Bosch, 2006; Foo et al., 2014). Furthermore, introducing Zld binding sites into an inactive enhancer can convert it into a morphogen-dependent responsive enhancer (Foo et al., 2014; Xu et al., 2014). Taken together, Zld primes the genome to potentiate morphogens to activate target gene enhancers in space but also influences gene expression timing.

However, it has been proposed that other pioneer-like factors also contribute to patterning in the early embryo. For example, Odd-paired (Opa) works coordinately with Zld at first and then takes over to support gene expression at the mid-blastula transition (MBT) and beyond when Zld's role is likely minimal (Koromila et al., 2020). Furthermore, Bcd and Dl morphogens, as transcription factors that bind DNA, may also contribute to the chromatin remodeling process directly to support differential gene expression.

While no studies, to date, have demonstrated whether Dl can directly regulate chromatin accessibility, a recent study using ATAC-seq analyzed the ability of Bcd to support chromatin accessibility (Hannon et al., 2017). Along the AP axis, chromatin accessibility associated with particular Bcd target genes is sensitive to levels of this morphogen (Hannon et al., 2017). When levels of Bcd are artificially increased, a subset of enhancers associated with Bcd target genes which are sensitive to Bcd levels exhibit increases in chromatin accessibility. These results suggest that high levels of Bcd promote remodeling of chromatin structure (Fig. 3B).

*Relationship of affinity of binding sites for Dl and Bcd within target gene enhancers to threshold-outputs*

Once enhancer regions of target genes become accessible, morphogens also play a role in promoting activation of transcription to support the patterning process. As discussed above, in the threshold-dependent model, morphogen-concentration provides positional information to drive cells to differentiate into distinct cell fates by activating domain-specific genes. In relation to *Drosophila* DV patterning, three Dl-dependent thresholds have been characterized: high, intermediate, and low (Fig. 2A) that correspondingly relate to three categories of genes: Type I, II and III expressed in distinct domains (Stathopoulos and Levine, 2002a; Stathopoulos et al., 2002). First, Type I genes are expressed in the ventral region of the embryo, where the highest levels of Dl are established through Toll signaling. In this region presumptive mesodermal target genes, including *snail (sna)* and *twist (twi)*, are activated. Next, Type II genes are expressed in the ventrolateral region of the embryo. In this region intermediate levels of Dl activate a different set of target genes, including genes *ventral nervous system defective (vnd)* and *vein (vn)*, to specify the presumptive neurogenic ectoderm. Finally, Type III genes are found in more dorsal regions and include two types of gene patterns: genes activated by low levels of Dl that are present in lateral regions [e.g. *short gastrulation (sog)*] and genes repressed by these same low levels that

exhibit expression limited to dorsal regions [e.g. *decapentaplegic (dpp)*]. These genes support neurogenic ectoderm and dorsal ectoderm fates, respectively (Fig. 2A) (Jiang and Levine, 1993; Liberman and Stathopoulos, 2009). Furthermore, mutant backgrounds that contain either high, intermediate or low levels of Dl exhibit broad expression of either Type I, II, or III genes, respectively, and were useful for the identification of 50+ Dl target genes and miRNAs using gene expression profiling (Biemar et al., 2006; Stathopoulos et al., 2002).

One of the proposed molecular models to explain threshold responses is that Dl binding site affinity sets the Dl level required for enhancer activation (Papatsenko and Levine, 2005; Reeves and Stathopoulos, 2009). In general, ventrally-expressed genes (i.e. Type I) have enhancers with low affinity Dl binding sites, so their activation requires high Dl levels. In contrast, ventrolaterally-expressed genes (i.e. Type II) generally have enhancers with high affinity Dl binding sites, so their activation is possible at lower levels of Dl (Fig. 3A). Additionally, recent studies found that these boundaries of gene expression along the DV axis also shift in time, presumably because of a real-time response to dynamic morphogen input (Reeves et al., 2012). For example, the *sna* gene is a canonical Type I Dl response gene as it requires high Dl levels for activation, and it is expressed in the ventral region of embryos. *sna* expression is detectable at nc13, at which point it can be deduced that the Dl concentration has risen above the threshold required to support its activation. Subsequently, during early nc14, *sna* expression dorsally expands as Dl levels continue to increase (Reeves et al., 2012), indicating that the spatial region receiving the threshold level of the Dl necessary to support its expression changes over time.

However, studies have shown limitations of the threshold model. For instance, *twist (twi)* is an early Dl target gene that encodes a bHLH transcription factor, which functions in a presumed feed-forward mechanism along with Dl to support genes in ventral regions (Kosman et al., 1991). Many of the enhancers associated with genes expressed along the DV axis are bound by both

Dorsal and Twist (Zeitlinger et al., 2007). When high levels of *Tw* are ectopically expressed, the Dl-threshold responses are spatially reversed (Stathopoulos and Levine, 2002b). This result suggests that while Dl threshold responses exist, they also likely receive input from other factors and, in particular, can be influenced by *Tw* levels. Furthermore, as many of Dl's target genes are coregulated by multiple enhancers, the affinity of Dl binding sites to one enhancer may not be a good predictor of the role of Dl in the context of multi-enhancer cis-regulatory systems, discussed below.

Similar to the DV patterning target genes, a set of AP target genes directly respond to Bcd in a concentration-dependent manner. In embryos in which the Bcd gradient was flattened using particular genetic backgrounds so that the levels across the embryo are uniform, *hb*, *Kr*, and *giant* (*gt*) are expressed in an on/off fashion (Fig. 2B) (Ochoa-Espinosa et al., 2009). Like ubiquitous expression of Dl target genes in mutant backgrounds that contain one level of Dl (e.g. Stathopoulos et al., 2002), genes *hb* and *gt* are responsive to particular levels of Bcd and are expressed broadly in embryos that contain low or intermediate levels of Bcd throughout, respectively (Ochoa-Espinosa et al., 2009). In contrast, the genes *orthodenticle* (*otd*), *empty spiracles* (*ems*), and *buttonhead* (*btd*) are not ubiquitously expressed in embryos with flattened Bcd gradients but instead these particular genes only exhibit shifts in the positions of their posterior boundaries of expression (e.g. Fig. 2B) (Ochoa-Espinosa et al., 2009). This result suggests that *otd*, *ems*, and *btd* boundaries are positioned through other mechanisms, not simply in response to Bcd levels. However, a more recent study suggested flattening of the Bcd gradient using mutants does not completely flatten Bcd levels and that a shallow gradient is still present. This other group using a transgenic approach to flattening the Bcd gradient, found that some target genes (e.g. *btd*) respond with all/none responses to Bcd levels whereas others exhibit a shift [e.g. *kni*] (Hannon et al., 2017). Taken together, the results suggest that, similar to the Dl target genes, some target genes can be

activated and positioned in a Bcd concentration-dependent manner and others are not; it is certain that additional factors also play roles during the patterning process along the AP axis.

In contrast to Dl target enhancers, an analysis of Bcd target gene enhancers found little correlation between the affinity of binding sites within target gene enhancers and the expression domains of many AP patterning genes arguing against a simple Bcd threshold-response that relates to binding site affinity (Briscoe and Small, 2015; Ochoa-Espinosa et al., 2005; Segal et al., 2008). Surprisingly, however, chromatin accessibility of Bcd target genes is sensitive to Bcd concentration and suggests that gene expression may be threshold-dependent even if the affinity of particular binding sites does not always provide a clear correlation (Fig. 3B) (Hannon et al., 2017). It is possible that not only the binding site affinity but also the length of exposure to a morphogen can influence threshold outputs.

*Combinatorial regulation of multiple factors in the establishment of correct boundaries along to AP and DV axis*

While Dl is a pivotal input to target genes expressed in ventral regions along the DV axis, these genes also receive input from an early Dl target gene, Twi (Ozdemir et al., 2011; Sandmann et al., 2007), as noted above. During DV axis patterning, boundaries of ventral target genes are supported by synergistic interactions between Dl and Twi (Szymanski and Levine, 1995). For instance, a 57 bp sequence within the *twi* proximal element (*twi*\_PE) enhancer, which is located adjacent to the promoter, has two weak Dl binding sites and drives expression within 12-14 ventral-most cells (Jiang and Levine, 1993). The dorsal boundaries of the expression supported by this element are expanded upon adding two Twi binding sites (E-box sequences), widening to encompass a domain of 20 cells in width (Fig. 3C).

In addition to synergistic input by morphogens and other activators, input from repressors also contributes to the spatial limits of target gene expression. The Dl gradient establishes the

initial expression pattern, but the early domains of expression are, in general, broader than the final patterns. Precise positioning of boundaries of DV genes require additional repressive inputs. In the case of *sna*, the transcription factor Suppressor of Hairless [Su(H)] sets the dorsal boundary through its binding at the *sna* distal enhancer (Fig. 3E) (Ozdemir et al., 2014; Schweisguth and Posakony, 1992). When the Su(H) binding sites are mutated in this enhancer, expression is expanded dorsally. Furthermore, in Su(H) mutants, many DV genes exhibit expanded expression dorsally (Ozdemir et al., 2014). Thus, this result suggests that Su(H) is a broadly-acting repressor that acts to set the dorsal boundaries of multiple DV genes.

Similarly, boundaries of AP patterning genes are set by combinatorial input from activators, including the Bcd morphogen and Zld pioneer factor, as well as repressors. Runt (Run) is one of the major repressors targeting AP patterning genes (Fig. 3D) (Chen et al., 2012). Run is expressed in the embryonic trunk and excluded from the poles, and limits Bcd-dependent activation (Chen et al., 2012; Gergen and Butler, 1988). In *run* mutants, the head gap gene *otd* posteriorly expands. Similarly, enhancers of *ems*, *sloppy paired 1 (slp1)*, and *sloppy paired 2 (slp2)* also exhibit posterior shift in the *run* mutant. At these enhancer regions, high affinity Run binding sites are enriched, and the mutation of these binding sites is associated with a posterior shift of the patterns. Furthermore, the posterior shifts become more severe when the enhancers are expressed in embryos that are mutant for multiple repressors including Run and also the maternal protein Capicua (Cic) and gap protein Kr (Fig. 3D) (Chen et al., 2012). In this manner, posterior boundaries of a subset of genes expressed along the AP axis are regulated spatially by these repressors. In addition, Run being broadly expressed throughout the trunk can also act to influence the timing of enhancer action; possibly acting as a counterbalance to pioneer activators such as Zld (Koromila and Stathopoulos, 2017).

Taken together, morphogens can activate target gene enhancers in a concentration-dependent manner, but input from morphogens alone is not sufficient to define target gene expression responses. Additional input from activators and/or repressors on the enhancer is key, and this combinatorial regulation by multiple factors modifies the concentration-dependent responses and sets the final expression domain of the target genes.

### **III. Morphogen inputs to multiple co-acting enhancers can vary**

The *Drosophila* embryo has been instrumental as a model system in demonstrating that gene expression patterns are supported through coordinate action of enhancer sequences. Some cis-regulatory systems are composed of multiple, distinct enhancer that function in an apparently additive manner to generate gene expression patterns. For example, five enhancer sequences combine to support the expression of seven stripes of expression for the gene *even skipped* (*eve*) (Fig. 4B; e.g. Fujioka et al., 1999; Small et al., 1992). Even though these enhancers support predominantly different patterns, surprisingly, they also can interact. For instance, deletion of the *eve* stripe 1 enhancer leads to precocious and expanded expression of *eve* stripe 2 (Lim et al., 2018) possibly due to sharing of repressor input.

In contrast, it also has become clear that some genes are regulated by multiple enhancers that support similar spatiotemporal activity (Barolo, 2012; Hong et al., 2008; McGregor et al., 2007). This is true even in the short period of time before embryos undergo gastrulation. Due to the similarity in expression output, the first enhancer identified at a locus was called the “primary enhancer” and the subsequent enhancers which exhibited similar spatio-temporal expression were named “shadow enhancers”. It was initially thought that the functions of *Drosophila* shadow enhancers, typically located at a distance, are redundant to counterpart primary enhancers as commonly the case in higher animals (e.g. Xiong et al., 2002). More recently, it has been suggested that the *Drosophila* shadow enhancers acting to support patterning in embryos have similar but

not identical functions compared to the associated primary enhancers (Dunipace et al., 2011; El-Sherif and Levine, 2016; Perry et al., 2011).

That the expression of many genes results from coordinate action of multiple enhancers raises the question whether morphogen input is required to modulate action of one or several enhancers in the context of cis-regulatory systems. Studies of Bcd-input to *hb* gene expression have suggested that Bcd provides input to early-acting enhancers and, subsequently, other Bcd-independent late-acting enhancers take over (Liu and Ma, 2013). It was suggested that this “hand-off” allows a window of opportunity for *hb* gene expression to benefit from the Bcd-gradient input and process gradient properties, including information pertaining to scaling. How this processing is accomplished is still unclear but may relate to regulation of sequential action of enhancers by promoter proximal regions (Dunipace et al., 2013).

#### *Dominant repression of multiple enhancers promotes dynamic shifts during patterning*

During patterning, accumulating evidence suggests that refinement of the expression domain is predominantly controlled by distal (“shadow”) enhancers (Dunipace et al., 2011; Perry et al., 2011). At the *sna* locus, the proximal enhancer supports a slightly expanded domain, whereas the distal enhancer supports a domain similar to that supported by the full *sna* locus (Dunipace et al., 2011). This is because the distal enhancer is responsive to input from repressors. Furthermore, the distal enhancer limits the proximal enhancer activity in the region where the gene is repressed. In this manner, the distal enhancer can dominantly-affect the final expression domain of the gene (Fig. 4A). It is possible that distal enhancers are associated, more generally, with dominant repression inputs and this allows genes to be silenced conditionally only when the enhancer is active.

Non-additive expression output is not only a feature of genes expressed along the DV axis, but is also a critical mode of regulation in genes expressed along the AP axis. For instance, the proximal and distal enhancers at the *hb* locus share similar spatial and temporal expression. Both are expressed at the anterior region in early embryos. However, unlike the proximal enhancer, the distal enhancer expression does not support expression at the anterior pole. At the anterior pole, Torso signaling represses *hb* expression, and only the distal enhancer is responsive to input from Torso signaling. To set the *hb* anterior boundary, the distal enhancer interferes with the proximal enhancer activity at the anterior pole region, resulting in repressed *hb* expression in the anterior pole (Fig. 4A) (Perry et al., 2011). Furthermore, non-additive effects are also observed in the cis-regulatory systems associated with other genes expressed along the AP axis including *kni* and *Kr* (El-Sherif and Levine, 2016). In this way, coordinate input from multiple enhancers is also necessary to set correct boundaries for a number of AP genes even in the cases where individual enhancer outputs appear similar (Barolo and Levine, 1997; El-Sherif and Levine, 2016; Perry et al., 2011).

In addition, enhancer interactions not only help to refine spatial patterns but also can function to modulate levels of gene expression. Unlike genes in which the levels are supported by the sum of individual enhancer activities, balanced input from two enhancers is necessary to regulate expression of genes *kni* and *sna* (Bothma et al., 2015). In the case of *kni*, two enhancers act in an additive way (Fig. 4B). Whereas in the case of *sna*, both distal (*sna\_dis*) and proximal (*sna\_prox*) enhancers act in a non-additive (sub-additive) manner to support proper levels of endogenous *sna* gene expression (Fig. 4C) (Bothma et al., 2015; Dunipace et al., 2011). Each enhancer is responsive to D1 levels at the early stages of DV patterning (Irizarry et al., 2020). However, these two enhancers support *sna* with different strengths: the distal enhancer drives high expression levels (“strong enhancer”), whereas the proximal enhancer supports low expression

levels (“weak enhancer”) (Bothma et al., 2015; Dunipace et al., 2011). Presumably, frequent interactions between *sna* distal enhancer and the *sna* promoter mediate RNA polymerase II binding and release at the maximal level, so additional input from the weak proximal enhancer does not further increase *sna* overall levels (Bothma et al., 2015). Instead, when the proximal enhancer is deleted, the *sna* levels increase relative to when both enhancers are present. In contrast, when the distal enhancer is deleted, the *sna* levels decrease in comparison to when both enhancers are present and active. Thus, these results suggest that the proximal enhancer attenuates the activity of distal enhancer and action of both enhancers is required to support correct *sna* levels (Fig. 4C) (Dunipace et al., 2011). Similar studies at the *Kr* locus have identified that also enhancers act in a sub-additive way to support *Kr* expression also (Scholes et al., 2019).

It is clear that the *cis*-regulatory logic inherent in target gene enhancer sequences interprets morphogen inputs to support both spatial as well as temporal-regulated outputs (Datta et al., 2018; Koromila et al., 2020; Yuh et al., 1998). We propose that co-acting enhancer pairs act to support precise and accurate gene expression (Bentovim et al., 2017) even under genetic- and environmental-induced variability including variability in Bcd and Dl morphogens.

## CONCLUSION

In summary, by dissecting the role of morphogen inputs directly as well as the ability of *cis*-regulatory modules to work in a coordinate manner, the field is making progress towards understanding the process of patterning, which relates not only to pivotal morphogen inputs but also is equally dependent on action of other components of the acting gene regulatory network. However, questions still remain about the role of morphogen dynamics.

It is likely that Bcd and Dl morphogens exhibit opposite trends in space and over time, and it is possible that nuclei read these dynamic inputs to infer 3D positional information. From the

perspective of a nucleus positioned at a ventral position in the embryo where peak Dl levels are present, the morphogen Dl appears to exhibit a larger change in nuclear levels from one nuclear cycle to the next (Fig. 1B); while nuclei located at a distance along the DV axis exhibit less of a difference (Fig. 5A). Moreover, the opposite trend is present for Bcd levels in nuclei located in different positions along the AP axis. Nuclei positioned more posteriorly may be exposed to larger differential in Bcd levels over time, either due to increased Bcd protein production overtime or as a result of prolonged Bcd diffusion from the anterior (Fig. 5C). In contrast, along the DV axis, ventral nuclei (see plot for trace) exhibit the largest change in Dl exposure overtime, and nuclei located along the DV axis exhibit decreasing differential the more dorsal their position (Fig. 5C). These opposite trends likely regulate the target gene threshold responses exhibited by Bcd and Dl, possibly by affecting burst kinetics as described for targets of another morphogen BMP (Hoppe et al., 2020). While it is unclear whether these Bcd and Dl dynamics differentially affect the expression of target genes, recent advances in single cell sequencing coupled to ATAC-seq analyses as well as the use of optogenetic approaches to inactivate these TFs with temporal precision should provide insight in the future. It is possible that dual-input from these two morphogens that pattern different embryonic axes coordinate in space and time to help cells identify their position in the embryo (Fig. 5B, D).

In addition, absolute morphogen levels do not necessarily act only to regulate the spatial domain of target gene expression patterns. In the case of Dl at the late blastoderm stage, many Dl target genes maintain constant expression domains while Dl levels continue to increase. We propose that during this late stage of the patterning process Dl may no longer impact patterning likely because other transcription factors, activators and/or repressors, take over expression and/or Dl switches its role. Our recent study of temporal regulation of *sna* expression identified that the Dl-target Twi transcription factor can suffice to activate expression at later timepoints through one

particular enhancer *sna\_Distal*, and serves as a molecularly-described example of hysteresis (Huang et al., 2017; Irizarry et al., 2020). In various systems, it is suggested that hysteresis (“memory”) is a mechanism to support robust patterning under inherent noisiness during development (Balaskas et al., 2012; Bollenbach et al., 2008; Manu et al., 2009). It will be of interest to determine if these maternal transcription factors play any role in gene regulation after cellularization when their nuclear gradients are still building, but they have both been shown to be expendable in expression of major target genes (i.e. *sna* and *hb*) (Irizarry et al., 2020).

Lastly, stage-specific change in responsiveness of target genes to morphogen levels is also observed in vertebrate systems. During neural tube patterning, Shh forms ventral-dorsal gradients to activate a series of target genes. The ventral most cells require high Shh levels, which can be achieved by increasing the duration of exposure to Shh signaling (Balaskas et al., 2012). However, cells become desensitized with increased exposure to this signal over time, and “hysteresis” allows the cells to maintain proper identities regardless of fluctuation of morphogen input (Balaskas et al., 2012; Dessaud et al., 2010). Similarly, during *Drosophila* wing disc patterning, *dpp* expression is maintained in the region where the transient Hh signal is received. It is supported by a memory module that is initiated by early Hh signaling, rather than autoregulation feedback (Nahmad and Stathopoulos, 2009). It will be of interest to determine whether specific enhancer action supports these and other examples of hysteresis described in other developmental systems.

## REFERENCES

- Anderson, K. V. and Nüsslein-Volhard, C.** (1984). Information for the dorsal–ventral pattern of the *Drosophila* embryo is stored as maternal mRNA. *Nature* **311**, 223–227.
- Balaskas, N., Ribeiro, A., Panovska, J., Dessaud, E., Sasai, N., Page, K. M., Briscoe, J. and Ribes, V.** (2012). Gene regulatory logic for reading the Sonic Hedgehog signaling gradient in the vertebrate neural tube. *Cell* **148**, 273–284.
- Barolo, S.** (2012). Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *Bioessays* **34**, 135–141.
- Barolo, S. and Levine, M.** (1997). hairy mediates dominant repression in the *Drosophila* embryo. *EMBO J.* **16**, 2883–2891.
- Belvin, M. P. and Anderson, K. V.** (1996). A conserved signaling pathway: the *Drosophila* toll-dorsal pathway. *Annu. Rev. Cell Dev. Biol.* **12**, 393–416.
- Belvin, M. P., Jin, Y. and Anderson, K. V.** (1995). Cactus protein degradation mediates *Drosophila* dorsal-ventral signaling. *Genes Dev.* **9**, 783–793.
- Bentovim, L., Harden, T. T. and DePace, A. H.** (2017). Transcriptional precision and accuracy in development: from measurements to models and mechanisms. *Development* **144**, 3855–3866.
- Bhatia, S., Bengani, H., Fish, M., Brown, A., Divizia, M. T., de Marco, R., Damante, G., Grainger, R., van Heyningen, V. and Kleinjan, D. A.** (2013). Disruption of Autoregulatory Feedback by a Mutation in a Remote, Ultraconserved PAX6 Enhancer Causes Aniridia. *The American Journal of Human Genetics* **93**, 1126–1134.
- Biemar, F., Nix, D. A., Piel, J., Peterson, B., Ronshaugen, M., Sementchenko, V., Bell, I., Manak, J. R. and Levine, M. S.** (2006). Comprehensive identification of *Drosophila* dorsal-ventral patterning genes using a whole-genome tiling array. *Proceedings of the National Academy of Sciences* **103**, 12763–12768.
- Bollenbach, T., Pantazis, P., Kicheva, A., Bökel, C., González-Gaitán, M. and Jülicher, F.** (2008). Precision of the Dpp gradient. *Development* **135**, 1137–1146.
- Borok, M. J., Tran, D. A., Ho, M. C. W. and Drewell, R. A.** (2010). Dissecting the regulatory switches of development: lessons from enhancer evolution in *Drosophila*. *Development* **137**, 5–13.
- Bosch, J. R. t. and t. Bosch, J. R.** (2006). The TAGteam DNA motif controls the timing of *Drosophila* pre-blastoderm transcription. *Development* **133**, 1967–1977.
- Bothma, J. P., Garcia, H. G., Ng, S., Perry, M. W., Gregor, T. and Levine, M.** (2015). Enhancer additivity and non-additivity are determined by enhancer strength in the *Drosophila* embryo. *Elife* **4**.

**Briscoe, J. and Small, S.** (2015). Morphogen rules: design principles of gradient-mediated embryo patterning. *Development* **142**, 3996–4009.

**Briscoe, J., Chen, Y., Jessell, T. M. and Struhl, G.** (2001). A hedgehog-insensitive form of patched provides evidence for direct long-range morphogen activity of sonic hedgehog in the neural tube. *Mol. Cell* **7**, 1279–1291.

**Chen, Y. and Schier, A. F.** (2001). The zebrafish Nodal signal Squint functions as a morphogen. *Nature* **411**, 607–610.

**Chen, H., Xu, Z., Mei, C., Yu, D. and Small, S.** (2012). A System of Repressor Gradients Spatially Organizes the Boundaries of Bicoid-Dependent Target Genes. *Cell* **149**, 618–629.

**Datta, R. R., Ling, J., Kurland, J., Ren, X., Xu, Z., Yucel, G., Moore, J., Shokri, L., Baker, I., Bishop, T., et al.** (2018). A feed-forward relay integrates the regulatory activities of Bicoid and Orthodenticle via sequential binding to suboptimal sites. *Genes Dev.* **32**, 723–736.

**Dessaud, E., Ribes, V., Balaskas, N., Yang, L. L., Pierani, A., Kicheva, A., Novitsch, B. G., Briscoe, J. and Sasai, N.** (2010). Dynamic assignment and maintenance of positional identity in the ventral neural tube by the morphogen sonic hedgehog. *PLoS Biol.* **8**, e1000382.

**Driever, W. and Nüsslein-Volhard, C.** (1988). A gradient of bicoid protein in *Drosophila* embryos. *Cell* **54**, 83–93.

**Dunipace, L., Ozdemir, A. and Stathopoulos, A.** (2011). Complex interactions between cis-regulatory modules in native conformation are critical for *Drosophila* snail expression. *Development* **138**, 4075–4084.

**Dunipace, L., Saunders, A., Ashe, H. L. and Stathopoulos, A.** (2013). Autoregulatory feedback controls sequential action of cis-regulatory modules at the brinker locus. *Dev. Cell* **26**, 536–543.

**El-Sherif, E. and Levine, M.** (2016). Shadow Enhancers Mediate Dynamic Shifts of Gap Gene Expression in the *Drosophila* Embryo. *Curr. Biol.* **26**, 1164–1169.

**Foo, S. M., Sun, Y., Lim, B., Ziukaite, R., O'Brien, K., Nien, C.-Y., Kirov, N., Shvartsman, S. Y. and Rushlow, C. A.** (2014). Zelda potentiates morphogen activity by increasing chromatin accessibility. *Curr. Biol.* **24**, 1341–1346.

**Fujioka, M., Emi-Sarker, Y., Yusibova, G. L., Goto, T. and Jaynes, J. B.** (1999). Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. *Development* **126**, 2527–2538.

**Furlong, E. E. M. and Levine, M.** (2018). Developmental enhancers and chromosome topology. *Science* **361**, 1341–1345.

**Gergen, J. P. and Butler, B. A.** (1988). Isolation of the *Drosophila* segmentation gene *runt* and analysis of its expression during embryogenesis. *Genes Dev.* **2**, 1179–1193.

- Gregor, T., Wieschaus, E. F., McGregor, A. P., Bialek, W. and Tank, D. W.** (2007). Stability and nuclear dynamics of the bicoid morphogen gradient. *Cell* **130**, 141–152.
- Hannon, C. E., Blythe, S. A. and Wieschaus, E. F.** (2017). Concentration dependent chromatin states induced by the bicoid morphogen gradient. *Elife* **6**,.
- Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R. and Eisen, M. B.** (2011). Zelda binding in the early *Drosophila melanogaster* embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet.* **7**, e1002266.
- Hong, J.-W., Hendrix, D. A. and Levine, M. S.** (2008). Shadow enhancers as a source of evolutionary novelty. *Science* **321**, 1314.
- Hoppe, C., Bowles, J. R., Minchington, T. G., Sutcliffe, C., Upadhyai, P., Rattray, M. and Ashe, H. L.** (2020). Modulation of the Promoter Activation Rate Dictates the Transcriptional Response to Graded BMP Signaling Levels in the *Drosophila* Embryo. *Dev. Cell*.
- Huang, A. and Saunders, T. E.** (2020). A matter of time: Formation and interpretation of the Bicoid morphogen gradient. *Curr. Top. Dev. Biol.* **137**, 79–117.
- Huang, A., Amourda, C., Zhang, S., Tolwinski, N. S. and Saunders, T. E.** (2017). Decoding temporal interpretation of the morphogen Bicoid in the early *Drosophila* embryo. *eLife* **6**,.
- Irizarry, J., McGehee, J., Kim, G., Stein, D. and Stathopoulos, A.** (2020). Twist-dependent ratchet functioning downstream from Dorsal revealed using a light-inducible degron. *Genes Dev.* **34**, 965–972.
- Jaeger, J., Surkova, S., Blagov, M., Janssens, H., Kosman, D., Kozlov, K. N., Manu, Myasnikova, E., Vanario-Alonso, C. E., Samsonova, M., et al.** (2004). Dynamic control of positional information in the early *Drosophila* embryo. *Nature* **430**, 368–371.
- Jiang, J. and Levine, M.** (1993). Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell* **72**, 741–752.
- Koromila, T. and Stathopoulos, A.** (2017). Broadly expressed repressors integrate patterning across orthogonal axes in embryos. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 8295–8300.
- Koromila, T., Gao, F., Iwasaki, Y., He, P., Pachter, L., Gergen, J. P. and Stathopoulos, A.** (2020). Odd-paired is a pioneer-like factor that coordinates with Zelda to control gene expression in embryos. *Elife* **9**,.
- Kosman, D., Ip, Y. T., Levine, M. and Arora, K.** (1991). Establishment of the mesoderm-neuroectoderm boundary in the *Drosophila* embryo. *Science* **254**, 118–122.
- Li, X.-Y. and Eisen, M. B.** Zelda potentiates transcription factor binding to zygotic enhancers by increasing local chromatin accessibility during early *Drosophila melanogaster* embryogenesis.
- Li, X.-Y., Harrison, M. M., Villalta, J. E., Kaplan, T. and Eisen, M. B.** (2014). Establishment of regions of genomic activity during the *Drosophila* maternal to zygotic transition. *Elife* **3**,.

- Liberman, L. M. and Stathopoulos, A.** (2009). Design flexibility in cis-regulatory control of gene expression: synthetic and comparative evidence. *Dev. Biol.* **327**, 578–589.
- Lim, B., Fukaya, T., Heist, T. and Levine, M.** (2018). Temporal dynamics of pair-rule stripes in living *Drosophila* embryos. *Proceedings of the National Academy of Sciences* **115**, 8376–8381.
- Little, S. C., Tkačik, G., Kneeland, T. B., Wieschaus, E. F. and Gregor, T.** (2011). The Formation of the Bicoid Morphogen Gradient Requires Protein Movement from Anteriorly Localized mRNA. *PLoS Biology* **9**, e1000596.
- Liu, J. and Ma, J.** (2013). Uncovering a dynamic feature of the transcriptional regulatory network for anterior-posterior patterning in the *Drosophila* embryo. *PLoS One* **8**, e62641.
- Liu, F., Morrison, A. H. and Gregor, T.** (2013). Dynamic interpretation of maternal inputs by the *Drosophila* segmentation gene network. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 6724–6729.
- Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J. M., Laxova, R., et al.** (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025.
- Manu, Surkova, S., Spirov, A. V., Gursky, V. V., Janssens, H., Kim, A.-R., Radulescu, O., Vanario-Alonso, C. E., Sharp, D. H., Samsonova, M., et al.** (2009). Canalization of gene expression and domain shifts in the *Drosophila* blastoderm by dynamical attractors. *PLoS Comput. Biol.* **5**, e1000303.
- McDowell, N. and Gurdon, J. B.** (1999). Activin as a morphogen in *Xenopus* mesoderm induction. *Semin. Cell Dev. Biol.* **10**, 311–317.
- McGregor, A. P., Orgogozo, V., Delon, I., Zanet, J., Srinivasan, D. G., Payre, F. and Stern, D. L.** (2007). Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature* **448**, 587–590.
- Nahmad, M. and Stathopoulos, A.** (2009). Dynamic interpretation of hedgehog signaling in the *Drosophila* wing disc. *PLoS Biol.* **7**, e1000202.
- Neumann, C. J. and Cohen, S. M.** (1997). Long-range action of Wingless organizes the dorsal-ventral axis of the *Drosophila* wing. *Development* **124**, 871–880.
- Nien, C.-Y., Liang, H.-L., Butcher, S., Sun, Y., Fu, S., Gocha, T., Kirov, N., Manak, J. R. and Rushlow, C.** (2011). Temporal coordination of gene networks by Zelda in the early *Drosophila* embryo. *PLoS Genet.* **7**, e1002339.
- Ochoa-Espinosa, A., Yucel, G., Kaplan, L., Pare, A., Pura, N., Oberstein, A., Papatsenko, D. and Small, S.** (2005). The role of binding site cluster strength in Bicoid-dependent patterning in *Drosophila*. *Proceedings of the National Academy of Sciences* **102**, 4960–4965.
- Ochoa-Espinosa, A., Yu, D., Tsirigos, A., Struffi, P. and Small, S.** (2009). Anterior-posterior positional information in the absence of a strong Bicoid gradient. *Proceedings of the National Academy of Sciences* **106**, 3823–3828.

- Ozdemir, A., Fisher-Aylor, K. I., Pepke, S., Samanta, M., Dunipace, L., McCue, K., Zeng, L., Ogawa, N., Wold, B. J. and Stathopoulos, A.** (2011). High resolution mapping of Twist to DNA in *Drosophila* embryos: Efficient functional analysis and evolutionary conservation. *Genome Research* **21**, 566–577.
- Ozdemir, A., Ma, L., White, K. P. and Stathopoulos, A.** (2014). Su(H)-Mediated Repression Positions Gene Boundaries along the Dorsal-Ventral Axis of *Drosophila* Embryos. *Developmental Cell* **31**, 100–113.
- Papatsenko, D. and Levine, M.** (2005). Quantitative analysis of binding motifs mediating diverse spatial readouts of the Dorsal gradient in the *Drosophila* embryo. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 4966–4971.
- Perry, M. W., Boettiger, A. N. and Levine, M.** (2011). Multiple enhancers ensure precision of gap gene-expression patterns in the *Drosophila* embryo. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 13570–13575.
- Perry, M. W., Bothma, J. P., Luu, R. D. and Levine, M.** (2012). Precision of Hunchback Expression in the *Drosophila* Embryo. *Current Biology* **22**, 2247–2252.
- Reeves, G. T. and Stathopoulos, A.** (2009). Graded Dorsal and Differential Gene Regulation in the *Drosophila* Embryo. *Cold Spring Harbor Perspectives in Biology* **1**, a000836–a000836.
- Reeves, G. T., Trisnadi, N., Truong, T. V., Nahmad, M., Katz, S. and Stathopoulos, A.** (2012). Dorsal-ventral gene expression in the *Drosophila* embryo reflects the dynamics and precision of the dorsal nuclear gradient. *Dev. Cell* **22**, 544–557.
- Rogers, K. W. and Schier, A. F.** (2011). Morphogen gradients: from generation to interpretation. *Annu. Rev. Cell Dev. Biol.* **27**, 377–407.
- Roth, S., Stein, D. and Nüsslein-Volhard, C.** (1989). A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the *Drosophila* embryo. *Cell* **59**, 1189–1202.
- Rushlow, C. A., Han, K., Manley, J. L. and Levine, M.** (1989). The graded distribution of the dorsal morphogen is initiated by selective nuclear transport in *Drosophila*. *Cell* **59**, 1165–1177.
- Sagner, A. and Briscoe, J.** (2017). Morphogen interpretation: concentration, time, competence, and signaling dynamics. *Wiley Interdiscip. Rev. Dev. Biol.* **6**,.
- Sandler, J. E., Irizarry, J., Stepanik, V., Dunipace, L., Amrhein, H. and Stathopoulos, A.** (2018). A Developmental Program Truncates Long Transcripts to Temporally Regulate Cell Signaling. *Developmental Cell* **47**, 773–784.e6.
- Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V. and Furlong, E. E. M.** (2007). A core transcriptional network for early mesoderm development in *Drosophila* melanogaster. *Genes Dev.* **21**, 436–449.
- Scholes, C., Biette, K. M., Harden, T. T. and DePace, A. H.** (2019). Signal Integration by Shadow Enhancers and Enhancer Duplications Varies across the *Drosophila* Embryo. *Cell Rep.* **26**, 2407–2418.e5.

**Schulz, K. N., Bondra, E. R., Moshe, A., Villalta, J. E., Lieb, J. D., Kaplan, T., McKay, D. J. and Harrison, M. M.** (2015). Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early *Drosophila* embryo. *Genome Res.* **25**, 1715–1726.

**Schweisguth, F. and Posakony, J. W.** (1992). Suppressor of Hairless, the *Drosophila* homolog of the mouse recombination signal-binding protein gene, controls sensory organ cell fates. *Cell* **69**, 1199–1212.

**Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U. and Gaul, U.** (2008). Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature* **451**, 535–540.

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

**Stathopoulos, A. and Levine, M.** (2002a). Dorsal gradient networks in the *Drosophila* embryo. *Dev. Biol.* **246**, 57–67.

**Stathopoulos, A. and Levine, M.** (2002b). Linear signaling in the Toll-Dorsal pathway of *Drosophila*: activated Pelle kinase specifies all threshold outputs of gene expression while the bHLH protein Twist specifies a subset. *Development* **129**, 3411–3419.

**Stathopoulos, A. and Newcomb, S.** (2020). Setting up for gastrulation: *D. melanogaster*. *Curr. Top. Dev. Biol.* **136**, 3–32.

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

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

**Sun, Y., Nien, C.-Y., Chen, K., Liu, H.-Y., Johnston, J., Zeitlinger, J. and Rushlow, C.** (2015). Zelda overcomes the high intrinsic nucleosome barrier at enhancers during *Drosophila* zygotic genome activation. *Genome Res.* **25**, 1703–1714.

**Szymanski, P. and Levine, M.** (1995). Multiple modes of dorsal-bHLH transcriptional synergy in the *Drosophila* embryo. *EMBO J.* **14**, 2229–2238.

**Tucker, J. A., Mintzer, K. A. and Mullins, M. C.** (2008). The BMP Signaling Gradient Patterns Dorsoventral Tissues in a Temporally Progressive Manner along the Anteroposterior Axis. *Developmental Cell* **14**, 108–119.

**Uslu, V. V., Petretich, M., Ruf, S., Langenfeld, K., Fonseca, N. A., Marioni, J. C. and Spitz, F.** (2014). Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nat. Genet.* **46**, 753–758.

**Wolpert, L.** (1969). Positional information and the spatial pattern of cellular differentiation. *Journal of Theoretical Biology* **25**, 1–47.

**Wolpert, L.** (1996). One hundred years of positional information. *Trends Genet.* **12**, 359–364.

**Xiong, N., Kang, C. and Raulet, D. H.** (2002). Redundant and unique roles of two enhancer elements in the TCRgamma locus in gene regulation and gammadelta T cell development. *Immunity* **16**, 453–463.

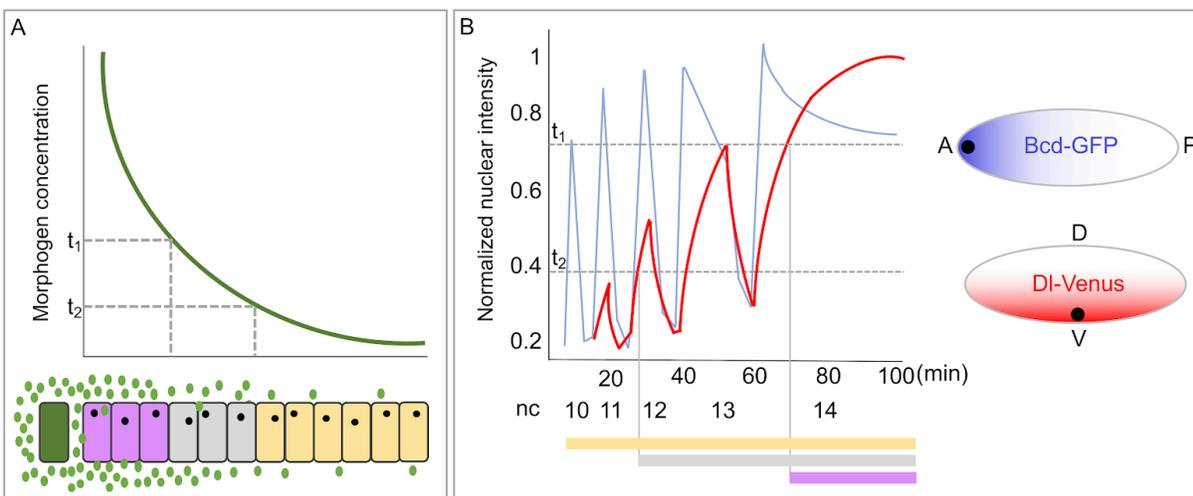
**Xu, Z., Chen, H., Ling, J., Yu, D., Struffi, P. and Small, S.** (2014). Impacts of the ubiquitous factor Zelda on Bicoid-dependent DNA binding and transcription in *Drosophila*. *Genes & Development* **28**, 608–621.

**Yuh, C. H., Bolouri, H. and Davidson, E. H.** (1998). Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science* **279**, 1896–1902.

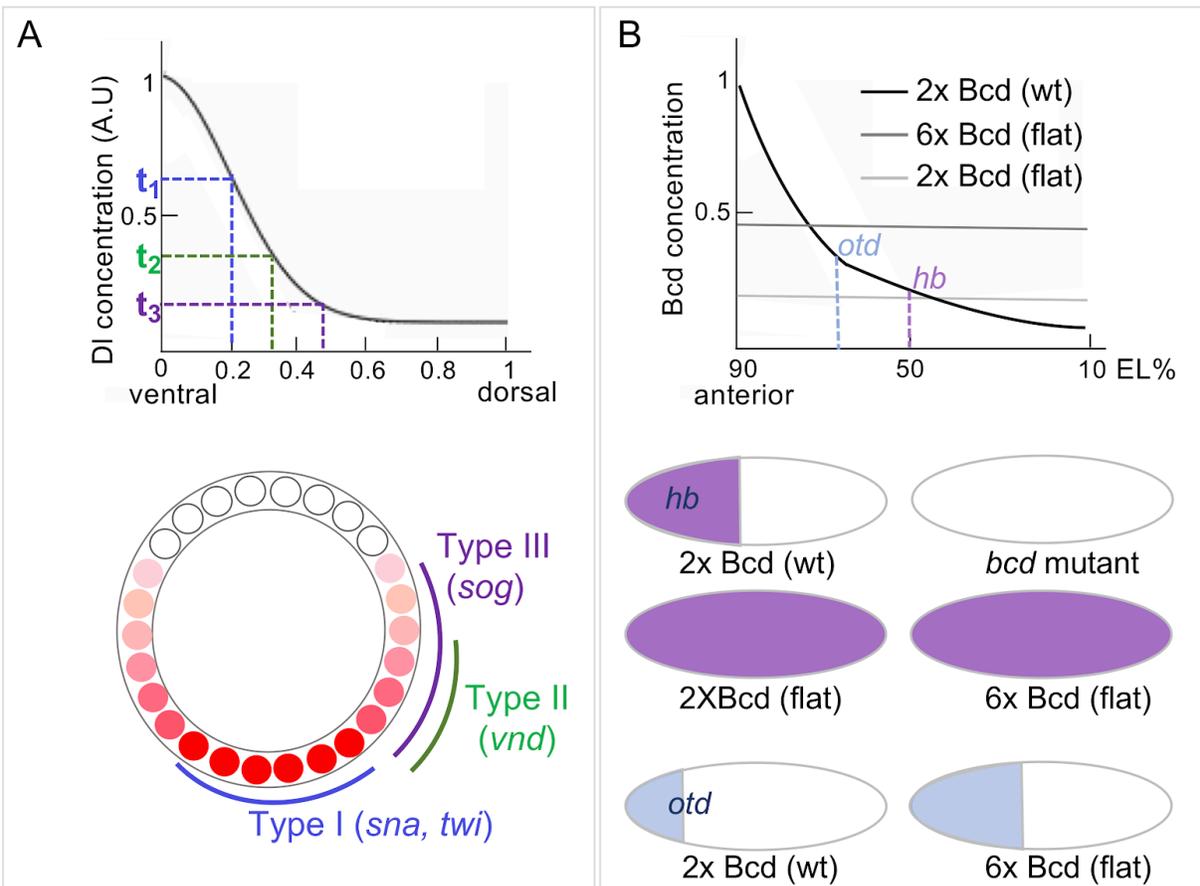
**Zaret, K. S. and Carroll, J. S.** (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241.

**Zeitlinger, J., Zinzen, R. P., Stark, A., Kellis, M., Zhang, H., Young, R. A. and Levine, M.** (2007). Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the *Drosophila* embryo. *Genes Dev.* **21**, 385–390.

## FIGURES & LEGENDS

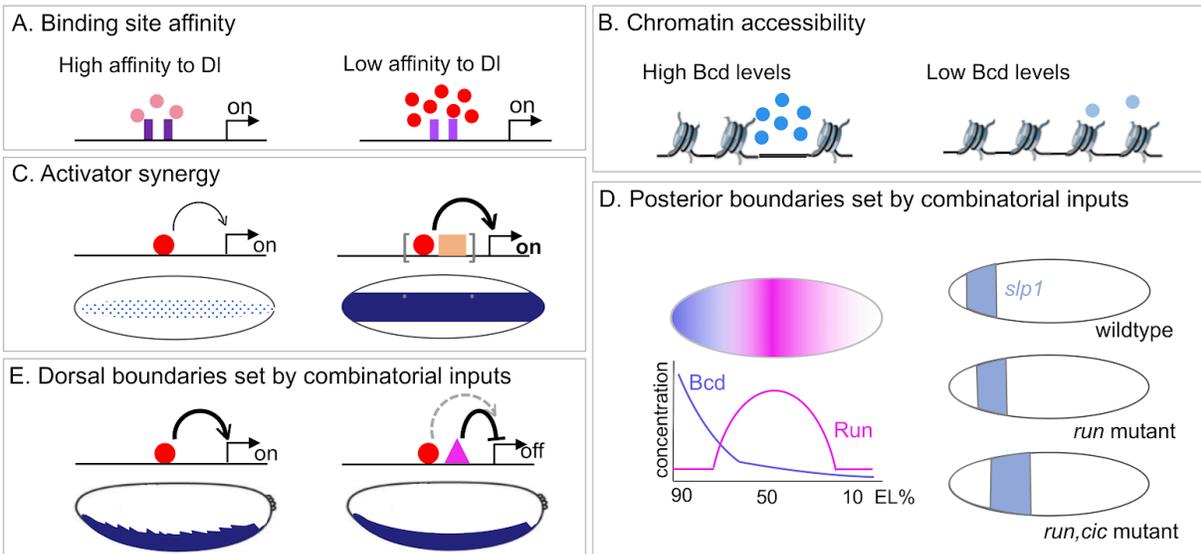


**Figure 1. Comparison of Bicoid and Dorsal morphogens in space and over time in *Drosophila* embryos.** (A) The French flag model for patterning (Wolpert, 1969) includes morphogen production from a source (green) along with its diffusion to neighboring cells, generating a morphogen gradient. A cell exposed to morphogen concentration above threshold 1 ( $t_1$ ) adopts one cell fate (purple); whereas cells at a distance adopt distinct cell fates (gray versus yellow) in response to lower morphogen concentrations above or below threshold 2 ( $t_2$ ). (B) A conceptual representation of the AP patterning morphogen Bicoid (Bcd; blue) and DV patterning morphogen Dorsal (Dl; red) from nuclear cycle 10 to 14. Fluorescence intensity was measured by monitoring Bcd-GFP or Dl-Venus. For Bcd-GFP, the measurements were taken at the single nucleus located at 10% along the AP axis. For Dl-Venus, the measurements were taken at the ventral-most region. The intensities were normalized to the maximum value (Gregor et al., 2007; Reeves et al., 2012). Scheme representing gradients of Bcd along the AP axis (top) and Dl along the DV axis (bottom) in *Drosophila* blastoderm embryos and hypothetical target gene threshold responses below graph. A: anterior pole, P: posterior pole, D: dorsal region, V: ventral region.

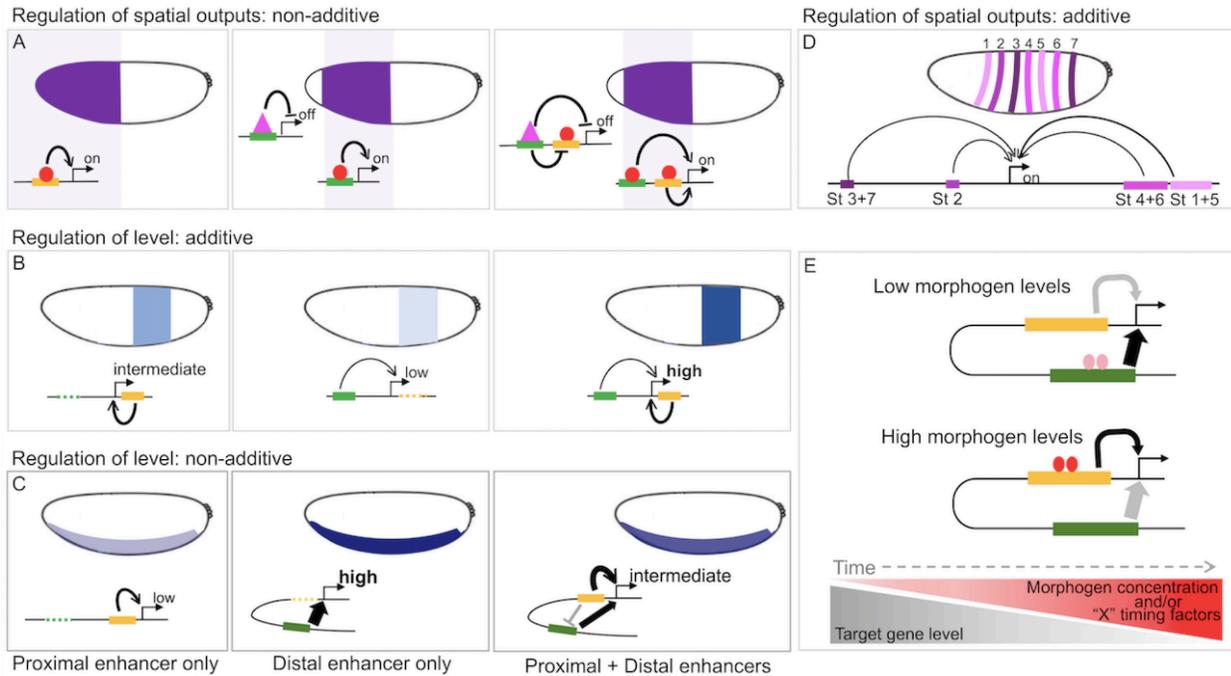


**Figure 2. Morphogen target genes are activated in a threshold-dependent manner. (A)** Dl threshold-dependent gene activation along the DV axis of *Drosophila* embryos. The graph shows how three different Dl thresholds ( $t_1$ ,  $t_2$ ,  $t_3$ ) establishes three gene expression domains along the DV axis. Schematic of embryo cross-section showing three gene expression domains (Types I-III) dictated by threshold dependent responses to graded nuclear Dl levels (red). **(B)** A set of Bcd target genes are controlled by Bcd in a threshold-dependent manner. In wildtype embryos with 2 copies of *bcd*, Bcd exists in a graded manner along the AP axis [2x BCD (wt), blue line] and also supports *hb* expression at the anterior of embryos. In a *bcd* mutant, *hb* loses its expression. When the Bcd gradient is flattened by assay in certain mutant backgrounds resulting in intermediate [6x BCD (flat)], purple line] or low [2x BCD (flat), green line] levels of Bcd throughout the embryo, *hb* expression expands to the posterior pole region. Another Bcd-target gene *otd* (light purple) also is

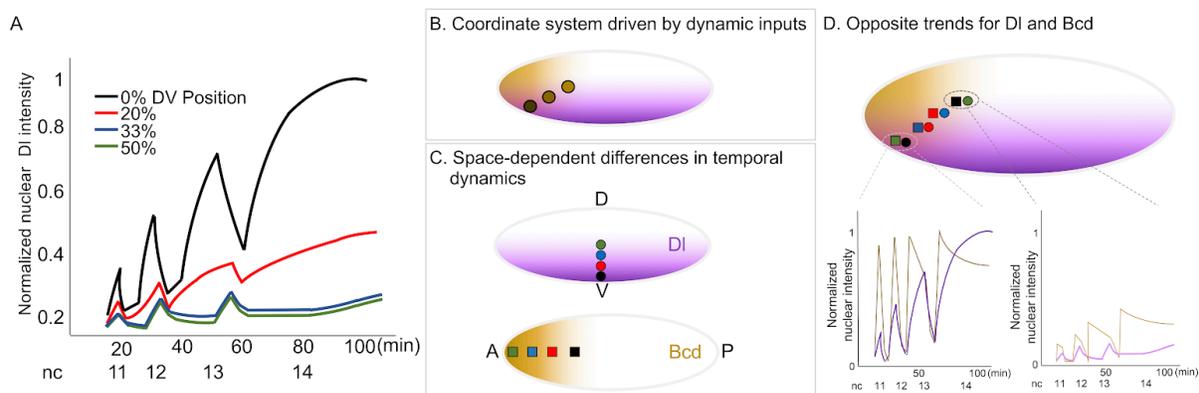
expressed at the anterior in wildtype embryos; but 6x BCD embryos only exhibit posterior expansion of *otd* not ubiquitous expression as found for *hb* (Ochoa-Espinosa et al., 2009).



**Figure 3. Factors influencing threshold outputs.** (A) Affinity of DI binding sites (high affinity, dark purple, and low affinity, light purple) on target enhancer regions dictates DI threshold levels (low, pink, and high, red). (B) Regulation of chromatin accessibility by Bcd. In nuclei of high Bcd concentration, chromatin is more accessible, compared to low Bcd concentration domain. (C) Cooperative inputs of DI (red) and activator Twi (orange) synergistically support target gene expression. (D) Posterior boundaries of AP patterning genes are controlled by Run (pink). In the *run* mutant, *slp1* slightly expands posteriorly, and expands even more posteriorly in a *run* and *cic* double mutant. (E) Dorsal boundaries of DV patterning genes are correctly set by repressive input, as from Su(H) (pink), which acts to limit DI (red)-dependent activation (Ozdemir et al., 2014). Without repressive Su(H) input, the boundaries of *sna* are imprecise. In comparison, combinatorial input from DI and Su(H) supports precise *sna* dorsal boundaries with clear on/off domains of expression.



**Figure 4. Multiple enhancers coordinate to support precise gene expression outputs. (A)** The proximal (yellow) and distal (green) enhancers driving *hb* gene expression are active in similar domains but the distal enhancer expression, specifically, has repressive input that acts in a dominant fashion (Perry et al., 2012). **(B)** *kni* levels are supported by both proximal and distal enhancers, and each enhancer equally contributes to support *Kni* levels (Bothma et al., 2015). **(C)** Proximal and distal enhancers at the *sna* locus act together to regulate levels of expression in a non-additive (or sub-additive) manner (Bothma et al., 2015; Dunipace et al., 2011). This response presumably relates to dominant repression by the proximal enhancer that acts to limit the distal enhancer activity. **(D)** *eve* expression is driven by multiple enhancers that support expression in predominantly distinct domains (rev. in Borok et al., 2010). **(E)** Schematic showing how morphogen concentration may differentially control individual enhancer action over time in response to building morphogen levels, as associated with Dorsal, and/or through input from timing factors that provide temporal information (“X”).



**Figure 5. Morphogen dynamics in time and space.** (A) A representation of morphogen Dorsal nuclear levels at four different positions along the DV axis (see key) from nuclear cycle 11 to 14. Fluorescence intensity was measured by monitoring DI-Venus (Reeves et al., 2012). (B) Nuclei in different positions along the DV or AP axes exhibit different levels of DI (purple) and Bcd (brown) morphogens. (C) Dynamics of morphogen inputs over time may vary along the respective axes. This is measured for DI: colored circles along DV axis in (C), top, correspond to (A); and hypothesized for Bcd, which may relate to increased protein production over time or diffusion. (D) Dual-input by two morphogens exhibiting opposite dynamic trends. Nuclei along the DV axis exhibit progressively smaller changes in levels of nuclear DI from one nuclear cycle to the next (see A), whereas we hypothesize that Bcd exhibits an opposite trend such that a nucleus located more posteriorly may see larger relative increase in Bcd levels from one nuclear cycle to the next (compare graphs).

## Chapter 2

Twist-dependent ratchet functioning downstream of Dorsal revealed using a light-inducible degron

This chapter was written by Jihyun Irizarry, James McGehee and Angelike Stathopoulos and published in *Genes & Development* in 2020.

## ABSTRACT

Graded transcription factors are pivotal regulators of embryonic patterning but whether their role changes over time is unclear. A light-regulated protein degradation system was used to assay temporal dependence of the transcription factor Dorsal in dorsal-ventral axis patterning of *Drosophila* embryos. Surprisingly, the high-threshold target gene *snail* only requires Dorsal input early but not late when Dorsal levels peak. Instead, late *snail* expression can be supported by action of the Twist transcription factor, specifically, through one enhancer, *sna.distal*. This study demonstrates that continuous input is not required for some Dorsal targets and downstream responses, such as *twist*, function as molecular ratchets.

## INTRODUCTION

The maternally-deposited transcription factor Dorsal (Dl) is considered a morphogen as it forms a nuclear gradient that specifies distinct cell fates along the dorsal-ventral (DV) axis of *Drosophila* embryos (rev. Reeves and Stathopoulos 2009). How Dorsal nuclear concentration, which varies along the DV axis, impacts target gene expression has been studied, but few studies have focused on the temporal action of this transcription factor on its targets (Lieberman et al. 2009; Reeves et al. 2012; Rushlow and Shvartsman 2012). Several recent studies have used optogenetic approaches to study the temporal contributions of other maternal transcription factors, Bicoid (Bcd) and Zelda, finding that they are continuously required in the early embryo to support expression of target genes (McDaniel et al. 2019; Huang et al. 2017). In particular, high-threshold targets of Bcd require continuous input early and late- both preceding and concurrent with cellularization. As Dl levels steadily increase over time, in contrast to Bcd levels which remain relatively constant (rev. Sandler and Stathopoulos 2016b), we hypothesized that target gene dependency on Dl may also be dynamic. In this study, we investigated whether Dl input to target

genes is required continuously, as for Bcd, or if instead Dl input is only required at particular timepoints.

## RESULTS

An optogenetic approach was used to examine the temporal action of Dl in supporting target gene expression, initially focusing analysis on the target gene *snail* (*sna*). *sna* is expressed in ventral regions of the embryo in cells that ultimately contain the highest levels of nuclear Dl (Kosman et al. 1991) and is therefore considered a high-threshold target. However, while Dl levels peak in ventral regions of the embryo during nuclear cycle (nc) 14, studies have shown that *sna* is expressed within ventral regions at nc13 suggesting lower levels of Dl are in fact sufficient for this high-threshold target (Reeves et al. 2012).

To assay the temporal dependence on Dl for expression of target genes, including the high-threshold response gene *sna* as well as low-threshold responses including genes *short gastrulation* (*sog*) and *decapentaplegic* (*dpp*) (rev. in Reeves and Stathopoulos 2009), an optogenetic Blue Light-Inducible Degron (BLID) sequence was fused to Dl in-frame at the C-terminus through modification of the endogenous gene locus using CRISPR/Cas9 technology (Fig. 1A). BLID consists of a LOV2 domain and a degron sequence, such that in the dark, when the alpha-helix of the LOV2 domain interacts with the LOV core domain, the degron is inaccessible; but upon illumination with blue light (~400-500nm), the helix dissociates from the LOV core domain, the degron is exposed, and the entire fusion protein, Dl-BLID in this case, is degraded (Bonger et al. 2014).

To assay the degradation efficiency of Dl-BLID, embryos laid by homozygous *dl-BLID* mothers were collected and illuminated with blue light for four hours (see Methods). Larval

cuticles were examined as (i) a proxy for changes to Dl levels that manifest as DV patterning defects (Roth et al. 1989) and (ii) to assay for any phenotypes induced indirectly by blue light treatment. The majority of *dl-BLID* embryos illuminated for four hours with blue light exhibit cuticles similar in phenotype to dorsalized embryos laid by *dl* null ( $dl^l/dl^A$ ) mothers (Fig. 1I, compared with 1K) suggesting that Dl-BLID is successfully degraded upon blue light illumination. However, while half the *dl-BLID* embryos that were not subjected to blue light exhibited normal cuticles (Fig. 1B,F), the remaining half exhibited a range of subtle defects including a small number with the more severe, dorsalized cuticle phenotype (Supplemental Fig. S1B-D). In contrast, wildtype and *dl* null mutant embryos appear unaltered when exposed to blue light for four hours (Supplemental Fig. S1A,E), supporting the view that differences in *dl-BLID* cuticles, in the light versus dark, result from light-induced degradation and not indirect effects of blue light exposure. These results suggest that blue light degrades Dl, but that the degradation process is likely leaky, occurring to some degree even in the dark.

To directly test if Dl is degraded upon illumination, we stained embryos with anti-Dl antibody and imaged cross-sections to assay for changes to the nuclear concentration gradient. As expected we found that levels of Dl in wildtype embryos containing an unmodified, native *dl* gene are unaltered both for embryos kept in the dark as well as those exposed to blue light for one hour (Fig. 1C,D). In the dark, the Dl gradient signal associated with *dl-BLID* embryos appears qualitatively lower compared to wildtype (Fig. 1C,G). However, when *dl-BLID* embryos were exposed to blue light for one hour almost all of the signal, especially the nuclear gradient, is lost (Fig. 1H). Taken together, Dl-BLID appears to support a relatively normal Dorsal nuclear gradient that is efficiently degradable with blue light illumination but exhibits increased variability in levels/shape compared to wildtype, even in the dark. We use this finding to our advantage, as lower levels of Dorsal initially are likely to be more easily manipulated by short light exposures.

To further confirm that Dl-BLID is being degraded, Dl protein levels in embryos were examined by Western blot using anti-Dl antibodies (Fig. 1E). After 30 min in the blue light, Dl-BLID protein levels were indeed reduced to barely detectable levels (Fig. 1E). For embryos that were kept in the dark, Dl-BLID proteins levels were lower compared to wildtype embryos, possibly due to leaky degradation of the degron (Fig. 1E). This lower Dl level may contribute to the broad range of cuticle phenotypes observed in *dl-BLID* embryos kept in the dark (Supplemental Fig. S1B). The results from the cuticle preparation, Dl antibody staining, and Western suggested that controlling Dl levels using blue light illumination with temporal resolution is feasible.

To directly observe Dl-BLID degradation by blue light, we created and assayed Dl-BLID fluorescent protein fusions. While we find that Dl-mCherry-BLID fusions do not retain Dl function, this fusion does permit visualization of the kinetics of blue light induced degradation. Embryos expressing Dl-mCherry and Dl-mCherry-BLID were imaged live using confocal microscopy. When control embryos are exposed to a high power (40%) blue laser of 488 nm wavelength for 10 min, Dl-mCherry embryos (lacking BLID sequence) exhibit little to no decrease in Dl signal (Fig. 1J,J'; Supplemental Movie S1). On the other hand, Dl-mCherry-BLID embryos undergo a dramatic decrease in Dl signal (Fig. 1L,L'; Supplemental Movie S1) indicating that Dl-BLID degradation is occurring in embryos, with appreciable degradation observable within minutes rather than hours observed in other systems (Baaske et al. 2018). Taken together, these results warrant use of the Dl-BLID system to finely assay temporal dependence of target genes on Dl over time during early embryonic development.

To determine whether high levels of Dl are required continuously throughout early embryonic development, we utilized confocal microscopy to illuminate individual embryos with blue light for either 20 min starting at nc14a (laser early, LE) or 20 min starting at nc14c (laser late, LL) (Fig. 2A). In addition, a triple fluorescent protein (FP) reporter system (H2A-BFP, MCP-

GFP, PCP-mCherry) (Bothma et al. 2015) was introduced by genetic crosses into the *dl-BLID* background in order to monitor embryonic development and gene expression responses. The H2A-BFP fusion identifies nuclei, which is useful for monitoring all cells in the developing embryos; whereas the MCP-GFP and PCP-mCherry fusions bind to particular RNA stem-loops, which can be used to monitor nascent transcription.

To start, H2A-BFP signal was used to assay whether blue light illumination affects developmental progression of embryos by observing gastrulation, which involves invagination of the presumptive mesoderm. *dl-BLID* embryos invaginate ventrally and proceed through gastrulation even when illuminated at the low power (0.8%) blue laser needed to image H2A-BFP, despite some low-level degradation of Dl-BLID (“dark”, Fig. 2C; Supplemental Movie S2). Alternatively, when additionally subjected to high power (40%) blue laser illumination during an early time window (“LE”, Fig. 2A), *dl-BLID* embryos fail to ventrally invaginate, and therefore do not gastrulate (Fig. 2D; Supplemental Movie S3). Embryos obtained from females lacking nuclear Dl also fail to undergo gastrulation (Leptin and Grunewald 1990) supporting the idea that the failure of *dl-BLID* embryos illuminated early to gastrulation is due to decrease in Dl levels. In contrast, *dl-BLID* embryos illuminated during a late time window (“LL”, Fig. 2A), surprisingly, are able to invaginate (Fig. 2E; Supplemental Movie S4). These differences in developmental progression between embryos illuminated early or late suggest that high levels of Dl achieved by late nc14 are not necessary for embryos to proceed through gastrulation.

To test how Dl target gene expression is altered by lower Dl levels, we performed fluorescent *in situ* hybridization (FISH) using riboprobes to monitor expressions of the genes *dpp*, *sog*, and *sna*, which span the DV axis (Reeves et al. 2012) comparing expression in the dark to that after illumination. In order to collect enough embryos to carry out FISH experiments, we illuminated embryos *en masse* on plates as opposed to using confocal microscope laser

illumination (Fig. 2B; see Methods). *dl-BLID* embryos kept in the dark were analyzed by FISH and show dorsal *dpp* expression at nc14a-b (Fig. 2G) but a narrower *sna* expression domain with increased variability at the anterior (Fig. 2H, Supplemental Fig. S2). In addition, *sog* expression is repressed in this more narrow domain encompassed by its repressor, Sna (Fig. 2F,H,I; Cowden and Levine 2002). Narrowing of the *sna* domain is likely due to lower levels of total DI present in the *dl-BLID* background, even in the dark (Fig. 1E,G).

Embryos illuminated for 30 min before being fixed at nc14a (likely illuminated between nc13-nc14a) exhibit ventrally expanded *dpp* (Fig. 2K) but retracted *sog* (Fig. 2J). As *dpp* and *sog* expression share a boundary, where *dpp* is repressed by DI and *sog* expression is supported by DI, these genes likely share the same threshold response but with opposite effect (rev. in Reeves and Stathopoulos 2009). Furthermore, *sna* expression is lost in embryos fixed at nc14b (likely illuminated between nc14a-b), but *sog* expression appears unaltered (Fig. 2L; see Discussion). *sog* transcription is absent from ventral-most regions, presumably due to the presence and action of Sna protein despite the lack of *sna* transcripts. As *sna* transcripts have a half-life of ~13 min (Boettiger and Levine 2013), Sna protein made before blue light illumination may perdure and continue to repress *sog* (at least partially) in ventral regions (Bothma et al. 2011).

In contrast, embryos illuminated for 30 min before being fixed at nc14d (likely illuminated later between nc14c-d), express both *sna* and *sog* similar to embryos kept in the dark (Fig. 2I,M). These results support the view that the decrease in DI levels upon illumination affects multiple target genes, but in a temporally dependent manner. Collectively, these results suggest that embryos exposed to light late (i.e. nc14c to nc14d) can still gastrulate because of maintained expression of target genes including *sna*, a critical regulator of gastrulation (Leptin and Grunewald 1990); whereas embryos exposed to light early (i.e. nc 14a to nc14b) fail to gastrulate due to loss of *sna*.

To distinguish whether maintenance of *sna* expression at the late timepoint relates to retention of transcripts made earlier or to an ability to produce new transcripts late, even when DI is degraded, we turned to live imaging. The *sna* transcripts identified by *in situ* hybridization within fixed embryos comprise both mature and nascent transcripts; it is difficult to distinguish nascent *sna* transcripts in part because this gene is expressed at high levels and transcripts accumulate. Instead, the MS2-MCP system was used to monitor nascent transcription *in vivo*. Combining the MS2-MCP system with *dl-BLID* allows nascent transcription to be assayed under different illumination schemes. Specifically, transgenic lines containing a previously defined *sna* MS2-based reporter were used to assay *sna* transcriptional activity (Bothma et al. 2015). In these constructs, ~20kB spanning the *sna* locus is used as a reporter in which *sna* is replaced with the *yellow* gene sequence including intronic MS2 RNA stem loop sequences (Fig. 3A; Bothma et al. 2015). When this reporter is actively transcribed, MCP-GFP fusion proteins bind to the stem loops and produce visible nuclear puncta, allowing live monitoring of *sna* expression.

An intermediate power laser setting (5%) was used to image the MS2-MCP signal, while the high power setting was used to degrade DI-BLID (Supplemental Fig. S3A; see Materials & Methods). Again, under these imaging conditions, illumination of *dl-BLID* embryos with high power at the early timepoint (i.e. nc14a-nc14b, “mLE”) leads to gastrulation failure, whereas illumination of embryos later (i.e. nc14c, “mLL”) has no effect on gastrulation despite the use of intermediate laser power to image the MS2-MCP signal for an extended period of time (Supplemental Fig S3B). We used this scheme, in which MS2-MCP imaging and DI-BLID degradation are compatible, to determine how *sna* transcription is affected by temporal changes in DI levels.

We found that wildtype *sna* MS2-MCP signal (*sna.wt*) was retained when embryos were illuminated with high power laser late (mLL; Fig. 3E,I; Supplemental Movie S7), but was

diminished when embryos were illuminated early (mLE; Fig. 3B,H; Supplemental Movie S5). Two enhancers are known to support early *sna* expression during embryogenesis, one proximal (*sna.prox*) and one distal (*sna.dis*) (Ip et al. 1992; Ozdemir et al. 2011; Perry et al. 2010; Dunipace et al. 2011). In order to understand, which cis-regulatory sequences drive *sna* gene expression even when Dl is degraded at the late timepoint, we also assayed two reporter variants in which portions of these two early embryonic enhancers had been deleted, constructed in a previous study (Fig. 3A; Bothma et al. 2015). The *sna.Δprox* reporter behaves as the *sna.wt* reporter: embryos illuminated early lose signal, whereas those illuminated late retain it (Fig. 3C,F,H,I; Supplemental Movies S6,S8). In contrast, the *sna.Δdis* reporter loses expression when illuminated at either timepoint (Fig. 3D,G,H,I; Supplemental Movie S6,S8). Thus, the distal enhancer is necessary for late *sna* expression when no or very little Dl is present, while the proximal enhancer cannot support late *sna* expression in the absence of Dl. This suggests that the proximal *sna* enhancer likely requires high Dl levels for activity. Taken together, these results support a model in which Dl acts through either enhancer (directly or indirectly) early, but that an additional input is required to sustain late *sna* expression through the *sna.dis* enhancer, specifically.

Another Dl target gene encoding a bHLH transcription factor, *twist* (*twi*), is expressed in ventral regions, and also provides input to *sna* (rev. in Reeves and Stathopoulos 2009). *sna* expression is either lost or greatly diminished in *dl* and *twi* mutants, respectively (Ip et al. 1992). *twi* transcript levels increase rapidly at the onset of *nc14* and activation of mesodermal genes follows (Sandler and Stathopoulos 2016a), suggesting that Twi may be an important input into these target genes. Furthermore, peak Dl levels are not required to support *sna* expression as ectopic Twi gradients can support its expression even in conditions of low, but not completely absent, Dl (Stathopoulos and Levine 2002). These previous studies had suggested that Twi may

suffice to support *sna* activation at the late timepoint, even in the absence of DI. However, it was previously not possible to remove DI but retain Twi as *twi* gene expression is DI-dependent.

We hypothesized that Twi is responsible for the late expression of *sna*, essentially taking over for DI. To test this idea, embryos were fixed after 30 min blue LED illumination (Fig. 2B) and assayed for DI and Twi proteins using antibodies, and for *sna* transcripts by FISH. Embryos exposed to light early or late exhibited low levels or no DI as expected but, surprisingly, retained Twi expression (Supplemental Fig. S4) demonstrating that even low levels of DI in nc14 are sufficient to support low levels of Twi expression. *sna* expression is also retained when embryos are illuminated late (Fig. 4A-B''), but not early (Supplemental Fig. 4A-B'') suggesting early nc14 *sna* expression is DI-dependent. However, when the *twi* mutant is recombined with *dl-BLID*, even when embryos are kept in the dark and high levels of DI are present, *sna* expression is lost if Twi is absent (Fig. 4C-C''). Taken together, these results suggested that Twi is a pivotal input for *sna* activation, particularly at late stages when *sna* expression is independent of high DI levels.

In order to understand the temporal relationship between DI and Twi transcription factor dynamics, we assayed Twi dynamics with fine time resolution in combination with temporally controlled DI-BLID levels. Twi levels were detected in *dl-BLID* embryos using a previously described Twi-mCherryLlamaTag fusion protein, which allows early zygotic proteins to be visualized without having to wait for fluorescence maturation (Bothma et al. 2018). When embryos are kept in the dark, mCherry signal intensifies throughout nc14, suggesting exponential production of Twi protein (Fig. 4D,G; Supplemental Movie S9). However, in embryos exposed to high power blue laser illumination at nc14a, no increase in Twi levels is observed ("LE", Fig. 4E,G; Supplemental Movie S10). In contrast, for embryos illuminated at nc14c, Twi levels increase ("LL", Fig. 4F,G; Supplemental Movie S10) similarly to embryos without illumination

(“dark”, Fig. 4D,G; Supplemental Movie S9). These results support the view that Twi is only responsive to Dl levels early, but is able to maintain its expression late even if Dl levels fall.

To examine how responsive Twi is to Dl levels early, we manipulated Dl levels using various durations of blue laser illumination at nc14b and measured the effect live using the Twi-mCherryLlamaTag as a proxy for Twi levels (see Methods). Short time window illuminations with blue laser (<5 min) early presumably lead to small or negligible changes in Dl-BLID levels and therefore had little or no effect on Twi levels throughout nc14 (Fig. 4H). However, with incremental increase in duration of blue laser illumination (5, 7 or 10min), Twi levels also fail to increase, with severity corresponding to the duration of illumination. The rate of change for Twi levels decreases substantially, most apparent with the 10 min exposure (Fig. 4H), instead of undergoing the exponential increase observed in dark, laser late, or short time window illuminations of 1 and 3 min (Fig. 4G,H). These results suggest that Twi levels are reflective of the underlying Dl levels early, and that levels of Dl early impact levels of Twi present later.

Interestingly, intermediate exposure of Dl-BLID to blue laser (e.g. 5 to 10 min) results in loss of the late nc14 exponential increase in Twi levels that is normally observed in control embryos (e.g. dark; Fig. 4G) as the rate of change in levels decreases, but it initially remained unclear why (Fig. 4H). In order to explain why Twi levels do not grow exponentially after intermediate duration Dl degradation, we hypothesized that low levels of Dl are retained that continue to support low levels of Twi. In this scenario, a second blue light illumination to knock-down the remaining Dl would be expected to further decrease Twi levels. However, we find that exposure to a second illumination (e.g. 15 min at nc14c) has no effect; Twi is maintained at levels similar to embryos exposed to a single 5 min illumination at nc14b (Fig. 4I). This observation suggests that in late nc14 activation of Twi shifts to a gene regulatory state that is independent of Dl levels. Collectively, these data support the view that a Twi-dependent threshold exists above

which *Tw* can activate its own expression independently of *Dl* at this late stage and supports a model where levels of *Dl* in early *nc14* determine *twi* expression, but during late *nc14*, *twi* expression is *Dl* independent.

## DISCUSSION

In this study, we have examined whether *Dl* is continuously required to activate target genes in the early embryo by utilizing a *Dl*-*BLID* fusion. *Dl* is required early for the initiation of expression of the *sna* target gene in ventral regions, but surprisingly is not needed late to maintain its expression. Like *sna*, expression of *htl*, *mes3* and *netA* are sustained in *dl-BLID* embryos illuminated with blue LED light late (Supplemental Fig. S7) suggesting that other target genes are similarly regulated. In contrast, we found that the lateral gene *sog* is still expressed no matter when *Dl* degradation occurs during *nc14*. This unexpected result appears to contradict the model where the *sog* dorsal boundary is formed by limiting levels of nuclear Dorsal. Although, one possible explanation that is consistent with this model is that low levels of Dorsal remain after illumination and are enough to activate *sog*. However, this should result in either a narrow *sog* expression domain or requires asymmetrical degradation of Dorsal, neither of which is observed (Supplemental Fig. S2, Fig. 1H, and Supplemental Movie S1). Another explanation is that once the *sog* domain is established by lower levels of *Dl*, *sog* does not require *Dl* to remain active because another factor acts to retain its expression. A simpler explanation is that *sog* transcripts are long and the detected signal could be from *sog* transcripts that were initiated at an earlier timepoint, when Dorsal was present. These possible explanations for how *sog* transcription fails to respond to Dorsal degradation upon illumination are not mutually exclusive. Addressing how *sog* transcription becomes *Dl* independent in future studies will be an important step forward in our understanding of how the *sog* dorsal boundary is set.

Our results also provide insight into how a transcriptional network may buffer changes in levels of a maternal patterning morphogen. In the case of *sna*, high levels of Dl are required early to activate *sna* gene expression. Dl acts both directly, and indirectly by controlling *twi* expression, as Twi is also an input to *sna* (Fig. 4J, left). In contrast, Dl is dispensable for *sna* activation at later timepoints. When Dl levels are reduced *sna* expression remains (Fig. 4J, right), likely maintained by Twi once Twi is expressed. The ability to retain expression of a morphogen target gene despite a decrease in morphogen levels has been termed a “ratchet-reponse”, and was demonstrated for targets of the activin morphogen in *Xenopus* (Gurdon et al. 1995). Twi can maintain its own expression through autoregulation (Kosman et al. 1991; Crews and Pearson 2009), and we propose this autoregulatory feedback serves to support this ratchet response that is able to buffer against decreases in Dl levels. However, simple Twi autoregulatory feedback would predict a single steady-state concentration for Twi. Instead, we observed that Twi levels increase exponentially or reach intermediate levels of Twi when varying the length of illumination. While this result would not support simple autoregulatory feedback as a mechanism for maintaining Twi expression in the absence of high Dl, it requires Twi levels to be at steady-state. It is possible the observed responses have not reached steady-state, and if given enough time they might all converge to the same steady-state concentration (i.e. a single response supported by autoregulation). It is likely that other factors contribute to *twi* regulation, however these results support the model that Dl activates *twi*, and Twi is able to maintain its own expression through autoregulation.

Taken together, we propose that once Twi reaches sufficient levels to support its own auto-activation, Dl is no longer required to support *sna* expression (Fig. 4J, right). This is in sharp contrast to the Bcd morphogen which patterns the anterior-posterior (AP) axis and to the early maternal pioneer factor Zelda (Huang et al. 2017; McDaniel et al. 2019). Both Bcd and Zelda have been found to be required continuously; perturbations at any stage cause loss of gene expression.

Alternatively, the DV gene regulatory network shifts from a state of high Dl-dependence to a state of Dl-independence for several target genes expressed in the presumptive mesoderm. It is possible that this ratchet-response relates to the ability of *twi* gene expression to buffer changes in Dl-concentration and allows the DV-patterning network to respond only to increasing Dl levels. Taken together, ratchet-like responses are crucial steps during animal development not only because they support morphogen-dependent patterning, but also they may serve to buffer expression of target genes against fluctuations in morphogen levels due to genetic and environmental changes.

## MATERIALS & METHODS

### *Fly stocks/ husbandry and plasmids*

All flies were kept at 18°C, unless otherwise noted. *yw* was used as wildtype. Fly stocks used: *dl<sup>4</sup>/CyO* (#7096, Bloomington *Drosophila* Stock Center, BDSC), *dl<sup>1</sup>/CyO* (#3236, BDSC), *twi<sup>1</sup>/CyO* (#2381, BDSC), *nos>MCP-GFP*, *nos>mCherry-PCP*, *His2Av-eBFP2* (from Michael Levine, Princeton University, US; Lim et al. 2018), *snailBAC>MS2* with both proximal and distal enhancers (WT, *sna.wt*), proximal deletion (NoPrimary, *sna.Δprox*), or distal deletion (NoShadow, *sna.Δprox*) (from Michael Levine, Princeton University, US; Bothma et al. 2015), *vasa-mCherry* and *TwimCherryLlamaTag* (from Hernan Garcia, UC Berkeley, US; Bothma et al. 2018). For details regarding fly crosses, see supplemental methods.

### *Genetic crosses*

To generate *dl* mutants, virgin *dl<sup>4</sup>/CyO* (#7096, Bloomington *Drosophila* Stock Center, BDSC) were crossed *dl<sup>1</sup>/CyO* (#3236, BDSC). For the cuticle preparation in *dl* mutant, female *dl<sup>1</sup>/dl<sup>4</sup>* were crossed with *yw* males. To test continuous requirements of high levels of Dl at blastoderm

stage, first, *dl-BLID* flies were recombined with *MCP-GFP*, *mCherry-PCP*, *His2Av-eBFP2* (from Michael Levine, Princeton University, US; Lim et al. 2018), and then *dl-BLID/CyO;MCP-GFP*, *mCherry-PCP*, *His2Av-eBFP2* male was crossed with *dl-BLID/CyO* virgin females to generate *dl-BLID/dl-BILD; MCP-GFP*, *mCherry-PCP*, *His2Av-eBFP2*. Female *dl-BLID/dl-BILD; MCP-GFP*, *mCherry-PCP*, *His2Av-eBFP2* flies were crossed with *yw* males to image nuclei to observe overall development upto gastrulation. Furthermore, virgin *dl-BLID/dl-BILD; MCP-GFP*, *mCherry-PCP*, *His2Av-eBFP2* flies were crossed male *snailBAC>MS2* with both proximal and distal enhancers (WT), proximal deletion (NoPrimary), or distal deletion (NoShadow) (from Michael Levine, Princeton University, US; Bothma et al. 2015). To examine Twi dynamics, first, *dl-BLID* flies were recombined with *vasa-mCherry* (from Hernan Garcia, UC Berkeley, US). Virgin *dl-BLID;vasa-mCherry* flies were crossed with male *Twi-mCherryLlamaTag* (from Hernan Garcia, UC Berkeley, US; Bothma et al. 2018). To recombine *dl-BLID* with *twi<sup>l</sup>*, *dl-BLID/CyO* females were crossed with male *twi<sup>l</sup>/CyO*. Individual *dl-BLID/twi<sup>l</sup>* virgin female was crossed with male double balancer flies. Recombinants were identified by examining brown eye color, crossing putative recombinant *dl-BLID,twi<sup>l</sup>/CyO* with *bw<sup>l</sup>*. Once recombinants were identified, virgin *dl-BLID* flies were crossed with male *dl-BLID,twi<sup>l</sup>/CyO* to generate *dl-BLID, twi<sup>l</sup>/dl-BLID*. Finally, virgin *dl-BLID,twi<sup>l</sup>/dl-BLID* flies were crossed with *twi<sup>l</sup>/CyO* males, and the embryos were collected and fixed.

#### *Generation of gRNAs and homologous repair template construct*

The guide RNAs (gRNA) were designed using the flyCRISPR Target Finder (Gratz et al. 2014). gRNAs that were upstream of the D1 stop codon and downstream of the 3'UTR were chosen (see Table S1 for sequences). The gRNAs were cloned into pCFD4 (Addgene Plasmid #49411) as done previously (Port et al. 2014). Briefly, primers were designed with the gRNA sequence, cut with

BsBI, and ligated into pCFD4. This plasmid was injected into *P{nos-phiC31}X;attP2 (III)* (NIG-FLY, TBX-0003). Integration of the gRNA was screened using  $v^+$ .

The *dl-BLID* homologous repair template was made by editing pHD-DsRed (Gratz et al. 2014). An 1160bp sequence was inserted using BglIII and XhoI sites, and served as the right homology arm. 984bp upstream of the stop codon were fused to BLID sequence from plasmid pBMN HAYFP-LOV24 (Addgene #49570; Bongler et al. 2014) and the 3'UTR using overlap PCR including a 6x Gly linker. This PCR product was inserted using EcoRI and NheI sites and serves as the left homology arm plus the insert.

### *Genome editing*

CRISPR was performed as described previously (Gratz et al. 2014). Briefly, the gRNA fly line (targeting before the c-term and after the 3'UTR of Dorsal, see supplemental methods and Table S1) and the Cas9 line *Sp/CyO, P{nos-Cas9}2A*, (NIG-FLY, CAS-0004) were mated. Embryos were collected and the homology-directed repair (HDR) template containing the C-term of *dl* fused to *BLID* (see supplemental methods) was injected into these embryos. Flies were screened for DsRed. The integration was confirmed by PCR and sequencing.

### *dl-BLID viability*

*dl-BLID* viability is severely reduced when allowed to develop at 25°C, as no larva hatch. All experiments were carried out at 18°C, where the viability is improved but roughly less than 50%.

### *Blue light illumination*

Embryos were collected at 18°C for 1 hr followed by 4 hr incubation for aging and illuminated with blue light using either a set of LEDs (2501BU Blue 225 LED 13.8 Watt Square Grow Light Panel 110) or the 488nm laser on a Zeiss LSM 800 confocal microscope. For blue LED light illumination, embryos on agar plates were placed 6.5 cm below the LED light panel and illuminated for appropriate time lengths. After blue light exposure, the embryos were fixed. For 488 nm blue laser illumination, the embryos were dechorionated and mounted on a heptane glued slide. The embryos were immersed in water, and a blue laser was applied using a 25x water immersion objective. All the embryos were prepared under red filtered light to avoid possible D1-BLID degradation by light coming from microscopes or other ambient sources.

#### *Cuticle preparations*

Embryos were collected at 18°C for 2 hours, aged 1.5 hours in the dark, and illuminated with blue LEDs for 4 hours. Subsequently, embryos were aged for an additional 36-40 h in the dark and then processed by standard cuticle preparation using lactic acid.

#### *Western blot analysis*

Aged embryos were dechorionated and mounted in Halocarbon 27 oil (Sigma-Aldrich). Embryos at nc14b were manually selected and illuminated for 30 min with LED blue light. After light exposure, embryos at nc14c were prepared for standard Western blot.

#### *Immunostaining and fluorescent in situ hybridization (FISH)*

Antisense RNA probes labeled with digoxigenin or FITC-UTP were utilized to examine *sna*, *sog*, *htl*, *mes3*, or *netA* transcripts. For *sna* probes, *sna* was transcribed from cDNA subcloned into pGEM-T vector. For *sog*, *htl*, *mes3*, and *netA* probes, primers were designed to target coding sequence of each gene. Immunostaining and FISH protocols were followed as previously described (Kosman et al. 2004). Sheep anti-digoxigenin (Life Technology PA185378), or rabbit anti-FITC (Invitrogen A889), mouse anti-Dl (1:10; Developmental Studies Hybridoma Bank 7A4) or guinea pig anti-Twi (1:200; Trisnadi and Stathopoulos 2014) were used together with Alexa conjugate secondaries (1:400; Thermo Fisher). DAPI staining (1:10,000 Molecular Probes) was used to mark nuclei.

#### *Live imaging and quantification*

To stage embryos for live imaging, individual embryos were manually dechorionated and mounted on a slide with heptane glue. Once embryos were immersed in water, nuclear morphology was observed live under a confocal microscope brightfield with 25x objective lens. To minimize possible degradation of Dl-BLID during staging, the light was filtered by red film (Neewer, 10087407).

To test efficiency of Dl-BLID degradation upon blue laser illumination, 488nm blue laser with 40% laser power (high power) was applied to the embryos heterozygous for either *dl-mCherry* or *dl-mCherry-BLID*; while also applying 555nm laser with 1.8% laser power to monitor mCherry signal. Images were taken in 14 Z-planes of interval distance 2.28 $\mu$ m.

To test whether high levels of Dl are continuously required at blastoderm stage, staging embryos and overall development of embryos were tested by imaging His2Av-eBFP2 excited at 0.8% of 405nm laser power (low power) between 28 Z-planes separated by 2.28 $\mu$ m. For 488nm

laser illumination (high power, e.g. early illumination started at nc14a, whereas late illumination started at nc14c), 40% of laser power was utilized with 33 Z-planes separated by 2.28 $\mu$ m while imaging His2Av-eBFP2 for 20 minutes. After blue laser illumination, His2Av-eBFP2 was imaged with the initial His2Av-eBFP2 settings up to gastrulation. For the dark condition, the embryos were imaged with His2Av-eBFP2 setting without 488nm laser illumination from the onset of nc14 to gastrulation (Lim et al. 2018).

To test *sna* transcriptional activities, the MS2-MCP system was used (Bothma et al. 2015) in combination with *dl-BLID* to optogenetically manipulate DI levels and assay target gene expression live. To both detect *sna*.MS2-MCP.GFP signals and degrade DI-BLID, a 488nm blue laser was used for both purposes but using different laser power: 5% (intermediate level) and 15% (high power), respectively. To distinguish this MS2-MCP imaging scheme from standard approach (i.e. Fig. 2A), we refer to MS2-MCP imaging laser early and laser late as “mLE” and “mLL” with exact conditions outlined in Fig. S3A.

To image Twi protein dynamics, Twi-mCherryLlamaTag system, which recognizes maternally-deposited mature mCherry fluorescent protein, was utilized (Bothma et al. 2018). mCherry was imaged live from the onset of nc14a to gastrulation while DI-BLID was degraded by 488nm blue laser at 40% laser power (high power) with varying lengths of time at appropriate developmental stages. (Bothma et al. 2018). To image mCherry proteins bound by Twi-mCherryLlamaTag, the fluorescent proteins were excited at 555nm with 5% of laser power. Images were taken in 30 Z-planes separated by 3 $\mu$ m. All images were taken using a 25x water immersion objective.

MS2-MCP.GFP foci were quantified using custom MATLAB functions. Images of MS2-MCP.GFP were first Z-projected and then segmented using a gaussian filter to smooth the image (standard deviation of 1 was used), then using Top-hat filtering to remove background (a disk

structuring element with a radius of 3 pixels was used). The resulting image was then segmented using a threshold. The threshold was determined using otsu's method on several of the images acquired, and then choosing a threshold that gave the best segmentation under different conditions. A threshold of 0.06 (on a scale of 0 to 1) was used, however threshold of 0.04 and 0.08 were also tested. Segmentation was verified manually. A threshold of 0.04 generally captured all foci but also included numerous regions where no real signal was present. A threshold of 0.08 generally did not include any regions without real signal, but also failed to include regions with clear signal. A threshold of 0.06 generally included most regions with real foci and the fewest regions without real foci. The number of foci or spots was determined by counting the number of unique (non-touching) regions detected. Although the segmentation is not perfectly accurate (including false positives, false negatives, and any foci that could not be separated), the differences in the number of spots detected between conditions were quite large, and any error introduced by false positives or negatives is likely negligible. The time of each acquisition was determined from the metadata of the image file. The number of spots detected was averaged for replicates. Blue light illumination time frames were noted manually and the time was determined from the corresponding time frames. The start and end timepoints of blue light illumination were averaged among all lines appearing on a single plot. Individual plots are listed in Supplementary Fig. S4.

Levels of mCherry associated with Twi-mCherryLlamaTag were quantified using custom MATLAB functions by first making a Z-projection, and then drawing a ROI within the Twi domain. The raw signal was calculated by taking the mean intensity of the ROI for each timepoint using the same ROI. Background levels were calculated by taking the mean of an ROI drawn outside of the Twi domain. The raw signal was normalized by first subtracting the background levels and then dividing by the background levels. The time of each acquisition was determined from the metadata of the image file. To align the timepoints for all images taken, the frame where

germ band elongation is observed was determined manually, and this timepoint was set to zero for each line by subtracting the time at that timepoint from all the other timepoints. This was done individually for each image series. The quantifications from replicates were averaged together by taking the mean of the normalized intensity values at each timepoint. Since the time for similar timepoints were not completely identical, but were very similar, the time was averaged as well. Standard error of the mean was determined and plotted as error bars. The frames where the embryo was undergoing blue light illumination were determined manually for each image series, and the time window of blue light was calculated as the starting time of the first frame where the embryo was illuminated to the start of the first frame after illumination ended. The time window of blue light illumination for the averaged normalized intensities were taken as the average of the start and end points of blue light illumination for the individual image series. Individual plots are listed in Supplementary Fig. S5.

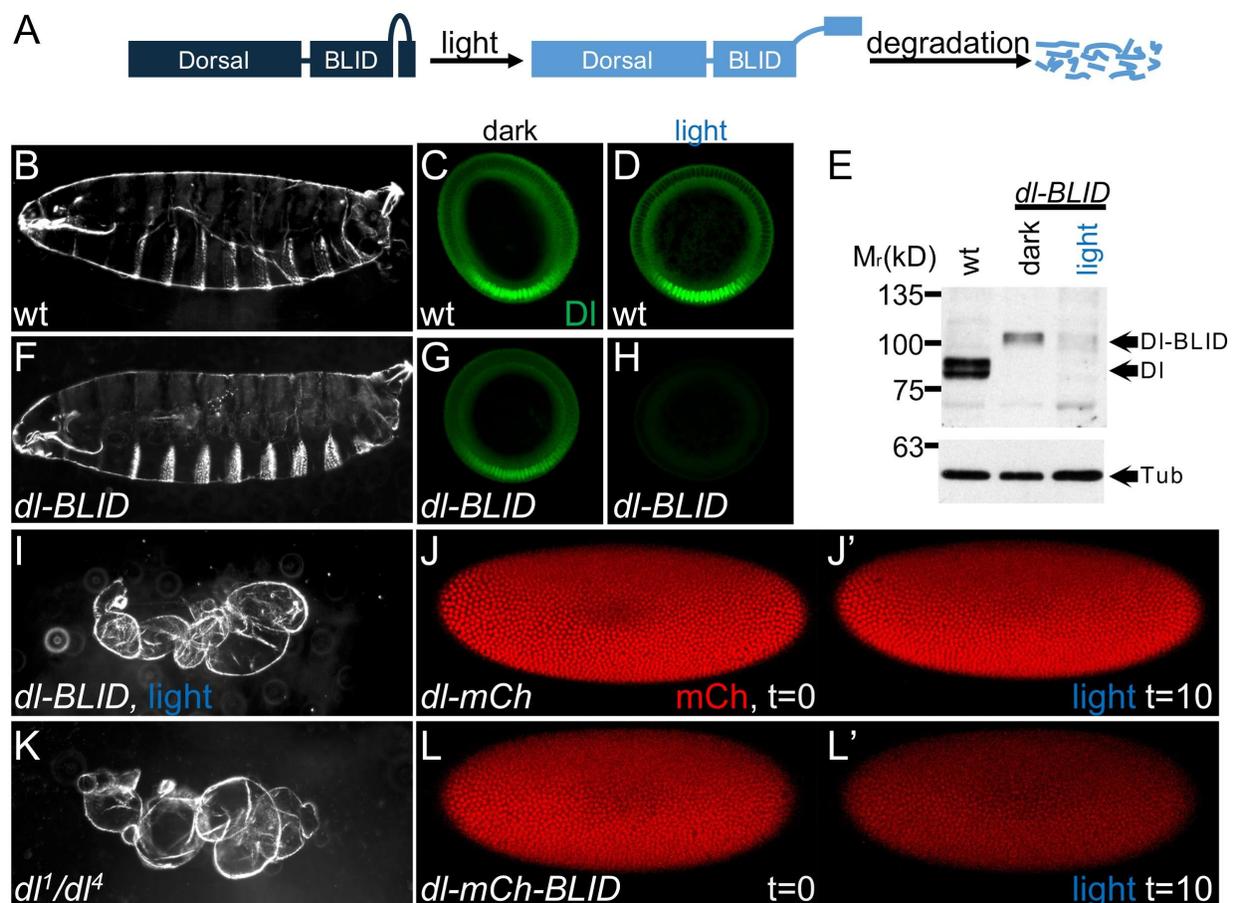
## REFERENCES

- Baaske J, Gonschorek P, Engesser R, Dominguez-Monedero A, Raute K, Fischbach P, Müller K, Cachat E, Schamel WWA, Minguet S, et al. 2018. Dual-controlled optogenetic system for the rapid down-regulation of protein levels in mammalian cells. *Sci Rep* **8**: 15024.
- Boettiger AN, Levine M. 2013. Rapid transcription fosters coordinate snail expression in the *Drosophila* embryo. *Cell Rep* **3**: 8–15.
- Bonger KM, Rakhit R, Payumo AY, Chen JK, Wandless TJ. 2014. General method for regulating protein stability with light. *ACS Chem Biol* **9**: 111–115.
- Bothma JP, Garcia HG, Ng S, Perry MW, Gregor T, Levine M. 2015. Enhancer additivity and non-additivity are determined by enhancer strength in the *Drosophila* embryo. *Elife* **4**. <http://dx.doi.org/10.7554/eLife.07956>.
- Bothma JP, Magliocco J, Levine M. 2011. The snail repressor inhibits release, not elongation, of paused Pol II in the *Drosophila* embryo. *Curr Biol* **21**: 1571–1577.
- Bothma JP, Norstad MR, Alamos S, Garcia HG. 2018. LlamaTags: A Versatile Tool to Image Transcription Factor Dynamics in Live Embryos. *Cell* **173**: 1810–1822.e16.
- Cowden J, Levine M. 2002. The Snail repressor positions Notch signaling in the *Drosophila* embryo. *Development* **129**: 1785–1793.
- Crews ST, Pearson JC. 2009. Transcriptional autoregulation in development. *Current Biology* **19**: R241–R246. <http://dx.doi.org/10.1016/j.cub.2009.01.015>.
- Dunipace L, Ozdemir A, Stathopoulos A. 2011. Complex interactions between cis-regulatory modules in native conformation are critical for *Drosophila* snail expression. *Development* **138**: 4075–4084.
- Ferguson EL, Anderson KV. 1992. Decapentaplegic acts as a morphogen to organize dorsal-ventral pattern in the *Drosophila* embryo. *Cell* **71**: 451–461.
- Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, O'Connor-Giles KM. 2014. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. *Genetics* **196**: 961–971.
- Gurdon JB, Harger P, Mitchell A, Lemaire P. 1994. Activin signalling and response to a morphogen gradient. *Nature* **371**: 487–492.
- Gurdon JB, Mitchell A, Mahony D. 1995. Direct and continuous assessment by cells of their position in a morphogen gradient. *Nature* **376**: 520–521. <http://dx.doi.org/10.1038/376520a0>.
- Huang A, Amourda C, Zhang S, Tolwinski NS, Saunders TE. 2017. Decoding temporal interpretation of the morphogen Bicoid in the early embryo. *Elife* **6**. <http://dx.doi.org/10.7554/eLife.26258>.

- Ip YT, Park RE, Kosman D, Yazdanbakhsh K, Levine M. 1992. Dorsal-twist interactions establish snail expression in the presumptive mesoderm of the *Drosophila* embryo. *Genes Dev* **6**: 1518–1530.
- Kosman D, Ip YT, Levine M, Arora K. 1991. Establishment of the mesoderm-neuroectoderm boundary in the *Drosophila* embryo. *Science* **254**: 118–122.
- Kosman D, Mizutani CM, Lemons D, Cox WG, McGinnis W, Bier E. 2004. Multiplex detection of RNA expression in *Drosophila* embryos. *Science* **305**: 846.
- Leptin M, Grunewald B. 1990. Cell shape changes during gastrulation in *Drosophila*. *Development* **110**: 73–84.
- Liberman LM, Reeves GT, Stathopoulos A. 2009. Quantitative imaging of the Dorsal nuclear gradient reveals limitations to threshold-dependent patterning in *Drosophila*. *Proc Natl Acad Sci U S A* **106**: 22317–22322.
- Liberman LM, Stathopoulos A. 2009. Design flexibility in cis-regulatory control of gene expression: synthetic and comparative evidence. *Dev Biol* **327**: 578–589.
- Lim B, Heist T, Levine M, Fukaya T. 2018. Visualization of Transvection in Living *Drosophila* Embryos. *Mol Cell* **70**: 287–296.e6.
- McDaniel SL, Gibson TJ, Schulz KN, Fernandez Garcia M, Nevil M, Jain SU, Lewis PW, Zaret KS, Harrison MM. 2019. Continued Activity of the Pioneer Factor Zelda Is Required to Drive Zygotic Genome Activation. *Mol Cell* **74**: 185–195.e4.
- Ozdemir A, Fisher-Aylor KI, Pepke S, Samanta M, Dunipace L, McCue K, Zeng L, Ogawa N, Wold BJ, Stathopoulos A. 2011. High resolution mapping of Twist to DNA in *Drosophila* embryos: Efficient functional analysis and evolutionary conservation. *Genome Research* **21**: 566–577. <http://dx.doi.org/10.1101/gr.104018.109>.
- Ozdemir A, Ma L, White KP, Stathopoulos A. 2014. Su(H)-Mediated Repression Positions Gene Boundaries along the Dorsal-Ventral Axis of *Drosophila* Embryos. *Developmental Cell* **31**: 100–113. <http://dx.doi.org/10.1016/j.devcel.2014.08.005>.
- Perry MW, Boettiger AN, Bothma JP, Levine M. 2010. Shadow enhancers foster robustness of *Drosophila* gastrulation. *Curr Biol* **20**: 1562–1567.
- Port F, Chen H-M, Lee T, Bullock SL. 2014. Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. *Proc Natl Acad Sci U S A* **111**: E2967–76.
- Reeves GT, Stathopoulos A. 2009. Graded dorsal and differential gene regulation in the *Drosophila* embryo. *Cold Spring Harb Perspect Biol* **1**: a000836.
- Reeves GT, Trisnadi N, Truong TV, Nahmad M, Katz S, Stathopoulos A. 2012. Dorsal-ventral gene expression in the *Drosophila* embryo reflects the dynamics and precision of the dorsal nuclear gradient. *Dev Cell* **22**: 544–557.
- Roth S, Stein D, Nüsslein-Volhard C. 1989. A gradient of nuclear localization of the dorsal protein

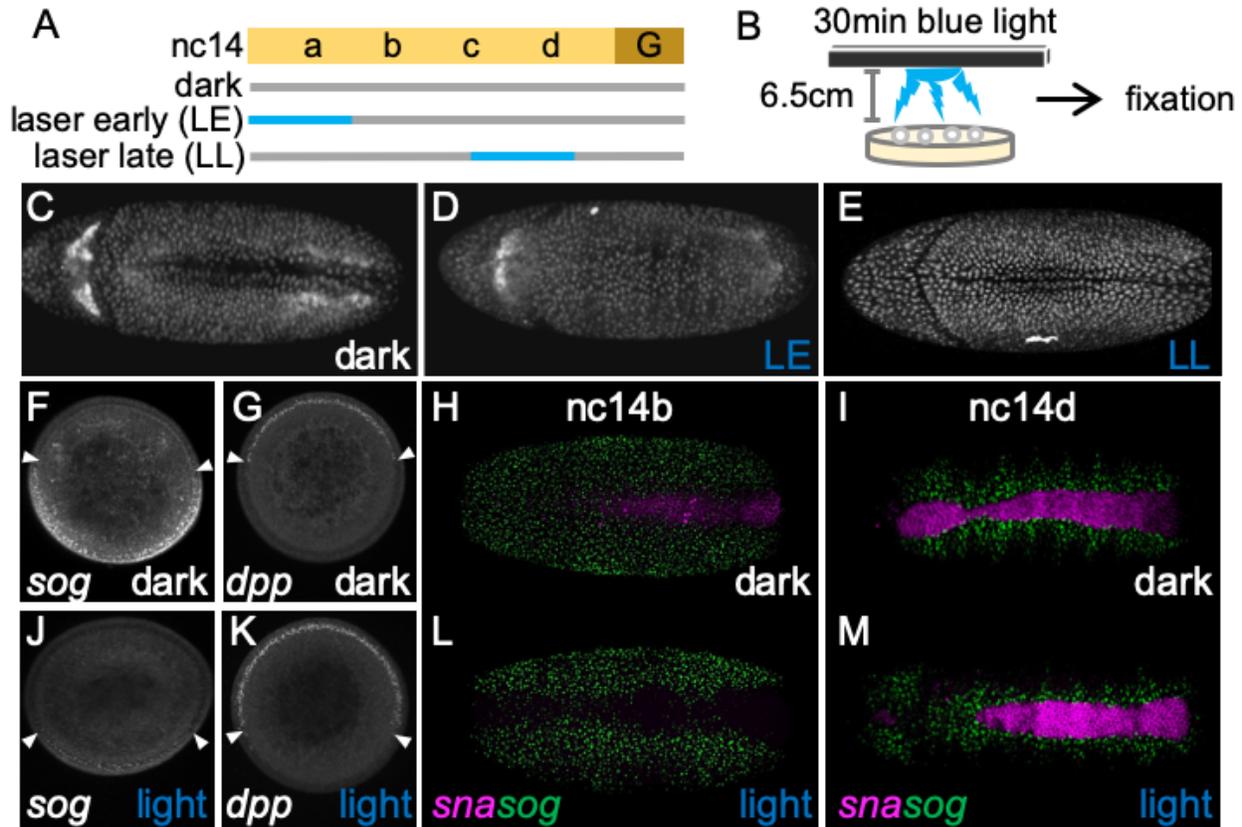
- determines dorsoventral pattern in the *Drosophila* embryo. *Cell* **59**: 1189–1202.
- Rushlow CA, Shvartsman SY. 2012. Temporal dynamics, spatial range, and transcriptional interpretation of the Dorsal morphogen gradient. *Curr Opin Genet Dev* **22**: 542–546.
- Sandler JE, Stathopoulos A. 2016a. Quantitative Single-Embryo Profile of *Drosophila* Genome Activation and the Dorsal–Ventral Patterning Network. *Genetics* **202**: 1575–1584. <http://dx.doi.org/10.1534/genetics.116.186783>.
- Sandler JE, Stathopoulos A. 2016b. Stepwise Progression of Embryonic Patterning. *Trends Genet* **32**: 432–443.
- Stathopoulos A, Levine M. 2002. Linear signaling in the Toll-Dorsal pathway of *Drosophila*: activated Pelle kinase specifies all threshold outputs of gene expression while the bHLH protein Twist specifies a subset. *Development* **129**: 3411–3419.
- Trisnadi N, Stathopoulos A. 2014. Ectopic expression screen identifies genes affecting *Drosophila* mesoderm development including the HSPG Trol. *G3* **5**: 301–313.
- Zeitlinger J, Zinzen RP, Stark A, Kellis M, Zhang H, Young RA, Levine M. 2007. Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the *Drosophila* embryo. *Genes & Development* **21**: 385–390. <http://dx.doi.org/10.1101/gad.1509607>.

## FIGURES &amp; LEGENDS



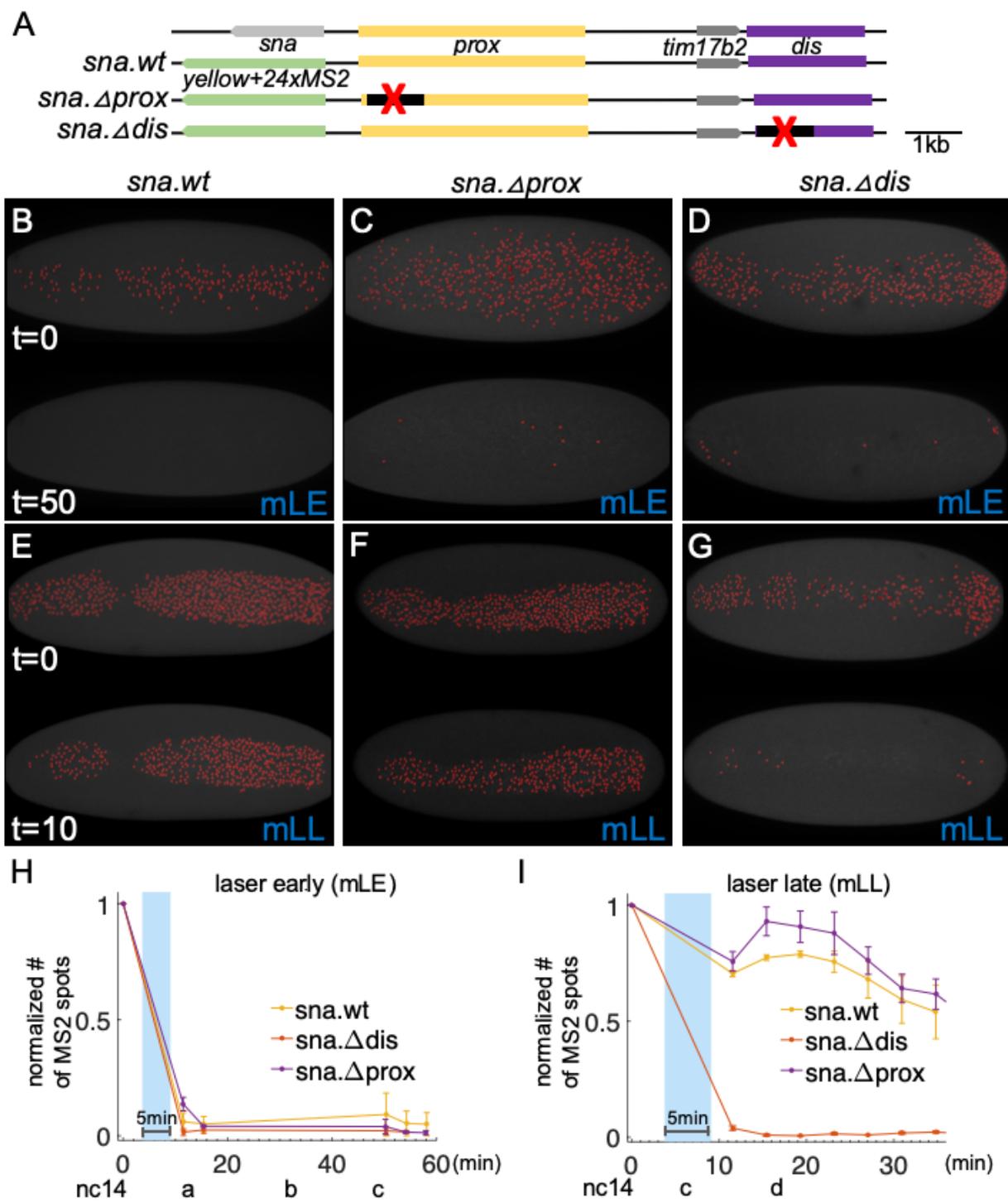
**Figure 1. Illumination with blue light induces degradation of DI-BLID fusion proteins.** (A) The DI-BLID construct. Blue light illumination causes a degradation sequence to be exposed, resulting in the degradation of the entire fusion protein. (B,F,I,K) Cuticle preparations of embryos derived from wildtype mothers without illumination (B, n=181/190), *dl-BLID* mothers without illumination (F, n=147/310), *dl-BLID* mothers with four hours of blue LED illumination (I, n=31/36), and *dl* null mutant (*dl<sup>1</sup>/dl<sup>4</sup>*) mothers without illumination (K, 142/142). (C,D,G,H) Manually cross sectioned embryos stained with anti-DI antibody (green) derived from wildtype

mothers without illumination (C, n=5/5), wildtype mothers with 1 hour of blue LED illumination (D, n=5/5), *dl-BLID* mothers without illumination (G, n=5/5), and *dl-BLID* mothers with one hour of blue LED illumination (H, n=4/5). All embryos in C,D,G,H were imaged at the same settings, demonstrating a clear decrease in D1 levels in H. (E) Western blot of wildtype (lane 1), *dl-BLID* without illumination (lane 2), and *dl-BLID* with 30 min blue LED illumination (lane 3). Top blot is probed with anti-D1 antibody. Bottom blot is probed with anti-Tubulin antibody to serve as a loading control. Arrows indicate the approximate locations of D1, D1-BLID, and Tubulin bands. (J,J',L,L') Snapshots from live imaging movies of *dl-mCherry* (n=1) and *dl-mCherry-BLID* (n=3) at the start (J, L t=0) and after 10 min of 40% power blue laser illumination (J', L' t=10). All embryos/larval cuticles are oriented with anterior to left and dorsal up, except cross sections which are oriented with the ventral side at the bottom and the dorsal side at the top.



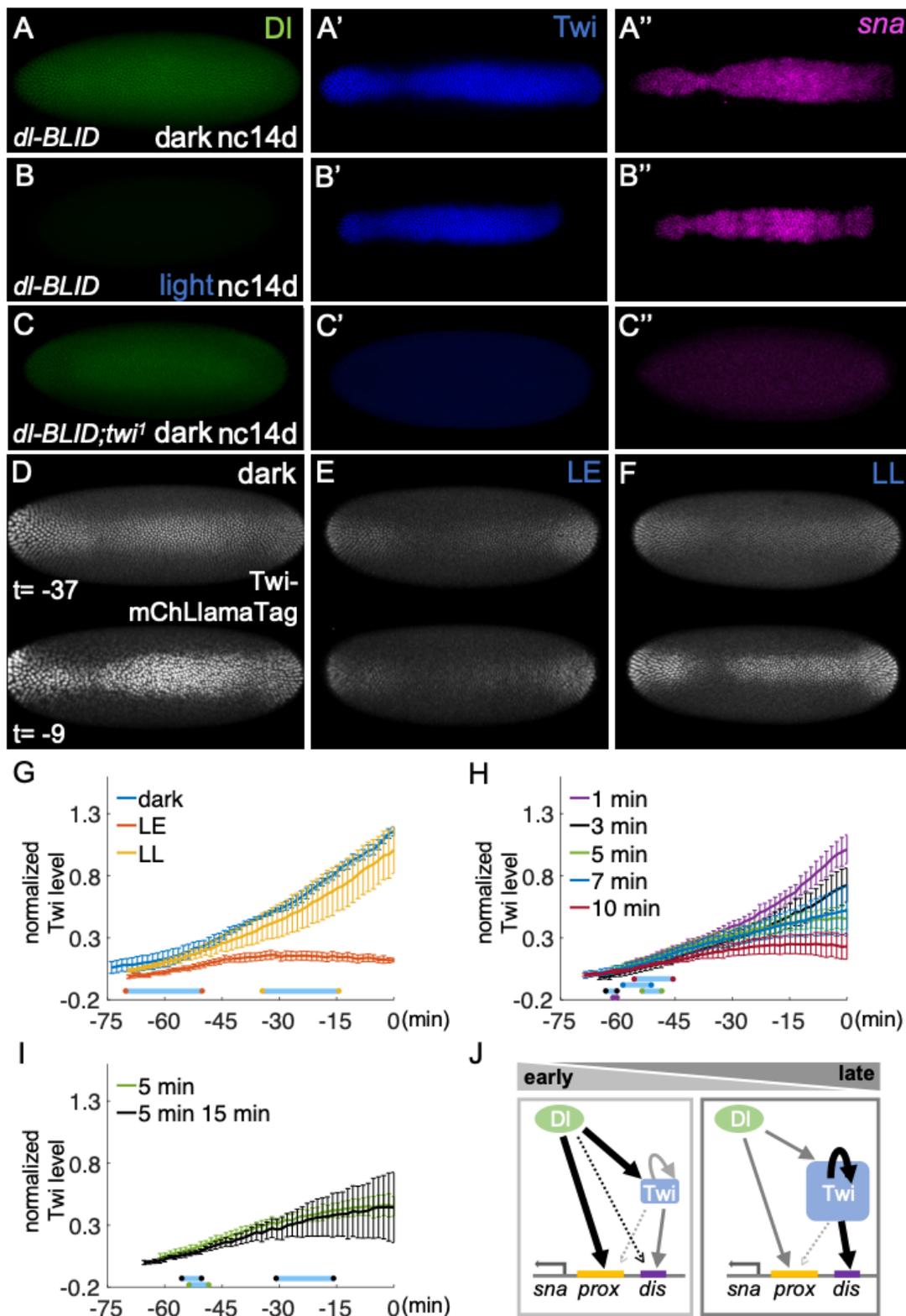
**Figure 2. High levels of D<sub>l</sub> at late stages are not required to support *sna* or gastrulation.** (A) Scheme of 20 min 40% blue laser illumination on single embryos using a confocal microscope. Gray bar: 0.8% (low power) 405nm laser to image H2A.BFP. Blue bar: 0.8% 405nm laser and 20 min 40% (high power) 488nm laser. (B) Scheme of 30 min blue LED illumination on a batch of embryos, which was followed by immediate fixation. (C-E) *dl-BLID* embryos at stage 6 illuminated using a laser (see A). Embryos (C) kept in the dark (“dark”, n=2), (D) with 20 min blue laser early illumination (“LE”, n=3), and (E) with 20 min blue laser late illumination (“LL”, n=3). (F,G,J,K) Manually cross-sectioned *nc14a* embryos stained for *sog* (F,J) or *dpp* (G,K) transcripts kept in the dark (F,G), or with 30 min blue LED illumination (“light”, J,K). White arrowheads mark the expression boundaries. (H,I,L,M) *sna* (purple) and *sog* (green) transcript expressions were assayed in *dl-BLID* embryos kept in the dark (H: n=4,L: n=6) or illuminated with a blue LED (see B) for 30 min (I: n=7,M: n=6). The stages of embryos at fixation were *nc14b*

(H,L) and nc14d (I,M). All whole mount images are a ventral view with anterior to the left. Cross-sectioned embryos are aligned with the ventral side at the bottom and the dorsal side at the top.



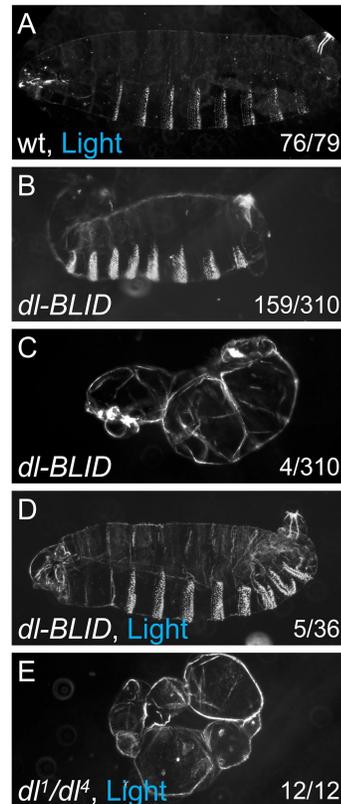
**Figure 3. High level of Dl is required for *sna* activation only at early stages, but not at late stages, in which *sna* expression is predominantly supported by the *sna* distal enhancer.**

(A) Scheme of large reporter constructs used to assay *sna* transcriptional activities by MS2-MCP system (Bothma et al. 2015). (B-G) MCP.GFP signals associated with the *sna* MS2 reporter were imaged (false-colored red dots) in *dl-BLID* with early (B-D) or late (E-G) blue laser illumination that is MS2-MCP imaging compatible (“mLE” and “mLL”, respectively; see also Fig. S3) in various *sna* regulatory conditions including wildtype (*sna.wt* B,E), proximal enhancer deletion (*sna.Δprox* C,F), and distal enhancer deletion (*sna.Δdis* D,G). Images are snapshots from movies, before illumination (top) and after illumination (bottom) of each panel. Three movies were taken for each condition. Ventral views of embryos are shown with anterior oriented to the left. (H,I) Quantitative analysis of the number of MCP.GFP dots associated with the *sna* MS2 reporter in *dl-BLID* embryos with *sna.wt*, *sna.Δprox*, or *sna.Δdis* *sna* regulatory condition. Number of MS2-MCP.GFP spots are counted in each time frame, and the values are normalized to the initial value detected in the first frame (before 5min blue laser illumination with 15% laser power) with early laser (H) or late laser (I) illumination. Blue shade indicates a time frame of 5min 15% blue laser illumination. Error bars represent standard error of the mean. For individual traces, see Supplemental Fig. S5. For details for detection of *sna*.MS2-MCP.GFP and blue laser illumination, see Supplementary Fig. S3.

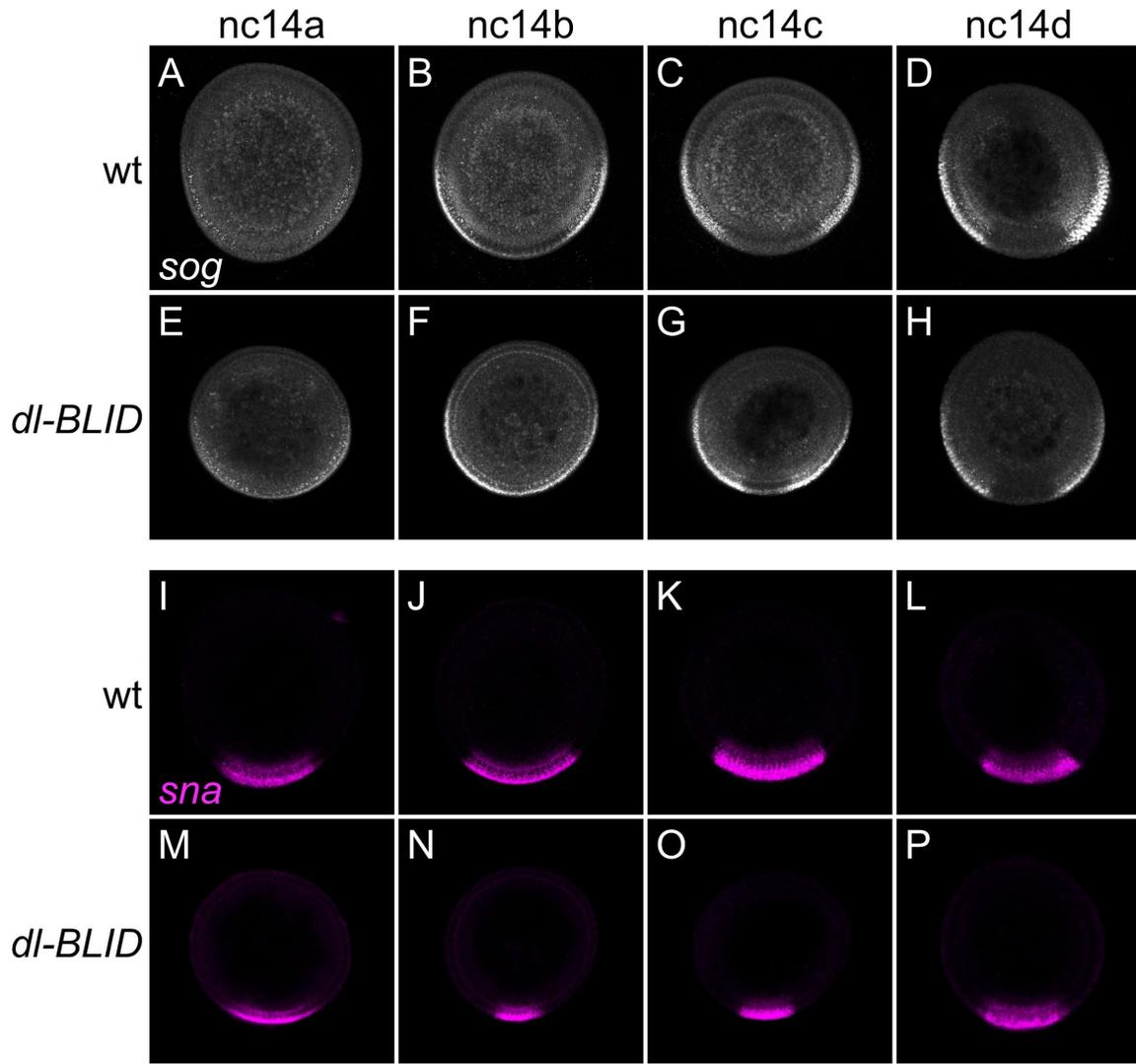


**Figure 4. Twi suffices to support *sna* expression at late stages in the absence of high levels of Dl.** (A-C) Expression of Dl proteins (green), Twi proteins (blue), and *sna* transcripts (pink) were examined in *dl-BLID* embryos with *twi* wildtype (A,B) or *twi<sup>l</sup>* mutant background at nc14d without (A,C) or with (B, n=11) 30min blue LED illumination. (D-F) Snapshots from movies showing mCherry signal associated with the Twi-mCherryLlamaTag (Twi-mChLlamaTag; Bothma et al. 2018) under various confocal 40% blue laser illumination conditions: (D) no illumination (dark), (E) 20min early illumination at nc14a (LE), and (F) 20min late illumination at nc14c (LL). Time indicates the time length preceding the germband extension. (G-I) Quantitative analysis of the levels of mCherry associated by Twi-mChLlamaTag with varying 40% blue laser illumination conditions: (G) no illumination (dark, blue), 20 min early illumination at nc14a (LE, red), and 20 min late illumination at nc14c (LL, yellow); (H) illumination at nc14b for: 1 min (purple), 3min (black), 5min (green), 7min (blue) and 10min (red); and (I) the 5 min (green) data replotted from H to compare with 5min at nc14b followed by additional 15 min illumination at nc14c (black). Three movies were taken for each condition. For the individual traces, see Supplemental Fig. S6. (J) A model of regulatory shift, such as from high level of Dl to high level of Twi dependent regulatory states, to support *sna* expression throughout early embryonic development. Dl proteins (green circle), Twi proteins (Blue square), *sna* proximal enhancer (yellow bar), and *sna* distal enhancer (purple bar). All embryo images are ventral views with anterior to the left. Blue bars in G-I represent the average time window of confocal blue laser illumination to their respective curves. Error bars represent standard error of the mean.

## SUPPLEMENTARY FIGURES &amp; LEGENDS

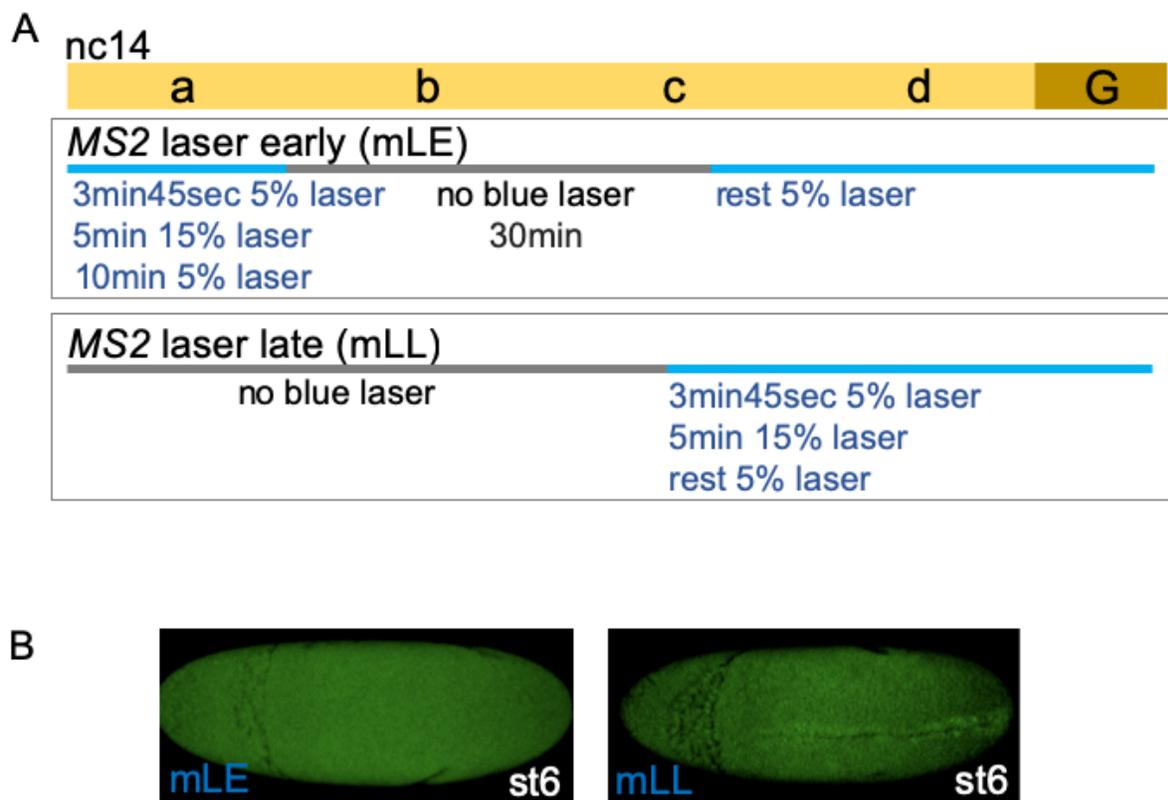


**Figure S1. Blue light has no effect on the cuticles of wildtype or *dl* mutant.** Larval cuticles with anterior to the left and dorsal side up. (A) Larval cuticles from wildtype mothers after four hours of blue LED illumination (n = 76/79). (B) An example of cuticle from *dl-BLID* mothers without illumination that exhibit abnormal cuticles (n=159/310). (C) Cuticles from *dl-BLID* mothers without illumination that appear lateralized or dorsalized (n=4/310). (D) A representative image of cuticles from *dl-BLID* mothers with four hours of blue LED illumination that did not exhibit lateralized or dorsalized cuticles (n=5/36). (E) Cuticles from *dl¹* (*dl¹/dl⁴*) mothers with four hours of blue LED illumination that are dorsalized (n = 12/12).



**Figure S2. Expression of *sog* and *sna* transcripts during nc14 for wt and *dl-BLID* embryos kept in the dark.** Manually crossed sectioned embryos co-stained for *sog* and *sna* transcripts using FISH. (A-D) *sog* expression in wt embryos kept in the dark at 14a (A), 14b (B), 14c (C), 14d (D). (E-H) *sog* expression in *dl-BLID* embryos kept in the dark at approximately the same stage as those in A-D. (I-L) *sna* expression in wt embryos kept in the dark, from the same embryos in A-D. (M-P) *sna* expression in *dl-BLID* embryos kept in the dark, from the same embryos in E-H.

Note the variability in the width of *sna* expression in *dl-BLID*, and also the differences in *sna* width between wt and *dl-BLID* (J compared with N). *sog* expression appears similar between wt and *dl-BLID*.



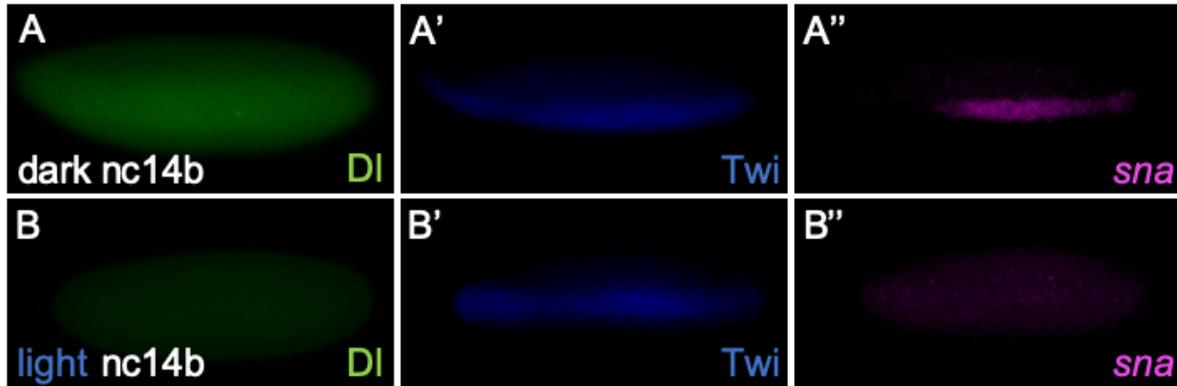
**Figure S3. Scheme of early laser exposure or late laser exposure to degrade DI-Blid while detecting *sna* transcriptional activities using MS2-MCP.GFP system.**

(A) Scheme of “MS2-MCP.GFP compatible Laser Early” (mLE) or “MS2-MCP.GFP compatible Laser Late” (mLL) exposures in *dl-BLID* recombined with *sna.MS2*. 5% laser power (intermediate power) was used to image MCP.GFP signals associating *sna.MS2*, while 15% (high power) laser power was used to degrade most DI-BLID. In mLE scheme, at nc14a, signals of MCP.GFP associating *sna.MS2* were imaged with 5% laser power during the first time frame to image the initial state of *sna* transcriptional activities. Then, a blue laser with 15% power was applied to the embryo to degrade DI-BLID for 5 min, followed by imaging MCP.GFP signals with 5% laser power for the next ten min. To avoid further DI-BLID degradation while imaging MCP.GFP

signals, imaging with blue laser was stopped for 30 min. Finally, MCP.GFP signals were imaged using 5% laser power for the rest. In the mLL scheme, *dl-BLID* embryos recombined with *sna.MS2* were aged upto nc14c. First, MCP.GFP signal interacting with *sna.MS2* were imaged using blue laser with 5% laser power. Then, a blue laser with 15% power was applied to the embryo for 5min to degrade DI-BLID, followed by imaging MCP.GFP associated with *sna.MS2* using a blue laser with 5% power for the rest of development.

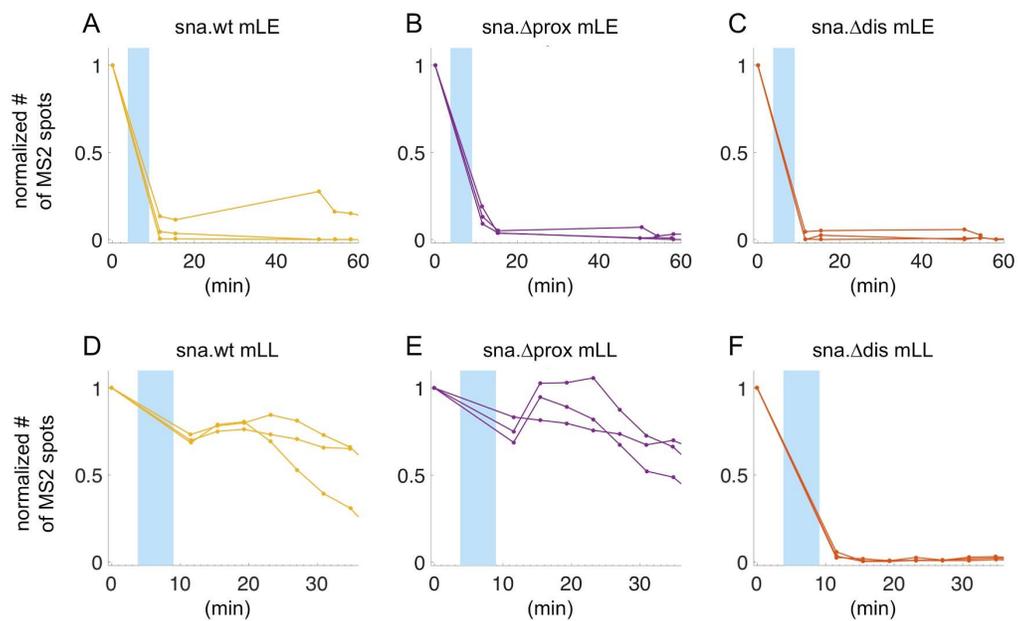
(B) Embryos at st6 to assay gastrulation defects after mLE (left) or mLL (right) blue laser illumination. Images of embryos in (B) are ventral views with anterior oriented to the left.

FIGURE S4

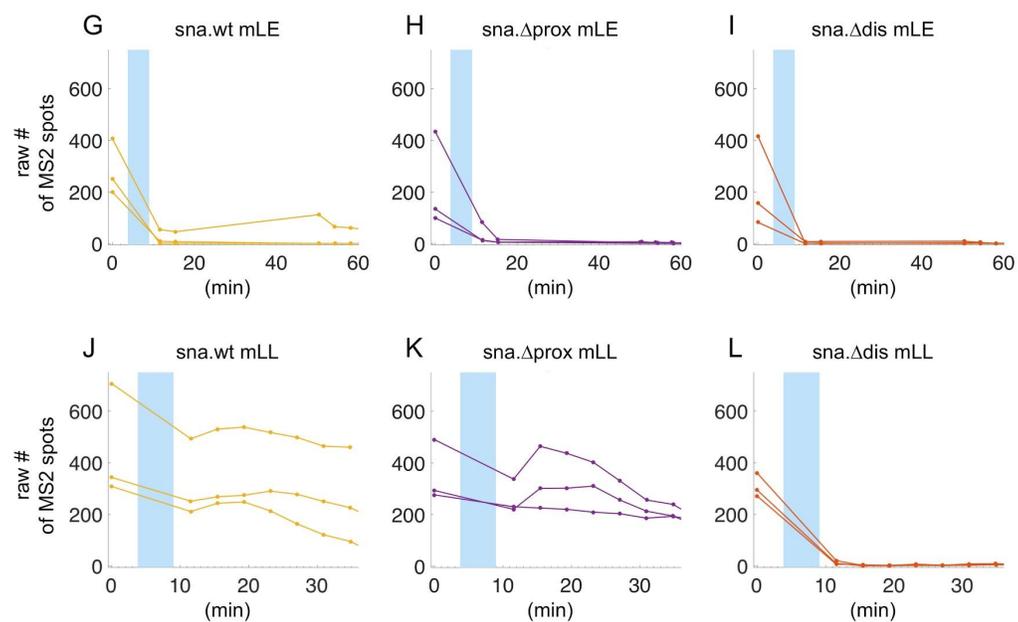


**Figure S4. Expression of D1 proteins, Twi proteins, and *sna* transcripts in *nc14b dl-BLID* embryos with or without 30 min blue LED illumination.** (A,B) *In situ* hybridization combined with immunostaining of *nc14b dl-BLID* embryos in the absence (dark, A-A'') or presence (light, B-B'') of 30 min of blue LED illumination to detect D1 protein (green), Twi protein (blue) and *sna* transcripts (pink). Images of embryos are ventral views with anterior oriented to the left.

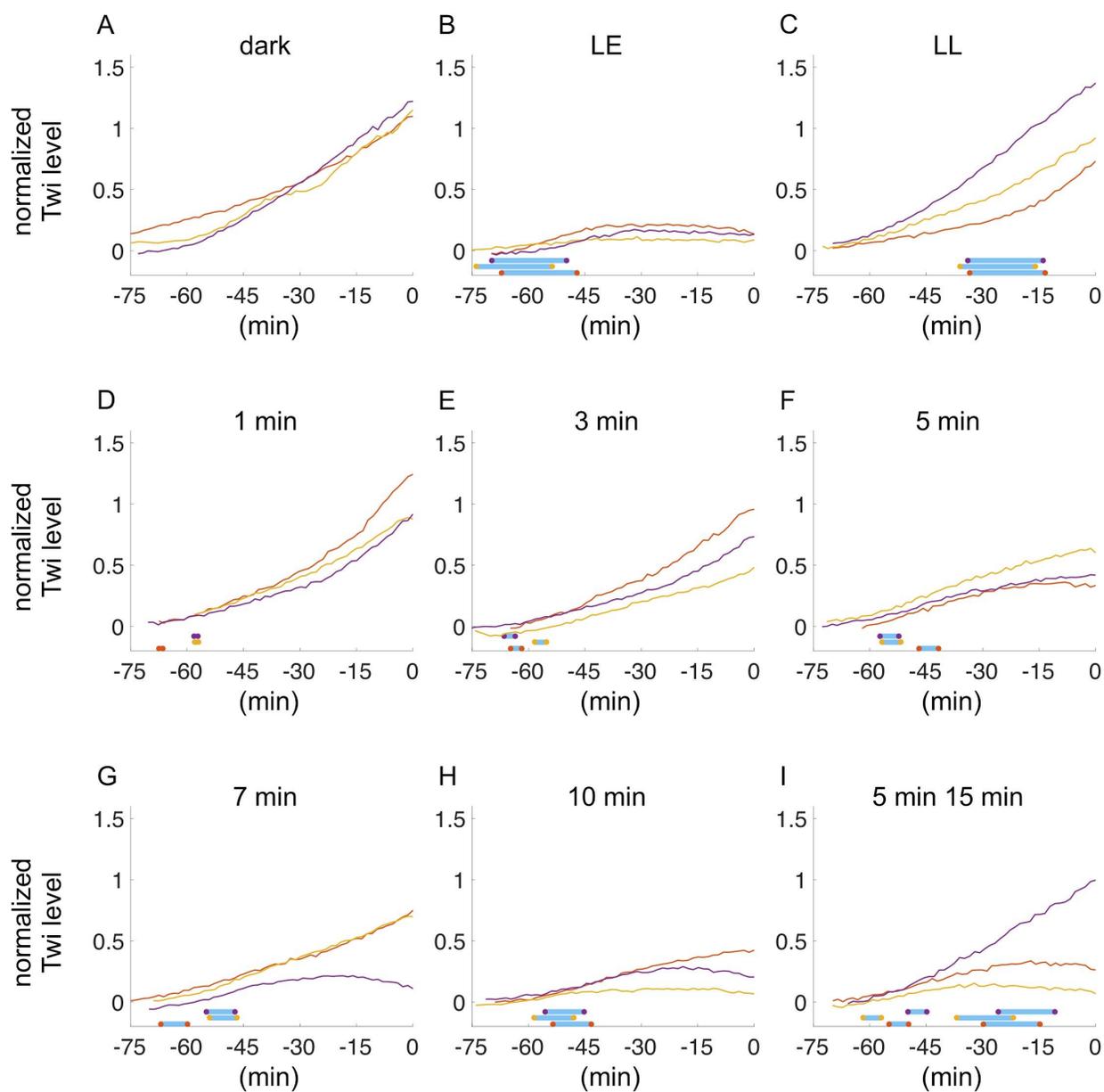
## Normalized Individual Curves



## Unnormalized Individual Curves

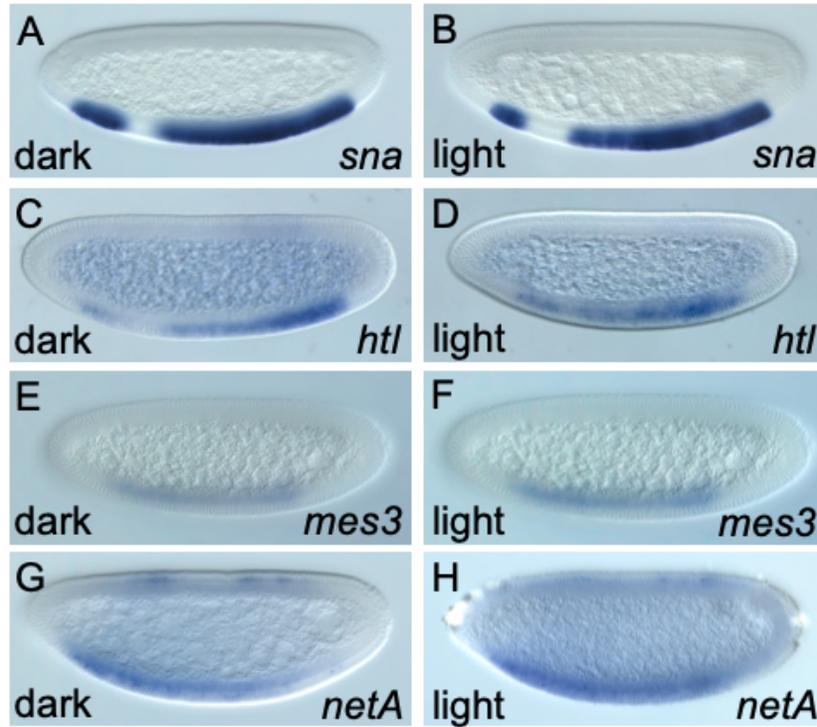


**Figure S5. Individual plots for the number of spots detected using MS2-MCP system for each of three assayed *sna* reporter genes.** (A-C) individual plots for the number of spots detected using MS2-MCP for the *sna* transcriptional reporters: (A) *sna.wt* (gold), (B) *sna.Δprox* (purple), and (C) *sna.Δdis* (red) with blue laser illumination applied early (mLE) and normalized by the number of spots detected in the first frame. (D-F) the same reporters and same normalization with blue laser illumination applied late (mLL). (G-I) The same data as in A-C but using the unnormalized number of spots. (J-L) The same data as in D-F but using the unnormalized number of spots. The blue windows represent the average of the start and end of blue light illumination for the lines on the respective plots, which are approximately 5 min in duration.



**Figure S6. Individual plots for the quantification of Twi using Twi-mChLlamaTag.** (A-I) individual plots for the quantification of Twi when embryos are (A) kept in the dark, (B) illuminated early (LE), (C) illuminated late (LL), (D) illuminated for 1min at nc14b, (E) illuminated for 3min at nc14b, (F) illuminated for 5min at nc14b, (G) illuminated for 7min at

nc14b, (H) illuminated for 10min at nc14b, and (I) illuminated for 5min at nc14b and illuminated again for 15min at nc14c. Gold, purple, and red lines represent individual embryos. Colors are repeated for each condition. Blue bars represent the time when embryos are illuminated with blue light, and correspond to the line matching the color of the two dots at the end of the blue bar.



**Figure S7. Target gene expression in ventral regions of nc14d *dl-BLID* embryos with or without 30 min blue LED illumination.** Lateral view of nc14d embryos stained for mesoderm targets without (left column, dark) or with 30 min blue LED illumination (right column, light). (A,B) *sna*, (C,D) *htl*, (E,F) *mes3*, and (G,H) *netA*. Expression remains for each gene tested. Embryos are oriented with anterior to the left and dorsal side up.

**Table S1**

NAME	SEQUENCE	COMMENTS
	gRNA	gRNA uppercase
gRNA F	tatataggaagatatccgggtgaactcgAATCTGCTTAGCTTCGATAGgttttagagctagaaatagcaag	
gRNA R	attttaactgctatttctagctctaaaacACGTTCCAGATTTACAACGcgacgttaaattgaaaatagtc	
	HDR Left Homology (LH) Primers	dl uppercase
LH dl F	ATTgaattcCTCGCTTCGCTTTGTAGATA	EcoRI lowercase
LH dl R	ATTgctagcAAAATTTAATTTGCAATAAGATCG	NheI lowercase
dl R	acceccaccgctctcccCGTGGATATGGACAGGTTCCG	6x Gly linker lowercase; with LH dl F
	dl-BLID	BLID uppercase
dl-BLID 1 F	ggaggagcggtgggggtTTCTTGCTACTACACTTGAACG	6x Gly linker lowercase
dl-BLID 1 R	gttgtgaaaaggtaCTAACCTCGCCGCTTGC	dl lowercase
dl-BLID 2 F	GCAAGGCGGCGAGGTTAGtaccttttcacaacgaacc	dl lowercase; with LH dl R
	dl-mCherry	mCh uppercase
dl-mCh 1 F	ggaggagcggtgggggtATGGTGAGCAAGGGCGAGGA	6x Gly linker lowercase
dl-mCh 1 R	ggttcgtgtgaaaaggtaTTACTTGACAGCTCGTCCATGC	dl lowercase
dl-mCh 2 F	TGGACGAGCTGTACAAGTAAaccttttcacaacgaacc	dl lowercase; used with LH dl R
	dl-mCherry-BLID	3x Gly linker lowercase
dl-mCh-BLID 1 R	AAGAAacceccaccCTTGTACAGCTCGTCCATGC	BLID, mCh uppercase; with LH dl F
dl-mCh-BLID 2 F	ACAAGggtgggggtTTCTTGCTACTACACTTGAACG	mCh, BLID uppercase; with LH dl R

---

HDR Right Homology (RH) Primers

RH F	ATTagatctTTTTGTTAATACTGTTATAAAAGATCC	BglII lowercase
RH R	ATTctegagCAAAGGCAAAGATTAGGAAA	XhoI lowercase

---



---

Sequencing Primers

F RH	TCACTGCATTCTAGTTGTGGT	in pHD-DsRed
R RH	CGCCCTTGAACCTCGATTGAC	in pHD-DsRed
F LH	GATGGTAGTGTGGGGACTCC	in pHD-DsRed
F1 dl LH	CCCACCAACAACAATGCCAA	in dl
R LH	GCCTCTATTTATACTCCGGCG	in pHD-DsRed
R BLID	TCTGGCAATCTTGGGTCAGT	in BLID
F2 dl LH	GCCATCGAGCAACTACAACC	in dl
R mCh	CATGTTATCCTCCTCGCCCT	in mCh

---



---

Primers for Probe Synthesis

sog int1 F	ATCTATTGCGCTCGTTGCTT
sog int1 R	AATTAATACGACTCACTATAGGGTTGCACAAAATGCCACAAAT
NetA int F	CCATCCTTCGCGTCCATCCC
NetA int R	AATTAATACGACTCACTATAGGGCCAAAACCAAGCGAACGCC
dpp F	ccagaactagaaaaccggaagc
dpp R	gaaatTAATACGACTCACTATAgggCGCCTGTGCTAAAGACCCTG

sog ex1 F TCAGGTCAGTCGCTCTTGA

sog ex1 R AATTAATACGACTCACTATAGGGGTGTCGGACTCCTCGAACAT

---

## Chapter 3

High Dorsal levels downregulate, not promote, ventral gene expression in *Drosophila* embryos by regulating enhancer action

## ABSTRACT

Morphogens instruct spatial expression of target genes, and morphogens are also dynamic therefore it is unclear whether this relationship holds over time or if they also have additional roles. For instance, genes expressed along the dorsal-ventral (DV) axis of the *Drosophila* embryo are responsive to the concentration of transcription factor Dorsal (Dl). However, levels of Dl steadily increase from one nuclear cycle (nc) to the next in this syncytial embryo, and Dl is required only early but not late in nc14 to support expression of genes in ventral regions like *snail* (*sna*). Here we studied Dl's dynamic action as it was unclear why this factor continues to build if expression of high-threshold targets like *sna* is already supported. Using a photosensitive degron to assay Dl's temporal role, we used MS2-MCP live imaging to study *sna* expression, specifically through the *sna.prox* enhancer. Our data demonstrate that Dl binding within the *sna.prox* enhancer functions as a molecular damper to limit activity of the *sna.dis* enhancer, to surprisingly downregulate *sna* gene expression total output when Dl levels are maximal.

## RESULTS & DISCUSSION

To assay the role of *sna.prox* and *sna.dis* in the native context of the endogenous *sna* gene locus, CRISPR/Cas9 was used to create deletions of genomic sequences associated with each enhancer. When either a 4.4kB segment encompassing the entire proximal enhancer or a 2kb region encompassing the distal enhancer (Fig. 1A,  $\Delta prox4.4$  or  $\Delta dis2.0$ , respectively) is deleted, *sna* expression is similar to wildtype (Fig. 1C, D, compared with B). While similar results were obtained with analysis of enhancer deletions in the context of large reporter constructs encompassing ~25 kB surrounding *sna* locus, deletion of the distal enhancer in the context of a large reporter was associated with derepression such that expression was expanded (Dunipace et

al., 2011). We reasoned that the reporter constructs may be more sensitive to loss of repression, for example, due to insertion at a genomic position permissive to activation.

To provide insight, we created a mutant that modulates activator-repressor balance. Available whole-genome chromatin immunoprecipitation data demonstrate input to the *sna.dis* enhancer by repressors Hucklebein (Hkb) and Suppressor of Hairless [Su(H)] as well as maternal activator Zelda (Zld) (Harrison et al., 2011; MacArthur et al., 2009; Ozdemir et al., 2014). Deletion of a 1.8 kB portion allows removal of repressor input but leaves Zld input intact (Fig. 1A). In this  $\Delta dis1.8$  mutant, *sna* is expanded within the trunk domain (Fig. 1E, compared with 1B) as expected by loss of input from Su(H) repressors. This demonstrates that the balanced activator-repressor inputs are important for establishing boundaries (Fig. 1G, and that similar but not identical results are obtained from deletions at the native locus versus large reporter constructs.

Results in both contexts (i.e. large reporter and native locus), however, demonstrate that the two enhancers are required to support normal *sna* levels.  $\Delta prox4.4$  mutants exhibit an increase in *sna* levels; whereas, in contrast,  $\Delta dis2.0$  or  $\Delta dis1.8$  mutants exhibit an increase in extent of the spatial expression but a decrease in *sna* levels (Fig. 1F). Though Zld input is left predominantly intact, the 1.8kB deletion removes input from another zygotic activator, Twist (Twi) transcription factor, which is also important for the support of *sna* expression. We focused further analysis on deciphering how the *sna.dis* and *sna.prox* enhancers coordinate to regulate *sna* expression levels.

To investigate whether activities of the two enhancers are dependent on each other, a mutant line was generated, in which a 3kB portion of the proximal enhancer was deleted (Fig. 3A,  $\Delta prox3.0$ ). Since the proximal enhancer is dependent on high Df levels (Irizarry et al., 2020), we maintained most of the regions that are bound by Df (Fig. 2A) (MacArthur et al., 2009). Furthermore, to investigate if Df inputs on the proximal enhancer affect activity of the distal enhancer, predicted Df binding sites were mutated [Fig. 2A,  $\Delta prox3.0(mDf)$ ; Fig. S2]. The

*Δprox3.0* and *Δprox3.0(mDl)* mutants do not affect the *sna* domain of expression (Fig. 2B,C compared with 1B) but do impact levels of expression (Fig. 2E). When the putative DI binding sites are mutated, the levels of *sna* expression increase [Fig. 2E, compare *Δprox3.0(mDl)* to *Δprox3.0*].

Since the distal enhancer drives strong *sna* expression levels, whereas the proximal enhancer drives weak *sna* levels (Fig. S1), we hypothesized that increased *sna* levels in *Δprox3.0(mDl)* is due to enhanced distal enhancer action (i.e. DI binding in the proximal enhancer acts to downregulate distal enhancer activity). To test the idea, 1.8kB of the distal enhancer was deleted additionally from this background (Fig. 2A; i.e. *Δprox3.0(mdl)*, *Δdis1.8*). In this double deletion, no *sna* domain changes are apparent (Fig. 2D), but levels are significantly decreased relative to *Δprox3.0(mDl)* (Fig. 2E). This result demonstrates that the increased expression associated with mutation of DI sites within the proximal enhancer stems from distal enhancer increased action.

To further investigate this idea that DI negatively control *sna* levels, first, *sna* levels were measured by monitoring *sna* nascent transcription in embryos with reduced DI levels (i.e. *dl<sup>l</sup>* or *dl<sup>4</sup>* heterozygote). To monitor *sna* nascent transcription, reporter constructs encompassing the ~20kb *sna* locus were used in which the *sna* gene is replaced by reporter gene *yellow* including intronic MS2 RNA stem loop sequences (Bothma et al., 2015). When the reporter gene is transcribed, the RNA stem loops associated with these transcripts interact with maternally deposited MCP-GFP fusion protein. In this manner, GFP signal within nuclei facilitates monitoring of *sna* transcription live *in vivo*. To investigate *sna* levels in *dl<sup>l</sup>* or *dl<sup>4</sup>* heterozygote, this *sna* MS2 reporter construct was introduced into a *dl<sup>l</sup>/+* or *dl<sup>4</sup>/+* embryos. At the onset of nc14, the numbers of MS2-MCP.GFP<sup>+</sup> spots, which reflects the number of nuclei with active *sna*, in these heterozygotes are lower than in wildtype (Fig. S3A). This lower output is observed

throughout nc14, but ultimately results in a similar number of spots by the end of nc14 in comparison to wildtype (Fig. S3A).

The mean spot size, which is correlated to the rate of transcription (Koromila and Stathopoulos, 2019), is lower in  $dl^l$  or  $dl^t$  heterozygote than wildtype at early nc14 (i.e. nc14a). In contrast, at late nc14 (i.e. nc14c-d), whereas the mean spot size is maintained in the heterozygotes, it continuously decreases in wildtype; the result is a higher mean number spot size in the  $dl^l$  or  $dl^t$  heterozygote compared to wildtype at the end of nc14 (Fig. S3B).

Taken together, unexpectedly, Df levels and *sna* levels exhibit an anti-correlation. *sna* is considered a Df high-threshold target gene, meaning the highest levels of Df are required to support its expression compared to other Df target genes; however, we found that when Df levels peak in nc14d, in which the embryos are about to initiate gastrulation, *sna* transcriptional activity (i.e rate of initiation) decreases, suggesting an alternative function of Df (e.g. “damper-like” function) at late nc14.

To test the idea further, an optogenetic approach was used that supports degradation of a protein of interest with fine-scale time resolution; in this case, degrading Df only at late nc14. In a previous study, we showed using an optogenetic approach that Df becomes dispensable for *sna* gene expression in late nc14, as Twist (Twi) can support *sna* activation after Df input is lost (Irizarry et al., 2020). Using this light-inducible degradation approach in the Blue Light-Inducible Degron (BLID) was fused to Df at the C-terminus (Df-BLID) at the endogenous locus (Irizarry et al., 2020), we examined a temporal requirement for Df in modulation of *sna* levels. In the dark, the degron domain is inaccessible and Df-BLID protein is present and functional; whereas, upon exposure to blue light, the degron is exposed, resulting in Df-BLID protein degradation (Fig. 3C)(Bonger et al., 2014). To monitor *sna* transcriptional activity while controlling Df levels, we combined blue light inducible degradation system with MS2-MCP.GFP system. Df-BLID is quite

sensitive to blue light exposure, and its light induced degradation happens in less than 5 min (Irizarry et al., 2020). To monitor *sna* gene expression, while also degrading sufficient amounts of DI-BLID within a narrow time window, a high laser power at 15% was used to initially degrade DI-BLID followed by imaging with a low laser power at 5% to image *sna.MS2* associated MCP.GFP signal (Fig. 3B). Furthermore, to investigate the individual role of each enhancer in the control of *sna* expression, *sna.MS2* reporter constructs containing deletions of each enhancer was performed (Fig. 3A)(Bothma et al., 2015). After 5 min of confocal blue laser illumination to degrade DI-BLID, *snaMS2.wt* and *snaMS2.Δprox* reporters remain active (Fig. 3D,E). In contrast, under the same conditions, *snaMS2.Δdis* reporter expression is lost (Fig. 3D). These results suggest that the proximal enhancer requires high levels of DI to maintain activity at late nc14, whereas the distal enhancer becomes independent of high DI levels.

To directly examine the anti-correlation observed between DI and *sna* levels, levels of DI were modulated in embryos and effects on *sna* transcription using the *snaMS.wt* reporter as a proxy were compared at nc14d. To see the trend in *sna* levels throughout nc14d, the mean spot sizes throughout nc14d were normalized to the initial level at the onset of nc14d. The embryos with high DI levels exhibit the fastest decrease in *sna* levels (*wt*, Fig. 4F); whereas embryos with intermediate DI levels exhibit a slower decrease in *sna* levels (*dl<sup>1</sup>/+* and *dl<sup>4</sup>/+*, Fig. 3F) in comparison. Furthermore, embryos with low to no DI show a slight increase in *sna* levels (*dl-BLID* background, after light exposure, Fig. 3F).

To further investigate the idea that DI levels are anti-correlated to *sna* levels, *sna* levels were examined by MS2-MCP.RFP system with embryos with optogenetically induced different levels of DI degradations at late nc14 . In contrast to MS2-MCP.GFP system in which DI degradations cannot be avoided while detecting MCP.GFP signals, detecting MCP.RFP signals associated with MS2 stem loops does not induce DI degradation. Starting with monitoring *sna*

transcriptional activity by detecting MCP.RFP signals associated with *sna*-MS2 stem loops for initial 4 min, subsequently Df proteins were degraded by illuminating embryos with confocal blue laser while detecting MCP.RFP signals. After blue laser illumination, MCP.RFP signals were continuously monitored at nc14d (Fig. 4A).

First, we tried to degrade Df in various degrees by illumination embryos with confocal blue laser with various time lengths. To confirm that different levels of Df degradation can be achieved, flies with mCherry-BLID domain fused to c-terminus of *dl* (i.e. *dl-mCh-BLID*) at endogenous locus were utilized. This mutant line allows us to detect Df levels by monitoring mCherry fluorescent signals while degrading Df by blue laser illumination. To generate intermediate or low Df levels, *dl-mCh-BLID* embryos were illuminated for 3 min and 5 min, respectively. As expected, *dl-mCh-BLID* embryo with 3 min blue laser illumination exhibited intermediate Df levels, whereas *dl-mCh-BLID* embryo with 5 min blue laser illumination exhibited low levels of Df (Fig. 4B).

Additionally, to confirm that the scheme used for blue laser illumination (i.e. 5 min of blue laser illumination with 20% laser power) does not affect on MCP.RFP signals, wildtype embryos were illuminated with blue lasers while detecting MCP.RFP signals associated with *sna* MS2 (Fig. 4A). No apparent differences were observed between embryos with/without blue laser illumination (Fig. 4C), suggesting that blue laser illumination used does not interfere with detecting *sna* MS2-MCP.RFP signals. Next, *dl-BLID* embryos were illuminated for 3 min or 5 min to degrade Df to intermediate or low levels, respectively, while detecting MCP-RFP signals associated with *sna* MS2. Surprisingly, while *dl-BLID* embryos were illuminated, mean spot sizes incrementally decrease (Fig. 4D), suggesting *sna* is losing its activity. Once the illumination ceases, the mean spot size increases and reaches above the initial level, in which the mean spot size decreases in the embryos without illumination (Fig. 4D). Furthermore, embryos with 5 min illumination (i.e. low

DI levels) reach higher levels than the embryos with 3 min illumination (i.e. intermediate DI levels). It suggests that, at late nc14, DI inputs negatively control *sna* levels.

Taken together, DI levels negatively affect *sna* levels. Furthermore, slowing down *sna* activity levels seen at late nc14 is dependent on proximal enhancer activity (Fig. 4E). Together with the previous observation that high levels of DI are required to maintain the proximal enhancer activity, these results support the view that in late nc14 DI functions as a damper to decrease *sna* transcription by acting through the *sna.prox* enhancer.

In summary, morphogens provide spatial cues for tissue formation, and the importance of achieving adequate levels of morphogens to activate concentration-dependent activation of target genes has been a focus of study; our data suggest that morphogens can also negatively influence levels of target gene expression. We propose that morphogens may commonly provide feedback to target genes that helps keep target gene expression robust. Animals have developed various defense mechanisms to overcome the challenges posed by fluctuating morphogen gradients that can result from genetic and environmental perturbations. As a defense mechanism, the *sna* gene has adapted to have two regulatory modules that differentially respond to morphogen DI concentration. In response, DI acquire temporally distinct functions, to precisely control gene expression levels. Furthermore, this regulatory mechanism, in which multiple enhancers are active and only a subset may require continuous morphogen input, may explain how morphogens accomplish hysteresis, the development phenomenon in which some target genes require morphogen exposures at higher levels, early but only require lower levels, later. It is possible that morphogen responses that perdure do not relate to prolonged function in supporting gene activation per se may be involved instead in the regulation of gene level control, possibly through multi-enhancer coordination.

## MATERIALS & METHODS

### *Fly stocks/ husbandry and crosses*

All flies were kept in 25°C, except for flies bearing *dl-BLID* allele which were kept in 18°C. *yw* is used for wildtype. To test *sna* transcriptional activity in *dl<sup>1</sup>* or *dl<sup>4</sup>* heterozygous background, *dl<sup>1</sup>/CyO* or *dl<sup>4</sup>/CyO* flies were crossed with *MCP-GFP*, *mCherry-PCP*, *His2Av-eBFP2* (from Michael Levine, Princeton University, US; Lim et al., 2018) to generate *dl<sup>1</sup>/CyO* (or *dl<sup>4</sup>/CyO*); *MCP-GFP*, *mCherry-PCP*, *His2Av-eBFP2*. The virgin flies with *MCP-GFP*, *mCherry-PCP*, *His2Av-eBFP2* in wildtype, *dl<sup>1</sup>* or *dl<sup>4</sup>* heterozygote or *dl-BLID* background (*dl-BLID*; *MCP-GFP*, *mCherry-PCP*, *His2Av-eBFP2*) were crossed with *sna* MS2 males for live imaging.

### *CRISPR-Cas9 Mediated Genome Modification*

To target a deletion of the various sizes of proximal distal enhancer [i.e. *sna.Δprox4.4*, *sna.Δprox3.0*, *sna.Δprox3.0(mDl)*] or distal enhancer (i.e. *sna.Δdis2.0* or *sna.Δdis1.8*), a transgenic line was generated expressing two guide RNAs (gRNAs) targeting the flanking region described as proximal enhancer or distal enhancer (Dunipace et al., 2011). First, the unique PAM recognition sites were identified flanking this region using the flyCRISPR optimal target finder (<https://flycrispr.org/target-finder>). These two sites were cloned into the plasmid pCFD4-U6:1\_U6:3tandemgRNAs (Addgene plasmid#49411). The plasmid was injected into *y2cho2v1; P{nos-phiC31\int.NLS}6X; attP2 (III)* (NIG-Fly #TBX-0003), resulting in phiC31-mediated site-integrated at landing site attP2 (Chr. III) (Kondo and Ueda, 2013). Integration in the genome at this position was confirmed by PCR/sequencing.

To generate the proximal deletion lines [*sna.Δprox4.4*, *sna.Δprox3.0*, *sna.Δprox3.0(mDl)*] and the full distal deletion (*sna.Δdis2.0*), Homology-Directed Repair (HDR)/CRISPR-Cas9

system was used. A donor construct was generated using pHD-DsRed vector (Addgene plasmid #51434). An ~1kb 5' or 3' homology arm to the regions either upstream or downstream of was cloned with SmaI/NheI or AscI/XhoI, respectively.

*y2cho2v1;sp/CyO;P {nos-Cas9,y+,v +} 2A* (NIG-Fly #Cas-0004) virgin flies were crossed with gRNA transgenic male flies. Embryos were collected and injected with 300 ng/μl of the donor vector. The deletion lines were screened by DsRed expression in adult fly eyes and sequencing. The DsRed marker was removed by crossing with a Cre expressing fly line (*y[1] w[67c23] P{y[+mDint2]=Crey}1b; D[\*]/TM3, Sb[1], BDSC #851*). The sequences after Cre-mediated DsRed marker excision were provided below; uppercase indicates genomic sequence, and lowercase indicates loxP remnant sequence.

>*sna.Δprox4.4*

CGACAAAGGATGTGACTCAGcggccgcgacatatgcacacctgcatcgtagtgcaccaactggggtaacctttgag  
ttctctcagttggggcgtagataacttcgtataatgtatgctatacgaagttatagaagagcactagtAGGGTGCGCCTGCGTC  
TGCT

> *sna.Δprox3.0*

CGACAAAGGATGTGACTCAGcggccgcgacatatgcacacctgcatcgtagtgcaccaactggggtaacctttgag  
ttctctcagttggggcgtagataacttcgtataatgtatgctatacgaagttatagaagagcactagtATGACCCACCAGGTAG  
GATG

> *sna.Δdis2.0*

TAAATTCCAACATTTTGCTGcatatgcacacctgcatcataacttcgtataatgtatgctatacgaagttatagaagagca  
ctagTAAATCCGTGTTAAATTGTT

To introduce mutated predicted Df binding sites in *sna.Δprox3.0* (see Fig. S2), first all three Df consensus sequence were replaced with the complementary base (i.e. A>T or G>C), and 1kb of the proximal enhancer sequence including all the mutated Df consensus sequence are synthesized and inserted into pUC57 (GenScript). This plasmid was used as a DNA template to generate the homology right arm for HDR donor construct.

> mDf in *sna.Δprox3.0* (uppercase: genomic sequence, lowercase: mutated Df consensus sequences)

```
ATGACCCACCAGGTAGGATGTGAGGACATAaacttttggggAGCCAGTTTTTCCACTCGTC
GTGGCTTGTTTTGCTTGAGTTTCGCTGACTGCGTAATTGGATAAGATGGGAAATTAC
TTTAAATCCTTCGCTGATCCACATCCGGACATTCGTCGAAGGAAAATCCATTGCAGG
GAAATACGAAATGGAAATGCGGCTGGGTTATTGGCTCGACATTTCCCATCTTCCCTC
ACGCCATTGGTTGCAGGATCGCGGGGAAaaccttaaggcgGCTGGAATTTTTTGTACCTCT
TGGGTTTATCAAACCTTTTGGGTTTGCTATcctaaaaagggtATTTTACC
```

To generate *sna.Δdis1.8*, non-homologous end joining (NHEJ)/CRISPR-Cas9 system was used. *y2cho2v1;sp/CyO;P {nos-Cas9,y+,v +} 2A* (NIG-Fly #Cas-0004) virgin flies were crossed with gRNA transgenic male flies. The individual progeny was confirmed for the deletion by PCR.

> *sna.Δdis1.8* (uppercase: genomic sequence, lowercase: sequence added after NHEJ)

```
TGAACTTGTTGTGAACTCAGtaccGTCTAATGGCCAGAACACCG
```

#### *In situ hybridization and imaging*

Embryos were collected for 2hr and age at 25°C to have 2-4 hr old embryos, followed by fixation and *in situ* hybridization using standard protocol. Antisense *sna* RNA probes were labeled

with digoxigenin. For fluorescent in situ hybridization, sheep anti-digoxigenin (Life Technology PA185378) was used with Alexa conjugate secondaries (1:400; Thermo Fisher). For alkaline phosphatase (AP) staining samples, Zeiss Axioxa 506 microscope was used, and images were taken with 20x objective. For FISH samples, Zeiss LSM800 microscope with a 20X objective were used.

#### *RNA extraction from single embryos and qPCR*

Embryos were collected for 1hr and aged for 2hr at 25°C. They were hand dechorionated and immersed in halocarbon 27 oil to determine the stage of the embryos. When the embryos reached the nc14c, in which 75% of cellularization processed, the individual embryo was collected into a 1.5ml microcentrifuge tube and proceeded for RNA extraction using standard Trizol protocol (ThermoFisher), followed by DNase I treatment (NEB) and cDNA synthesis using SuperScript III First-Strand Synthesis System kit (ThermoFisher). qPCR was performed on cDNA using SYBR Green I Master Mix (Roche) on a StepOnePlus Real-Time PCR System (Applied Biosciences).

#### *Confocal blue laser illumination*

*dl-BLID* embryos were collected for 1 hr and aged for 4hr at 18°C. Embryos were hand dechorionated, and placed onto a heptane glued slide. The embryos were immersed in water and the stages of embryos were determined using Zeiss LSM 800 confocal microscope using 25x water immersion objective lens. At 40 min after onset of nc14, in which embryos reach nc14 mid c, a 488nm confocal blue laser with 15% laser power was applied using a 25x water immersion objective (see Figure 4C).

### *Live imaging*

To test *sna* transcriptional activity changes responded to DI levels, the MS2-MCP system was used. In wildtype (high level of DI), *dl* heterozygote ( $dl^1/+$  or  $dl^4/+$ , intermediate level of DI) embryos, nuclei were marked by imaging H2A-BFP (i.e. His2Av.eBFP2) with 405nm with 1.5% laser power, and MCP-GFP associated with *sna* MS2 were detected with 488nm with 2.4% laser power. For dl-BLID embryos, nuclei were marked by imaging H2A-BFP using 405nm with 1.5% of laser power, and MCP-GFP associated with *sna* MS2 were detected using 488nm with 5% laser power before and after blue laser illumination (see Fig. 3B). All images were taken by 30 Z-planes separated by 3.0 $\mu$ m.

### *Mean Intensity Quantification*

To measure the mean intensity of FISH samples using Fiji software, max projection of scans was used. A square of 345  $\mu$ m<sup>2</sup> regions were selected within an embryo and analyzed for the mean intensity. This process was repeated three times in each embryo within *sna* expression domain.

### *MS2 Foci detection and Quantification*

MS2 foci were detected and quantified similar to previously described methods (Irizarry et al., 2020). Briefly, a gaussian filter was used to smooth the image and background was removed using Top-hat filtering. A threshold was applied to detect MS2 foci. Foci consisting of only two pixels were removed as often these represented individual bright pixels not associated with real signal, as confirmed manually. For each focus detected, the number of pixels were counted, and the mean intensity of each focus was determined. The number of foci, the mean area (or mean number of pixels per focus), and the average of the mean intensities was computed at every time

point for each individual embryo. The number of foci detected and mean area or mean number of pixels per focus were plotted. Little difference was detected in the average of mean intensities (not shown). Error bars were calculated as the standard error of the mean.

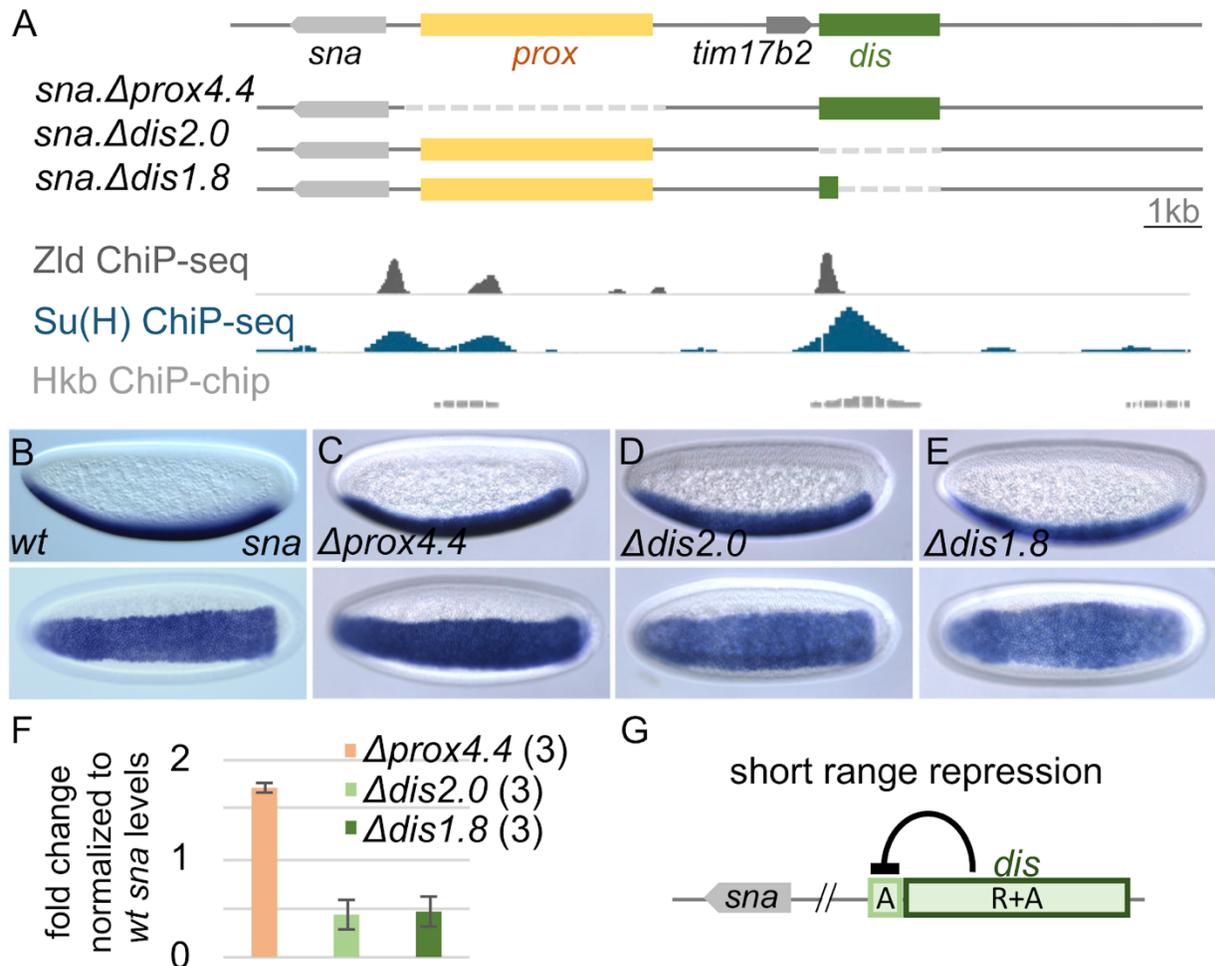
## REFERENCES

- Bonger, K. M., Rakhit, R., Payumo, A. Y., Chen, J. K. and Wandless, T. J.** (2014). General method for regulating protein stability with light. *ACS Chem. Biol.* **9**, 111–115.
- Bothma, J. P., Garcia, H. G., Ng, S., Perry, M. W., Gregor, T. and Levine, M.** (2015). Enhancer additivity and non-additivity are determined by enhancer strength in the *Drosophila* embryo. *Elife* **4**,.
- Dunipace, L., Ozdemir, A. and Stathopoulos, A.** (2011). Complex interactions between cis-regulatory modules in native conformation are critical for *Drosophila* snail expression. *Development* **138**, 4075–4084.
- Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R. and Eisen, M. B.** (2011). Zelda binding in the early *Drosophila melanogaster* embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet.* **7**, e1002266.
- Ip, Y. T., Park, R. E., Kosman, D., Yazdanbakhsh, K. and Levine, M.** (1992). dorsal-twist interactions establish snail expression in the presumptive mesoderm of the *Drosophila* embryo. *Genes & Development* **6**, 1518–1530.
- Irizarry, J., McGehee, J., Kim, G., Stein, D. and Stathopoulos, A.** (2020). Twist-dependent ratchet functioning downstream from Dorsal revealed using a light-inducible degron. *Genes Dev.* **34**, 965–972.
- Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J. A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S. R., Tan, G., et al.** (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D1284.
- Kondo, S. and Ueda, R.** (2013). Highly improved gene targeting by germline-specific Cas9 expression in *Drosophila*. *Genetics* **195**, 715–721.
- Koromila, T. and Stathopoulos, A.** (2019). Distinct Roles of Broadly Expressed Repressors Support Dynamic Enhancer Action and Change in Time. *Cell Rep.* **28**, 855–863.e5.
- Lim, B., Heist, T., Levine, M. and Fukaya, T.** (2018). Visualization of Transvection in Living *Drosophila* Embryos. *Mol. Cell* **70**, 287–296.e6.
- MacArthur, S., Li, X.-Y., Li, J., Brown, J. B., Chu, H. C., Zeng, L., Grondona, B. P., Hechmer, A., Simirenko, L., Keränen, S. V. E., et al.** (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 Biology* **10**, R80.
- Ozdemir, A., Ma, L., White, K. P. and Stathopoulos, A.** (2014). Su(H)-mediated repression positions gene boundaries along the dorsal-ventral axis of *Drosophila* embryos. *Dev. Cell* **31**, 100–113.

**Reeves, G. T., Trisnadi, N., Truong, T. V., Nahmad, M., Katz, S. and Stathopoulos, A.** (2012). Dorsal-ventral gene expression in the *Drosophila* embryo reflects the dynamics and precision of the dorsal nuclear gradient. *Dev. Cell* **22**, 544–557.

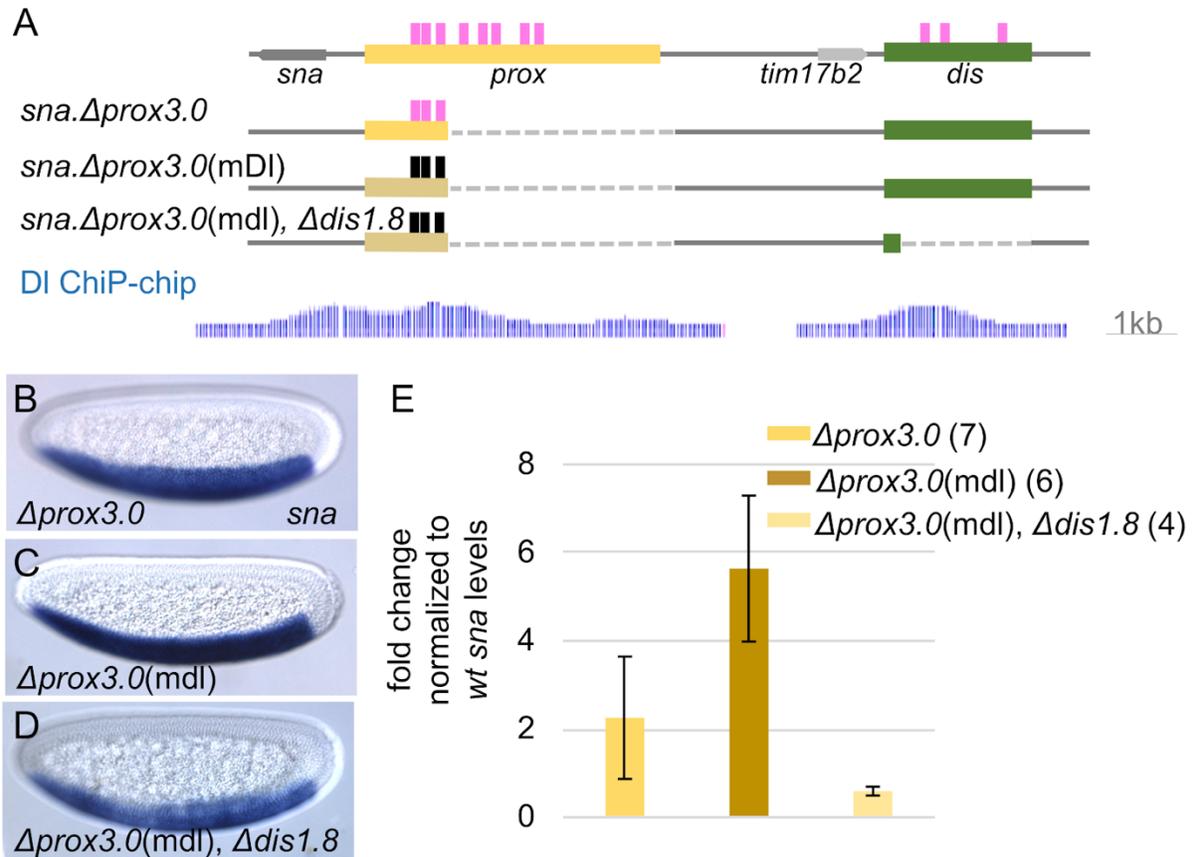
**Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al.** (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676–682.

## FIGURES &amp; LEGENDS

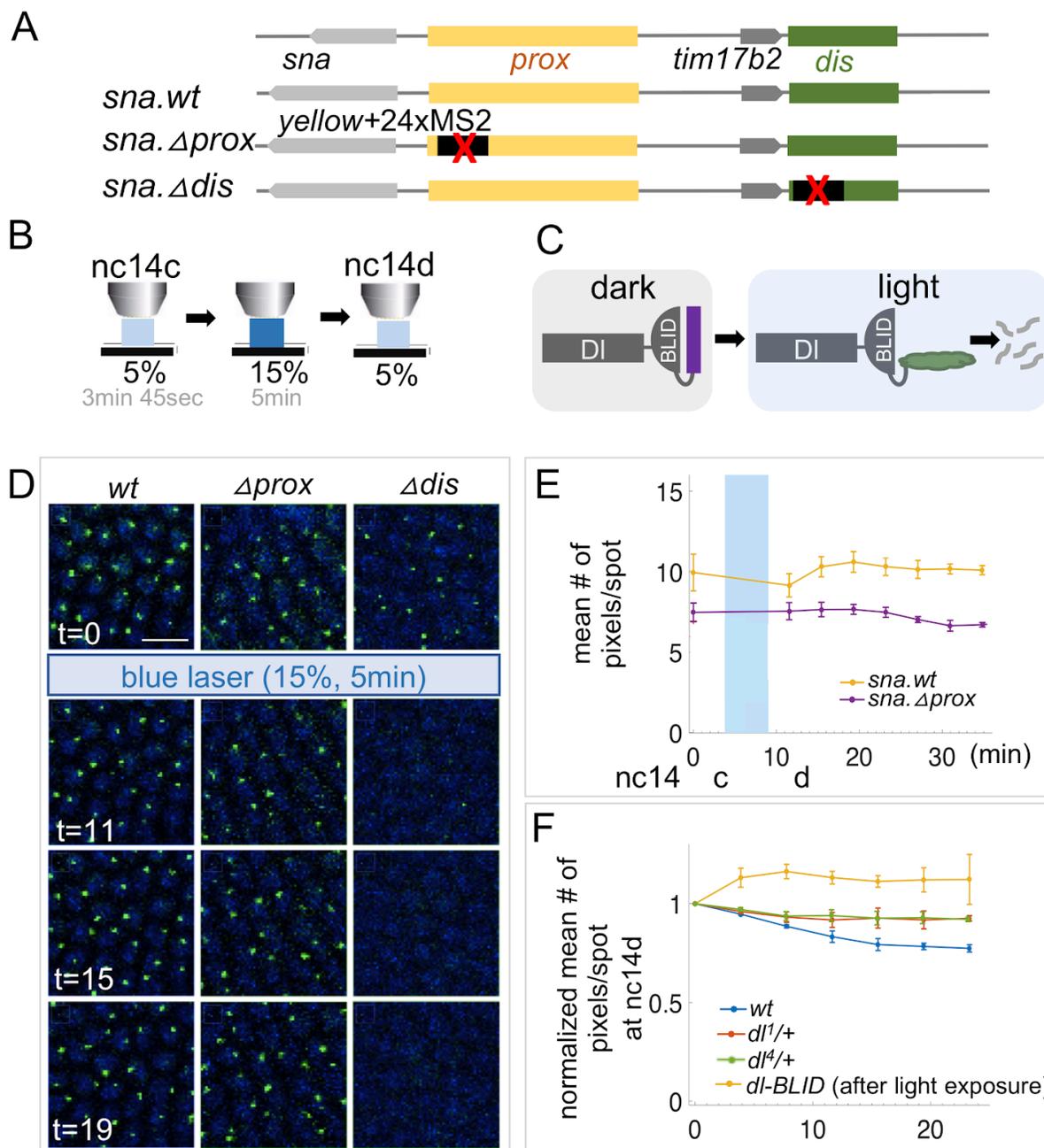


**Figure 1. *sna* boundaries are supported by the distal enhancer, which supports higher levels of expression, compared to the proximal enhancer.** (A) Schematic of enhancer deletion, proximal (*prox*, yellow) or distal (*dis*, green), by CRISPR-Cas9 system. Binding peaks by Zelda (Zld, dark gray)(Harrison et al., 2011), Suppressor of Hairless [Su(H), dark blue] (Ozdemir et al., 2014), and Hucklebein (Hkb, light gray) (MacArthur et al., 2009) are shown. (B-E) *sna* expression by *in situ* hybridization at nc14d. Lateral view of embryos (top row), Ventral view of embryos (bottom row). All embryos are oriented with anterior to left. (F) Normalized *sna* transcript levels

to wildtype at nc14c by qPCR. Sample sizes are indicated next to the genotype. (I) Model of short range repression in the distal enhancer to set *sna* expression boundaries. A: activation domain. R+A: repression/activation domain.

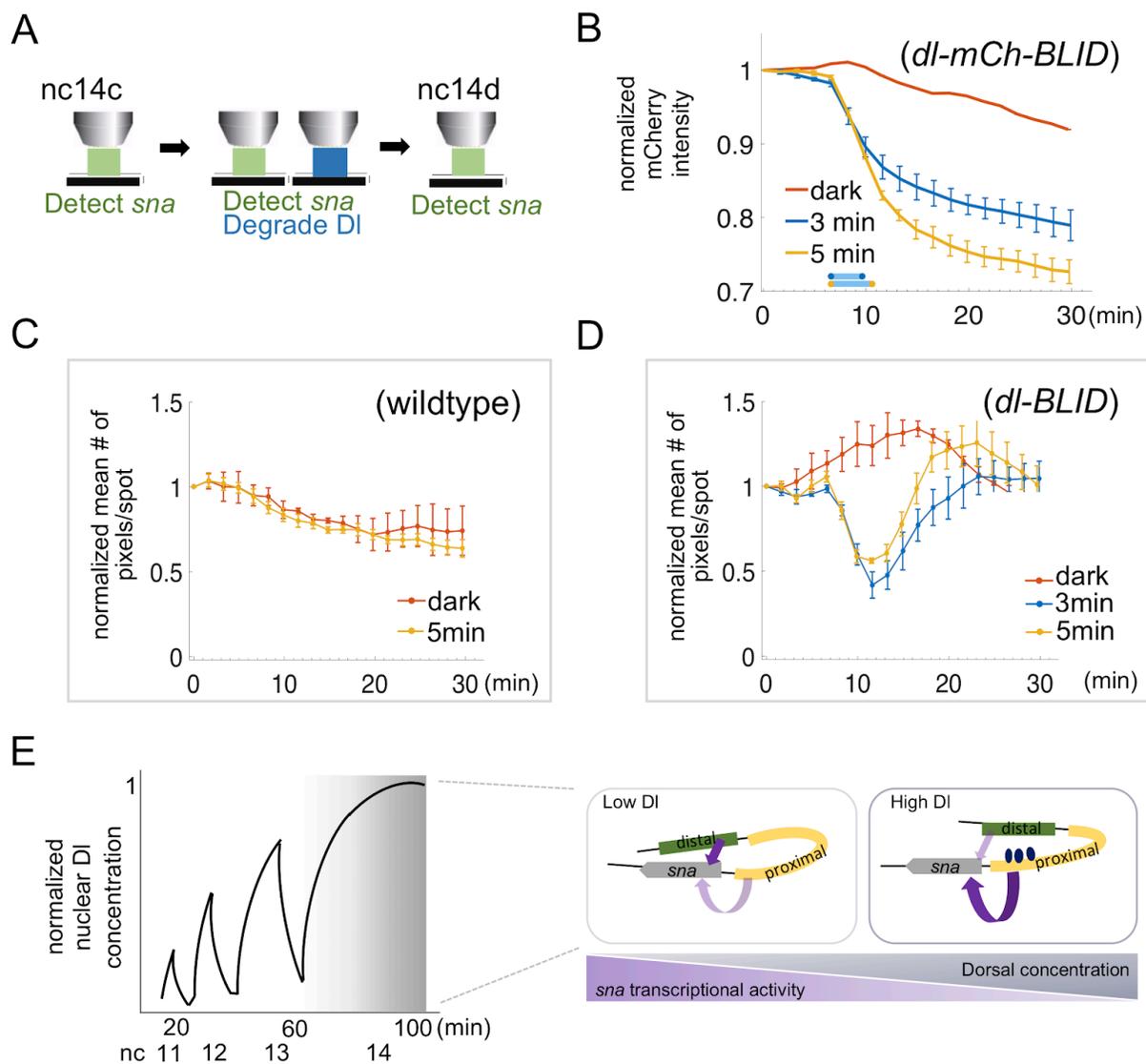


**Figure 2. DI bindings on the proximal enhancer limit the distal enhancer activity.** (A) Schematic of manipulations in the proximal or distal enhancer, 3.0kB proximal deletion r (*sna.Δprox3.0*), 3.0kB proximal enhancer deletion with mutated putative DI binding sites [*sna.Δprox3.0(mDI)*], 1.8kB of distal enhancer deletion in *Δprox 3.0(mDI)* [*sna.Δprox 3.0(mDI)*]. Binding peaks by DI (blue) (MacArthur et al., 2009)(B-D) *sna* expression at nc14d by *in situ* hybridization. All embryos are oriented with anterior to left and dorsal to top. (E) Normalized *sna* transcript levels to wildtype at nc14c by qPCR.



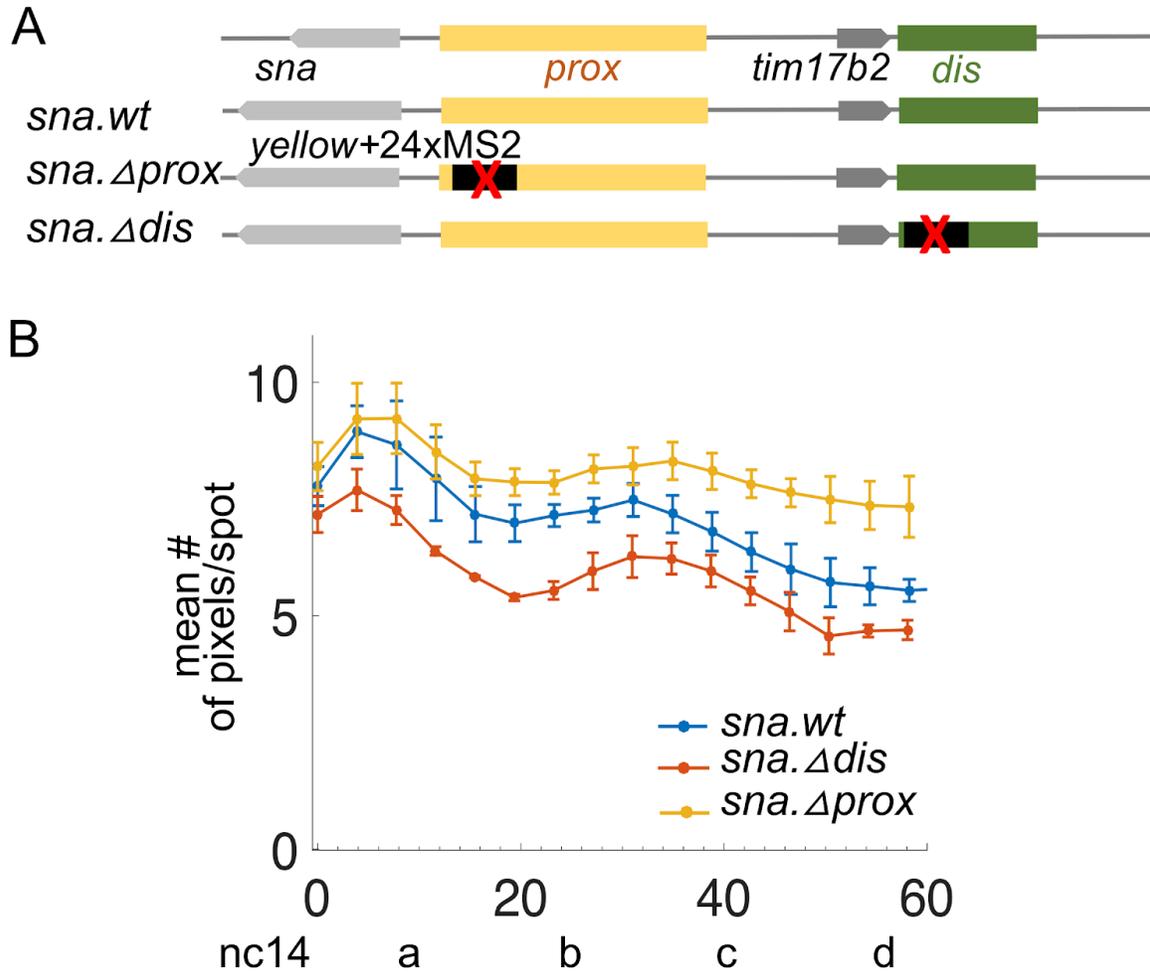
**Figure 3. High levels of DI function as a damper for *sna* levels, that supports the proximal enhancer activation limiting the distal enhancer activity.** (A) Schematic of reporter constructs used to monitor MS2-MCP.GFP to assay *sna* transcriptional activities (Bothma et al., 2015). (B) Scheme of *sna* MS2-MCP.GFP system combined with DI-BLID degradation system at late nc14

using confocal blue laser. (C) Schematic of D1-blue light inducible degron (dl-BLID) system (Bonger et al., 2014). (D) MCP.GFP signals associated with *sna* MS2 (green dot) driven by both enhancers (*snaMS2.wt*, left column) or either enhancer [*snaMS2.Aprox* (middle column) or *snaMS2.Δdis* (right column)] are shown before (t=0 min) and after (t=11,15,19 min) D1-BLID degradation by 15% confocal blue laser for 5 min. Nuclei are visualized by H2A-BFP (blue). Scale bar: 10μm. (E) Quantitative analysis of mean number of pixels per MCP.GFP spot associated with the *sna* MS2 reporter in dl-BLID embryos with *snaMS2.wt* or *snaMS2.Aprox* regulatory conditions. Error bars represent standard error of the mean. Blue shade represents the 5 min of 15% blue laser illumination. (F) Quantitative analysis of mean number of pixels per MCP.GFP spot associated with *snaMS2.wt* in wildtype (blue, high D1 levels), *dl<sup>1/+</sup>* (red, intermediate D1 levels), *dl<sup>4/+</sup>* (green, intermediate D1 levels) and *dl-BLID* after blue laser exposure (yellow, low to zero D1 levels). Each value was normalized to the initial value at the onset of nc14d. Error bars represent standard error of the mean. Three movies were taken for each condition.



**Figure 4. DI levels anti-correlate to *sna* levels.** (A) Schematic of *sna* MS2-MCP.RFP system combined with DI-BLID degradation system at late nc14. (B) Quantitative analysis of levels of mCherry fused to DI-BLID proteins (*dl-mCh-BLID*). *dl-mCh-BLID* embryos were kept in dark or illuminated by confocal blue laser for 3 or 5 min. (C,D) Quantitative analysis of mean number of pixels per MCP.GFP spot associated with *sna*MS2.wt in wildtype (C) or *dl-BLID* (D) embryos. Embryos were kept in dark or illuminated for 3 or 5 min. Error bars represent standard error of the

mean. Three movies were taken for each genetic background. (E) A representation of Df nuclear levels at ventral-most region of embryos from nc 11 to 14 (Reeves et al., 2012) (left). Model showing anti-correlation between Df levels and *sna* transcriptional activity at nc14, in which Df bindings on the proximal enhancer dampen the distal activities to fine tune *sna* levels (right).

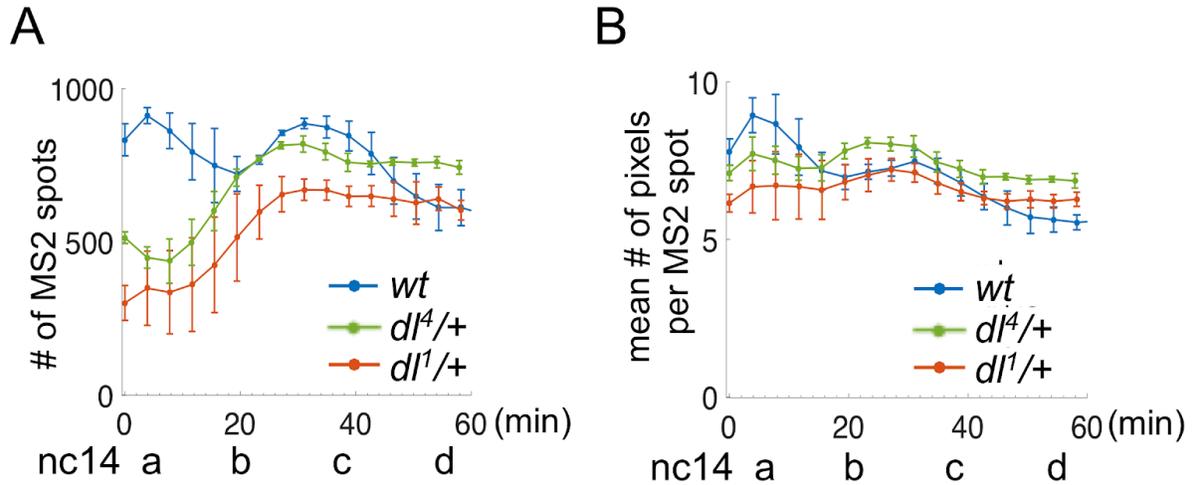


**Figure S1. Inputs from the proximal and distal enhancers are required to support normal *sna* expression levels.** (A) Schematic of reporter constructs used to monitor MS2-MCP.GFP to assay *sna* transcriptional activities (Bothma et al., 2015). (B) Quantitative analysis of mean number of pixels per MCP.GFP spot associated with *sna* MS2 reporter in wildtype embryos with *snaMS2.wt*, *snaMS2.Δprox*, or *snaMS2.Δdis* regulatory conditions at nc 14. Error bars represent standard error of the mean. Three movies were taken for each condition.

A

aaccgagatatccgggatgaccaccaggtaggatgtgaggacata**ATCTTTTGGGGG**  
**TAGAAAACCCCC**agccagttttccact  
 cgtcgtggcttgtttgcttgagtttcgctgactgcgtaattggataagatgggaaattacttaaatccttcgctgatccacatcc  
 ggacattcgtcgaaggaaaaatccattgcagggaaatacgaaatggaaatgcggtgggttattggctcgacattcccatc  
 ttccctcagccattggttgcaggatcgcggggaa**AACCTTAAGGCG**  
**TTGGAATTCCGC**gctggaatTTTTgtcacctctgggttatca  
**CCTAAAAAAGGT**  
 aaactttgggttgctat**GGATTTTTTCCA**attttacc

**Figure S2. Mutating predicted DI binding sites in the proximal enhancer, generating *sna.Aprox3.0(mDI)* line.** (A) Predicted DI binding sites by JASPR were indicated in bold font. The putative DI binding sites were mutated by switching the sequences to complementary sequences (red). Sequences with blue underline show identified DI binding sites by DNase I footprint assay (Ip et al., 1992)



**Figure S3. Comparison of MCP.GFP signals associated with *sna* MS2 reporter between wildtype and *dl<sup>1</sup>/+* or *dl<sup>4</sup>/+* heterozygote at nc 14.** (A) Quantitative analysis of the number of MCP.GFP dots associated with the *sna* MS2 reporter in wildtype and *dl<sup>1</sup>/+* or *dl<sup>4</sup>/+* heterozygote at nc 14. (B) Quantitative analysis of mean number of pixels per MCP.GFP spot associated with *sna* MS2 reporter in wildtype and *dl<sup>1</sup>/+* or *dl<sup>4</sup>/+* heterozygote at nc 14.

## **Chapter 4**

### Discussion

The goal of this study was to provide better understanding in temporal roles of morphogens in early embryonic development. In order to provide mechanical insights, we used DV patterning gene *dl* in *Drosophila* embryos as a system. Dl is dynamic, and the levels continuously increase. To manipulate Dl levels with fine time resolutions, we generated a blue light inducible Dl degradation (Dl-BLID) system. Furthermore, for deeper understanding in how dynamically changing Dl levels in time are interpreted by target genes at the *cis*-regulatory level, we generated various mutant lines with various enhancer conditions at the *sna* endogenous locus by CRISPR-Cas9 system.

Dl is required to provide positional cues for DV patterning genes at early nc14. Upon the blue light illuminations at early nc14, *sna* loses its expressions. Low to zero Dl levels at the early stage inactivate the *sna* proximal and distal enhancers. Interestingly, *sog*, a lateral target gene, was activated even with the blue light illumination at the early stages. It suggests two scenarios: (1) *sog* gene senses extremely low levels of Dl, so inability to completely degrade Dl proteins allows *sog* gene to get activated. (2) *sog* activity, perhaps, become independent of Dl at nc14. Possibly, *sog* activation before nc14 may advance to the next regulatory networks, and it allows *sog* to be continuously expressed with absence of normal Dl levels. However, to elucidate molecular mechanism of temporal roles of Dl in *sog*, further investigation is required.

Unlike early nc14, *sna* activation becomes independent of high Dl levels at late nc14. Surprisingly, when Dl levels decrease upon blue light illumination at late nc14, *sna* activation remains. This is an unexpected result since *sna* is known to be a gene that requires high Dl levels. After analyzing *sna* activation at the *cis*-regulatory level, we found that the *sna* distal enhancer becomes independent of high Dl levels, whereas the *sna* proximal enhancer still requires high Dl levels. It suggests that the distal enhancer alone can support *sna* expression at a later time point. Furthermore, we found that Twi is a transcriptional factor supporting the *sna* expression at the late

nc14. Once Twi advances to an auto-regulatory state, its activity remains active regardless of Dl levels. Furthermore, high Twi levels supported by auto-regulation, functioning as a molecular ratchet, are sufficient to support various target genes in the ventral region.

A functional shift of Dl when its levels are at the highest ensures precision in *sna* levels at late nc14. Up to date, we did not understand why Dl levels continuously increase even after Dl becomes dispensable for multiple presumptive mesodermal gene activations, including *sna*. At this stage, in wildtype embryos, we found that Dl levels and *sna* activity levels become anticorrelated. As Dl levels increase, the *sna* levels decrease. However, when the high Dl levels are manipulated to lower levels at late nc14, *sna* transcriptional activity fails to slow down and increase net *sna* levels. Furthermore, with various mutagenesis in the *sna* proximal enhancers, we identified the 345bp region in the proximal enhancer that attenuates the distal enhancer activity to reduce *sna* levels. We proposed that Dl binding on the 345bp region on the proximal enhancer interferes with the distal enhancer activity, and it, in turn, dampens net *sna* levels at late nc14.

*Maintenance of gene expressions under genetic or environmental perturbations in morphogen levels*

*sna* proximal and distal enhancers are sensitive to morphogen Dl levels at early nc14. However, the distal enhancer activity becomes independent of Dl while the proximal enhancer remains sensitive to Dl levels at late nc14. The distal enhancer is predominantly controlled by Twi, allowing *sna* to remain active without high Dl levels at late nc14. These morphogen dependency changes in the distal enhancer may allow *sna* gene to be robustly expressed under genetic or environmental perturbations in Dl concentrations. This defense mechanism is well conserved among invertebrate and vertebrate systems. At *Xenopus* blastula stage, morphogen activin supports ratchet-like mechanism. Once cells are exposed to high activin levels, the cells maintain adapted

cell fates even when activin levels decrease (Gurdon et al. 1995). However, how activin promotes the process at *cis*-regulatory levels is not well understood. It is interesting to know whether a ratchet-like phenomenon is controlled by a single enhancer in vertebrate systems, which allows a gene to remain active under any types of morphogen level perturbations.

*Perduring morphogens may be required to support robust gene expression levels*

Morphogens provide positional cues to set body patterning at early embryonic development. After the initial inputs from morphogens, multilayered controls (i.e. feed-forward loop, cross repression, positive feedback, auto-regulation) become integrated to ensure precious gene expressions (Ashe and Briscoe 2006). However, some morphogen expressions persist even after its inputs become dispensable to maintain the gene expressions (i.e. perduring expression of Dl at late nc14). We demonstrated that a persisting Dl has an additional function in fine-tuning of gene expression levels. Thus, it is possible that morphogen responses that perdure do not relate to prolonged function in supporting gene activation. It may be involved in gene level controls, possibly through multi-enhancer coordination in other systems.

*Mechanisms during level fine-tuning*

Gene expression levels are controlled by transcriptional bursting. Combinatorial inputs of transcriptional factors on enhancers control frequencies of interactions between the enhancers and target promoter, dictating transcriptional bursting rates. Recent studies have proposed several molecular mechanisms controlling bursting rates. Strong enhancers have high bursting rates by high frequencies of interaction between enhancer and promoter, whereas weak enhancers have lower rates due to low frequencies of interaction between enhancer and promoter. Furthermore, distance between an enhancer and promoter also can control bursting rates, such that an enhancer

that is located proximal to the promoter has higher rates (Fukaya et al. 2016). However, an enhancer located distal to promoter can increase the interaction frequencies by looping (Lewis 1978; Lettice 2003; Fukaya et al. 2016). However, this looping between distal enhancer and promoter can be interfered by the insulator (Fukaya et al. 2016). Thus, mechanisms for transcriptional level control can be quite complex.

At late nc14, Df acting on the *sna* proximal enhancer dampens *sna* transcriptional activity levels by attenuating the *sna* distal enhancer activity. How does the proximal enhancer interfere the distal enhancer activity? Few scenarios are possible. Perhaps, to support *sna* expression, the enhancers compete with each other. At early nc14 in which both enhancers are sensitive to Df, possibly, the proximal enhancer activity dominantly contributes to net *sna* levels due to the proximity to the promoter. However, as Twi levels build up, Twi potentiate the distal enhancer (i.e. mediating looping processes of the distal enhancer to the promoter), and inputs from the distal enhancer become the major contributors controlling *sna* levels. Since the distal enhancer drives high *sna* levels, net *sna* levels increase fast. Finally, when Df levels reach the maximum levels at late nc14, high levels of Df convert the proximal enhancer to a hyper-active enhancer, and the *sna* levels are mainly mediated by the proximal enhancer. Since the proximal enhancer drives low levels of *sna*, it results in decreasing transcriptional activities at the late stages.

Alternatively, the proximal enhancer may directly interfere with the distal enhancer activity. Df can physically interact with Twi, forming heterodimers (Shirokawa and Courey 1997). When the proximal enhancer is occupied with Df proteins, Twi bound onto the distal enhancer may favorably interact with the Df cluster on the proximal enhancer. This “Molecular Velcro” mechanism interferes with interactions between the distal enhancer and the promoter by trapping the distal enhancer at the proximal enhancers by heterodimerization of Df and Twi. Cross-repression between enhancers have been proposed during patterning. However, how one enhancer

directly represses another enhancer is not well understood. Thus, Molecular Velcro can be one of mechanisms utilized during cross-repression processes.

## Future Directions

### *Early morphogen function as a chromatin modifier*

In the *Drosophila* embryo, body axis patterning morphogens are detected at quite early embryonic stages. Bcd is detected at nc 6, and Dl is detected at nc 10 (Little et al. 2011; Reeves et al. 2012). In this thesis, we demonstrate that a morphogen changes its roles in time. However, the developmental stage we examined is late blastoderm stage. Therefore, it is possible that early detectable morphogens may have additional functions we have not examined. When embryos lose expressions of Bcd or Dl at early stages, many target genes fail to become activated. Loss of target gene expressions could be simply due to loss of activators. However, we cannot exclude possibilities that failure to prime genome without Bcd or Dl causes the loss of target gene expressions.

### *Mechanisms that support twist expression*

During *Drosophila* body patterning, a set of morphogen sensitive genes become independent of morphogens at later time points. We demonstrated that *sna* activation becomes independent of high Dl levels at late nc14. *Twi* becomes a major input for *sna* activation at the late timepoint. Like *sna*, *twi* activity also becomes independent of high Dl levels at late nc14, possibly supported by autoregulation feedback. However, simple autoregulation feedback raises questions in observations we made in this thesis. We were able to dial down *Twi* levels by increasing Dl degradation levels by increasing the length of blue light illumination at nc14 (Chapter 2, Figure 4H). With short illumination (i.e. less than 5 min of blue light illumination), the embryos exhibited

exponential increases in Twi levels, similar to wildtype embryos. In contrast, with longer blue light illumination, Twi levels reach plateau. If Twi activity is controlled by simple autoregulation feedback without any inputs from additional activators, the levels will reach a steady state. Thus, the two modes of Twi levels suggest that complex regulations may be involved in *twi* expression control rather than a simple autoregulation feedback.

Alternatively, *twi* activity may be supported by a memory module (e.g. positive feedforward loop). The memory module emphasizes on the history of exposure to inputs for gene activation, allowing it to be independent of inputs at later timepoints. In the case of *dpp* expression in *Drosophila* wing disc, *dpp* expression is maintained in the region where the transient Hh signal is received. It is supported by a memory module that is initiated by early Hh signaling, rather than autoregulation feedback (Nahmad and Stathopoulos 2009). It is interesting to know whether *twi* share a similar mechanism, which may explain better how two modes of *twi* existed in our experiments. Furthermore, identifying factors involved in the memory module will provide another layer of gene regulatory networks involved in *Drosophila* body patterning processes.

## References

- Ashe HL, Briscoe J. 2006. The interpretation of morphogen gradients. *Development* **133**: 385–394.
- Fukaya T, Lim B, Levine M. 2016. Enhancer Control of Transcriptional Bursting. *Cell* **166**: 358–368.
- Gurdon JB, Mitchell A, Mahony D. 1995. Direct and continuous assessment by cells of their position in a morphogen gradient. *Nature* **376**: 520–521. <http://dx.doi.org/10.1038/376520a0>
- Lettice LA. 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics* **12**: 1725–1735. <http://dx.doi.org/10.1093/hmg/ddg180>
- Lewis EB. 1978. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**: 565–570.
- Little SC, Tkačik G, Kneeland TB, Wieschaus EF, Gregor T. 2011. The Formation of the Bicoid Morphogen Gradient Requires Protein Movement from Anteriorly Localized mRNA. *PLoS Biology* **9**: e1000596. <http://dx.doi.org/10.1371/journal.pbio.1000596>
- Nahmad M, Stathopoulos A. 2009. Dynamic interpretation of hedgehog signaling in the *Drosophila* wing disc. *PLoS Biol* **7**: e1000202.
- Reeves GT, Trisnadi N, Truong TV, Nahmad M, Katz S, Stathopoulos A. 2012. Dorsal-ventral gene expression in the *Drosophila* embryo reflects the dynamics and precision of the dorsal nuclear gradient. *Dev Cell* **22**: 544–557.
- Shirokawa JM, Courey AJ. 1997. A direct contact between the dorsal rel homology domain and Twist may mediate transcriptional synergy. *Mol Cell Biol* **17**: 3345–3355.

## Appendices

A.

FGF signaling regulates cell proliferation and differentiation during morphogenesis of the *Drosophila* ovary to support fertility

This paper was written by Jihyun Irizarry and Angelike Stathopoulos and published in *Development Biology* in 2011.

## ABSTRACT

The *thisbe* (*ths*) gene encodes a *Drosophila* fibroblast growth factor (FGF), and mutant females are viable but sterile suggesting a link between FGF signaling and fertility. Ovaries exhibit abnormal morphology including lack of epithelial sheaths and muscle tissues that surround ovarioles. Here we investigated how FGF influences *Drosophila* ovary morphogenesis and identified several roles. Heartless (Htl) FGF receptor was found to be expressed within somatic cells at the larval and pupal stages, and phenotypes were uncovered using RNAi. Differentiation of terminal filament cells was affected, but this effect did not alter the ovariole number. In addition, proliferation of epithelial sheath progenitors, the apical cells, was decreased in both *htl* and *ths* mutants, while ectopic expression of the *Ths* ligand led to these cells' over-proliferation suggesting that FGF signaling supports ovarian muscle sheath formation by controlling apical cell number in the developing gonad. Additionally, live imaging of adult ovaries was used to show that *htl* RNAi mutants, hypomorphic mutants in which epithelial sheaths are present, exhibit abnormal muscle contractions. Collectively, our results demonstrate that proper formation of ovarian muscle tissues is regulated by FGF signaling in the larval and pupal stages through control of apical cell proliferation and is required to support fertility.

## INTRODUCTION

The *Drosophila melanogaster* ovary is a highly studied developmental system that has already provided many important insights into the biology of organ development. In particular, the *Drosophila* adult ovary has served as an excellent model for the interaction of germline stem cells (GSCs) with their somatic support cells known as the niche. In the ovary, the GSCs and niche facilitate egg production throughout the lifetime of the *Drosophila* female (Song et al., 2007, Xie and Spradling, 2000). Less is known regarding how the ovary is formed, but tight regulation of cell

proliferation, differentiation, and survival by signaling pathways appears critical. Regulated signaling ensures that all cell types within this organ develop in a balanced manner during this period of major growth of the ovary at the larval and pupal stages.

Each ovary in the *Drosophila* adult consists of 15–20 ovarioles that contain GSCs, their associated niche, and a chain of oocytes at various stages of development. At the apical region of an ovariole, a unique structure called the germarium resides. It is within this structure that two to three GSCs reside at the apical tip next to their niche composed of terminal filament (TF) cells and cap cells (Eliazer and Buszczak, 2011). Much insight into the mechanisms controlling GSC maintenance and differentiation has been uncovered in *Drosophila* owing to the ease of accessibility of these cells within adult ovaries and because the system is amenable to genetic manipulation (Kirilly and Xie, 2007). In contrast, less is known regarding how GSCs, their somatic niche, and muscle tissues that encapsulate the ovarioles (the epithelial and peritoneal sheaths) are formed as these events occur earlier, at the larval and pupal stages, as the ovaries develop.

Previous studies of ovary morphogenesis at the larval and pupal stages have focused on the role of signaling pathways in regulating cell number, proliferation, differentiation, and survival (review in Gilboa (2015); Sarikaya and Extavour (2015)). EGFR, JAK/STAT, and Hippo signaling is important in mediating cellular homeostasis during the period of extreme cell growth of the gonad at the larval stage. Specifically, EGFR regulates the number of primordial germ cells (PGCs) as well as their somatic support cells, the interstitial cells (ICs) (Gilboa and Lehmann, 2006, Matsuoka et al., 2013). Ecdysone hormone also has been shown to trigger cell proliferation and to control growth of the ovary through effects on the insulin receptor (InR) and Target of rapamycin (Tor) pathway, as well (Gancz and Gilboa, 2013). Additionally, Bone morphogenetic protein (BMP) and Hedgehog (Hh) signaling pathways positively regulate PGC cell division at the

larval stage (Sato et al., 2010). However, whether FGF signaling impacts ovary homeostasis and morphogenesis had not been previously investigated.

FGF signaling is involved in a multitude of important biological processes. FGF receptors (FGFRs) are a family of receptor tyrosine kinases. Upon receptor activation by ligand binding, various intracellular signaling pathways are induced (Feldman et al., 1995, Powers et al., 2000, Rottinger et al., 2008). To define a role for FGF signaling or to identify the specific molecular mechanisms involved can be challenging due to the complexity of the pathway. In humans and mice, for instance, 24 FGF and four FGFR genes have been discovered (Ornitz and Itoh, 2001), which support over one hundred possible FGF–FGFR complexes (Ornitz et al., 1996). Conversely, invertebrate systems have much simpler FGF signaling systems (Tulin and Stathopoulos, 2010a). In the case of *Drosophila*, three FGF and two FGFR genes have been discovered, supporting only three functional FGF–FGFR combinations (Kadam et al., 2009). The role of FGFR signaling in *Drosophila* as well as ligand choice varies and is context-specific (review in Bae et al. (2012)).

A role for fibroblast growth factor (FGF) signaling pathway in supporting ovarian development has been suggested in vertebrates, but no previous study has directly examined the role of FGF signaling in the *Drosophila* ovary. In vertebrates, it has been shown that both FGF ligands and receptors are expressed within follicular cells of vertebrate ovaries, including human (Berisha et al., 2006, Buratini et al., 2007). Furthermore, FGF addition to cultured ovarian tissues leads to cells' proliferation, and high levels of FGF signaling are correlated with many cancers including that of the ovary (Basu et al., 2014, Ropiquet et al., 2000). In particular, vertebrate FGF-8 exhibits gonad-specific expression, within the ovary and testes, suggesting that this signaling pathway plays an important, yet currently uncharacterized role in supporting gonad development (Valve et al., 1997). Keeping FGF signaling properly regulated is important for

normal ovary development, but its exact role in supporting gonad development is unclear. Furthermore, FGF signaling is conserved as its biological roles and structural properties appear similar in *Drosophila* and higher vertebrates (Huang and Stern, 2005, Tulin and Stathopoulos, 2010b). Studies of how FGF signaling impacts *Drosophila* ovary morphogenesis have the potential to provide novel insights into conserved functions and/or regulatory mechanisms acting in other organisms, including vertebrates.

In the current study, we investigated the role of FGF signaling in supporting *Drosophila* ovary morphogenesis and found that this signaling pathway has several roles spanning multiple stages of development. At the larval stage, our results demonstrate a role for the Htl FGFR in controlling specification of the adult stem cell niche through regulation of TF cell differentiation; in the larval and pupal stages, this pathway also supports migration of a somatic cell population in the ovary, the apical cells, through regulation of these cells' proliferation. These earlier functions are necessary for the proper specification of the epithelial sheaths that surround individual ovarioles to support proper oocyte development and, thus, fertility.

## MATERIALS AND METHODS

### *Fly stocks*

*Drosophila* stocks were kept at 25°C, unless otherwise noted. The *yw* stock is used as wildtype. To generate *ths* mutants viable flies, *ths*<sup>e02026</sup>/*CyO ftz-lacZ* (CFLZ) (Stathopoulos et al., 2004) and *Df(2R)ths238/CFLZ* (Kadam et al., 2009) were crossed to generate transheterozygotes. GAL4 lines used for genetic analysis were: *ptc*GAL4; +/*SM6;TM6B,Tb* (from Dr. Sean Carroll, University of Wisconsin, USA), *c587*-GAL4 (from Dr. Hilary Ashe, University of Manchester, UK), and *nos*-GAL4 (Dr. Hilary Ashe). UAS lines utilized for genetic analysis were: UAS.*htl.RNAi* [40627, Vienna *Drosophila* Research Center (VDRC)], UAS.*thsRNAi/CyO* (24538, VDRC),

UAS.*pyr.RNAi* (36523, VDRC), UAS.*ths* (AMS289-22; (Stathopoulos et al., 2004)). For temporal control, UAS.*htl.RNAi* was crossed with  $w^+$ ; *Sco/Cyo;tub-GAL80<sup>ts</sup>* [7018, Bloomington Drosophila Stock Center (BDSC)]. As necessary, *If/CyO,actin-gfp; MKRS/Tm3,Ser,actin-gfp* (from Dr. Kai Zinn, California Institute of Technology, US) was used as a marked balancer at the larval stage.

To examine expression patterns of *htl* or *ths*, *htl*-GAL4 or *ths*-GAL4 lines generated by another study were crossed with UAS.gfp (Pfeiffer et al., 2008). Numerous GAL4 lines were assayed but only a subset (underlined) drove expression in the ovary: *htl*-GAL4 lines (BDSC lines: 47240, 40668, 40669, 48004, 40706, 47277, 40707, 40708, 48431, 47278,47279), *ths*-GAL4 lines (BDSC lines: 40051, 47051, 40049, 40050, 40052, 48624, 48355).

We also generated five *pyr*-GAL4 lines, by cloning 1-3 kB fragments of DNA flanking the *pyr* gene into the Gateway donor vector and pBGUw vector (Pfeiffer et al., 2008 #1883) to create GAL4 drivers HV01-05; of these, only HV03 and HV04 supported expression in the developing ovary. Primer sequences are provided (Table S1).

To examine Htl localization, an inframe insertion of the Cherry reporter was inserted into a large *htl* rescue construct “*htl-mcherry*”. The 52 kb *htl* P[acman] construct was generated using recombineering mediated gap repair performed as in Venken (2006) using SW105 cells. Insertion of Cherry just before the stop codon of *htl* was performed by standard recombineering techniques, using the Cherry-SV40-frt-kan-frt,plasmid modified from the GFP-SV40-frt-kan-frt plasmid kindly provided by Dr. Eric Davidson (Caltech). The kan cassette was flipped out by arabinose induction of Flp in the SW105 cells (Warming et al., 2005).

### *Collection and aging*

Eggs were collected in fresh vials for two hours to prevent overcrowding. Once flies were removed, vials were incubated at 25°C for adequate length of time: 72 hours (h) for early-larval third instar, 96h for mid-larval third instar, 120h for late-larval third instar, 144h for early pupae, 168h for middle pupae, and 192 h for late pupae. For UAS.*htl*.RNAi;GAL80<sup>ts</sup>, after a two-hour-egg collection, vials were incubated at 18°C for 72h, transferred to 29°C, and incubated for additional 72-96h to examine mid-larval third instar or late-larval third instar, respectively. To examine ovaries at the early pupal stage, vials were incubated at 18°C for 216h, transferred to 29°C, and incubated for an additional 24h.

*Fixation, immunocytochemistry, and in-situ hybridization*

Dissected ovaries were fixed in 33% paraaldehyde in PBT solution for 20 minutes (min) at a room temperature (RT). The fixed ovaries were washed with PBS three times and incubated in blocking solution (10% BSA in PBT) for 1h. After the blocking, they were stained with adequate primary antibodies for ~18 h at 4°C. After the primary antibodies staining, the samples were washed with 1:10 diluted blocking solution 4x, with 30 min incubation for each wash. Secondary antibodies in 1:100 diluted blocking solution were added into the sample, and incubated further for ~18h at 4°C. The samples were washed with PBT 3x and mounted in Vecta-shield mounting medium (Vector Laboratories). Images were captured with a Pascal confocal microscope (Zeiss).

The primary antibodies utilized were rabbit anti-GFP (1:400; Molecular Probes), rabbit anti-RFP (1:1000, Rockland), rat anti-Vasa (1:40; Developmental Studies Hydridoma Bank (DSHB)), mouse anti- $\alpha$ -Spectrin (1:133; DSHB), mouse anti-Fasciclin III (1:60; DSHB), mouse anti-Engrailed (1:400; DSHB), rabbit anti-Phospho-Histone-H3 (1:1000; Rockland), rabbit anti-Dof (gift from Dr. Maria Leptin, Univ. of Cologne, Germany). All the secondary antibodies were from Invitrogen.

Rhodamine phalloidin-Alexa488 (Invitrogen) and propidium iodide (PI; Sigma) were used to detect ovarian muscle tissues and cell death, respectively. For these detections, standard protocols provided by the manufacture were used.

For detection of *htl*, *ths*, or *pyr* transcripts, *in-situ* hybridization using gene specific riboprobes were used as previously described (Stathopoulos et al., 2004). Ovaries were mounted in Permount (Fisher Scientific).

## RESULTS

### ***ths* mutants are sterile and have defective ovarian muscle sheaths**

A role for FGF signaling in female reproduction was suggested by the finding that hypomorphic mutants for one Htl-receptor ligand, Ths, are viable but sterile. *ths* mutant adult females were obtained by crossing *ths<sup>e02026</sup>*, a piggyback mutation that is semi-lethal, to *Df(2R)ths238*, a small deficiency that deletes the region containing the *ths* gene (Kadam et al., 2009). Each of the two ovaries present in the abdomen of wildtype *Drosophila melanogaster* females consists of ~15 ovarioles (Fig. 1A). However, in FGF mutant females [i.e., *ths<sup>e02026</sup>/Df(2R)ths238*], two ovaries are present, but their overall structure appears disorganized (Fig. 1B). Furthermore, *ths* mutant females rarely deposit eggs (Fig. 1E), and yet oocytes do not accumulate within their abdomens suggesting a block to oogenesis (data not shown). To confirm that the observed phenotypes relate to loss of FGF signaling, a *ths* rescue transgene was introduced into the *ths* mutant background (Fig. 1C; see Materials and methods). Both the morphology and fertility phenotypes were rescued (Fig. 1C; data not shown), supporting the view that FGF signaling through the Ths FGF ligand is required to support ovary function.

Next, we examined the Htl-receptor loss-of-function phenotype and compared it with that of the Ths ligand. *htl* null mutants are zygotically lethal, therefore RNAi was used. *htl* levels were

reduced by driving a UAS.RNAi hairpin construct with a somatic cell specific driver, *c587.GAL4*. Upon *htl* knock-down (KD), the organization of the ovary was disrupted (Fig. 1D), and the egg/embryo deposit percentage was significantly decreased compared to the control (i.e. driver alone; Fig. 1E). The *htl* RNAi phenotype was not as severe as *ths* mutants and may relate to the observation that few *c587>htl.RNAi* progeny make it to the adult stage. Survival may select for weak *htl* KD. The similarity of *ths* and *htl* mutant phenotypes suggested that FGF signaling through the Htl receptor and Ths ligand regulates ovary morphogenesis.

To provide insight into these defects, we examined the expression of the FGF receptor, *htl*, to infer the location of FGF signaling action. In ovarioles isolated from the adult ovary, *htl* transcripts were identified, specifically, within cells of the epithelial sheath using in situ hybridization with gene-specific riboprobes (Fig. 1G). The sheath covers each ovariole and is composed of a layer of squamous epithelium surrounded by bands of muscle (Hudson et al., 2008). The epithelial sheath secretes a thick basement membrane and provides structural support to the ovariole (Fig. 1F; Cummings, 1974).

To confirm the expression domain of Htl protein within the epithelial sheath, we used several approaches. First, a fusion gene was created in which the monomeric Cherry fluorescent protein (FP) sequence was inserted into the *htl* gene in the context of a 54 kb rescue construct to create a C-terminal fusion of this FP to Htl (“Htl-mCherry”) so that the Cherry reporter could be used as a proxy for Htl FGFR protein expression. In ovaries isolated from Htl-mCherry transgenic females, strong anti-RFP staining was detected in the epithelial sheaths (Fig. 1H), as well as in the peritoneal sheaths, a distinct muscle tissue that surrounds the entire ovariole collective (data not shown; Fig. 1F). Next, we examined cis-regulatory sequences acting to support *htl* expression. In a previous genome-wide study of non-coding DNA sequences supporting expression in the brain (Pfeiffer et al., 2008), sequences flanking the *htl* gene were placed upstream of a heterologous

gene, GAL4, encoding a transcription factor able to support ectopic expression through UAS sequences (Phelps and Brand, 1998). We found that a particular *htl*.GAL4 driver (*htl*.GAL4<sup>GMR93H07</sup>), only one out of 11 tested, supports expression in the ovary and that this expression was present in the epithelial sheath (Fig. 1I). Results from all three expression assays (in situ, FP fusion, and cis-regulatory activity) support the view that *htl* is expressed in the ovarian muscle sheaths, an ovarian tissue that has been little studied (Cummings, 1974, Hudson et al., 2008). We hypothesized that FGF signaling regulates some aspect of ovarian muscle sheaths function and/or morphogenesis that relates to the phenotypes observed in *ths* mutants, namely loss of overall structural organization within adult ovaries (Fig. 1B).

As a result, the epithelial sheaths of wildtype and *ths* mutant ovaries were examined. In *ths* mutant ovarioles, no epithelial sheath was observed as assayed by localization of  $\alpha$ -Spectrin, a constituent of the sub-membrane cytoskeleton of epithelial cells (Fig. 1K, compare with Fig 1J). Moreover, while organized actinfilaments are clearly visible within epithelial sheaths of wildtype ovaries (Fig. 1L), in *ths* mutant ovaries, only a limited amount of actin staining was identified and it was confined to a few disorganized actin-rich masses (Fig. 1M). No evidence of peritoneal sheaths was present in *ths* mutants (data not shown). Previous studies have shown that apical cells (ACs), a type of somatic gonadal cell residing at the anterior of the developing ovary, are responsible for specification of both epithelial and peritoneal sheaths during gonadal development (Cohen et al., 2002, King, 1970). Therefore, we investigated whether these identified muscle tissue malformations in the adult ovary of *ths* mutants might relate to defects at earlier stages during ovary morphogenesis.

**Htl FGF receptor is expressed within apical cells throughout the course of their migration**

Morphogenesis of the adult ovary starts in the embryo and continues throughout the larval and pupal stages and involves many changes (Fig. 2A; review in Gilboa (2015)). To start, in embryos, primordial germ cells (PGCs) undergo a migration to the somatic gonadal mesoderm. These somatic gonadal mesoderm cells serve as the precursor population for several types of cells present in the developing ovary, later, at the larval stage. Specifically, at mid-larval third instar (ML3), somatic cells include ACs (discussed above), ICs, and swarm cells (Fig. 2A, left). These cell types all actively undergo mitotic cell divisions throughout larval and pupal development and further differentiate into more specialized cell types (Couderc et al., 2002). For example, by late-larval third instar (LL3) following ML3 but before pupation, some of the ACs differentiate into terminal filament (TF) cells to specify a component of the adult germline stem cell niche. In addition, cell movements support morphogenesis of the ovary. For instance, at the early pupal stage, a subset of ACs initiates migration from the anterior of the gonad towards the posterior (Fig. 2A, middle). By ~24 h after puparium formation, at the middle pupal stage, the migration is complete. As a result of the migration, a subset of ACs make direct contact with the primordial ovarioles to specify the epithelial sheath, whereas others migrate to the periphery of the gonad to specify the peritoneal sheath (Fig. 2A, right).

With the aid of the *htl-mcherry* reporter, expression of Htl-mCherry protein was identified in somatic tissues of the developing gonad (Fig. 2B–D). Using an anti-RFP antibody, staining was detected in ACs present at the anterior of the gonad at ML3 (Fig. 2B). At later stages, expression remained detectable in the ACs as they migrate from the anterior toward the posterior end of the developing ovary at the early pupal stage (Fig. 2C) as well as at the end of their migration at the middle pupal stage (Fig. 2D). In addition, weak expression of Htl-mCherry was identified in other somatic cells, possibly the swarm cells, which also undergo a posteriorly-directed migration to form basal cells, precursors of basal stalk cells (Fig. 2B, arrowhead).

To provide additional insight into the role of FGF signaling in the developing ovary, other reporters of FGFR-activation were examined. Intracellular signaling downstream of FGFR-activation requires Downstream of FGFR (Dof, also known as Stumps/Heartbroken), an intracellular adaptor protein (Imam et al., 1999, Michelson et al., 1998, Vincent et al., 1998). Dof expression was examined using an anti-Dof antibody and localized to differentiating TF cells at the ML3 stage (Fig. 2E), migrating ACs at the early pupae stage (Fig. 2F), as well as ACs having completed their migration at the middle pupae stage (Fig. 2G). This pattern overlaps with the domain of Htl FGFR expression inferred using the Htl-mCherry fusion. However, the Dof expression domain encompassed only a subset of the receptor expression domain, suggesting that Dof might possibly differentially influence downstream signaling pathway activation.

To provide additional insight into the activation domains of Htl FGFR, a receptor tyrosine kinase, we assayed where signaling was active using an antibody against the dual-phosphorylated form of MAPK (dpERK) (Gabay et al., 1997). At the early larval third instar (EL3) stage, before TFs are specified, dpERK staining was found broadly distributed in somatic cells including ACs (Fig. S1AA", arrow); whereas, in slightly older gonads (i.e., ML3), dpERK staining was detected in differentiating TF cells as well as in non-differentiated ACs (Fig. S1BB"; arrowhead and arrow, respectively). Previous studies have detected dpERK staining within the IC population, somatic support cells for PGCs present in the medial domain of developing ovaries, and demonstrated that this activation relates to intracellular signaling downstream of EGFR activation (Gilboa and Lehmann, 2006). As we detected dpERK in anterior regions (Fig. S1A') including the domain of TF specification (Fig. S1B'), these results suggest that signaling downstream of other receptor tyrosine kinases, possibly FGFR, may also occur at this stage.

Collectively, these data show that FGF signaling is spatially and temporally positioned within domains able to impact differentiating TF cells and ACs.

### **Htl-receptor activation by the Pyr FGF ligand regulates differentiation of TF cells**

To assay a role for FGF signaling in the developing ovary, we assayed phenotypes resulting from loss of *htl*. *htl* mutants could not be assayed directly because null mutants are zygotically lethal at the embryo/early larval stages (data not shown; Gisselbrecht et al., 1996). Therefore an RNAi approach was taken. Specifically, the pan-somatic driver *c587.GAL4* was used to support expression of two UAS-RNAi transgenes (VDRC40627 and VDRC6692) targeted to distinct regions of the *htl* gene.

To start, the role of FGF signaling in supporting TF development was investigated. The earliest Dof expression was observed at the EL3 stage, in a domain in which TF cells are specified at the later larval stage (Fig. 3A), suggesting a role for FGF in TF development. TF cells start off lens-shaped at the ML3 stage (Fig. 3D), and they later transform into blunt-edged disc shaped cells as the differentiation process progresses resulting in complete separation between TF stacks (Godt and Laski, 1995, Sahut-Barnola et al., 1995). At the LL3 stage, TF cells resolve into a number of individual stacks that each contain approximately 7–9 cells exhibiting flattened nuclei. Differentiated TF cells can be identified by expression of the transcription factor Engrailed (En) within nuclei (Fig. 3G), which is maintained throughout gonadal development. Upon *htl* KD, the number of En<sup>+</sup> TF cells was reduced at LL3 stage (Fig. 3H, white arrowhead). In contrast, when a constitutively active form of the Htl receptor (“*htl lambda*”) was expressed within somatic tissues, ectopic expression of En was observed (Fig. 3J, green arrow). When a germline specific GAL4 driver (*nos.Gal4*) was used to support either *htl*.RNAi or expression of constitutively active Htl receptor, no effect on En expression within TFs was observed (data not shown). These results

suggest that the Htl FGF receptor acts in somatic cells to regulate TF specification, possibly, by controlling gene activation.

We further examined a role for each of the two ligands for the Htl receptor, Pyr or Ths, in supporting FGF signaling in this function. Neither *ths* mutant nor RNAi exhibited any change in TF morphology (Fig. S2E, compared with Fig. S2C; data not shown) or En expression (Fig. S2G, compared with Fig. S2F; data now shown) and suggested, instead, that Pyr is fulfilling Htl receptor activation in this role. The expression domain of Pyr FGF ligand was examined through assay of associated cis-regulatory domains. *pyr.GAL4<sup>HA04</sup>* supports expression in somatic cells (Figs. 3B and S2A), while another cis-regulatory sequence from the vicinity of the *pyr* gene was found to support expression in PGCs (*pyr.GAL4<sup>HV03</sup>*; Fig. S2A and B).

To assay a role for Pyr, *pyr* KD by RNAi, through GAL4-mediated expression of a UAS-driven hairpin construct, was chosen due to zygotic lethality of all assayed *pyr* mutants at the embryonic/early larval stage. However, the only available RNAi construct directed to *pyr* (VDRC36523) has a number of off-targets. While *pyr* KD resulted in a decreased number of En<sup>+</sup> cells at the LL3 stage (Fig. 3I), a phenotype shared with *htl* KD, earlier phenotypes were observed as well that were not shared with *htl* KD. Upon *pyr* KD, TF morphology was affected; few lens-shaped cells were observed at stage ML3 (Fig. 3F), which is a more severe effect than *htl* KD (Fig. 3E, compare with Fig. 3D), suggesting either that off-targets of *pyr*.RNAi36523 are responsible and/or that the *htl* RNAi presents only a partial loss-of-function phenotype. In contrast, when *pyr* was knocked-down in the germline cells, TF morphology appeared normal (Fig. S2D, compared with Fig. S2C). Taken together, FGF signaling, likely mediated by somatically expressed Pyr at the larval stage, contributes to ovarian morphogenesis by directly promoting TF cell differentiation.

Next, we investigated whether these observed differentiation defects, changes in En<sup>+</sup> cell numbers within TF precursor cells, affected TF number at later stages. The number of TF stack per ovary and the number of TF cells per TF stack were counted at the LL3 stage based on staining with rhodamine phalloidin and TO-PRO3 (marking TF cells and nuclei, respectively; Fig. S3). No effects on TF cell or stack numbers were observed in *htl* KD or *ths* mutant ovary at these earlier stages (Fig. 3K); neither were any effects on the adult stem cell niche identified at the adult stage (data not shown). These results suggested that the identified role for FGF in regulating gene expression within TF precursor cells is unlikely to relate to the fertility defects observed in FGF mutants. We focused instead on whether FGF signaling has additional roles in the developing gonad at later stages that relate to fertility.

### **Htl-receptor activation by the Ths FGF ligand regulates apical cell proliferation**

At the transition stage from larva to pupa, *htl* transcripts were detected in the ACs by in situ hybridization (Fig. 4A). The domain of FGF signaling in developing gonads also was investigated by assaying expression supported by the various *htl*.GAL4 and *ths*.GAL4 drivers (Fig. 4H), which were found to support expression in different somatic cell types. Once again, only one driver *htl*.GAL4<sup>GMR93H07</sup> supported expression in the developing gonad within ACs at the ML3 and early pupae stage (Fig. 4B and C, respectively); the same construct that was found to support expression in the epithelial sheath of adult ovaries (Fig. 1I).

*ths* transcripts were detected in TF and presumptive basal stalk cells by in situ hybridization (Fig. 4D). Moreover, we identified cis-regulatory sequences flanking the *ths* gene able to drive expression in these regions. *ths*.GAL4<sup>GMR79H07</sup> and *ths*.GAL4<sup>GMR79G11</sup> drivers both support expression in TF cells (Fig. 4E and F, respectively), whereas the *ths*.GAL4<sup>GMR79G11</sup> driver alone

supports expression, additionally, within the basal/ basal stalk cells (Fig. 4F). Based on (i) the specific expression of *Htl* within ACs; (ii) the abutting domains of expression of the *htl* and *ths* genes (Fig. 4G); as well as (iii) the finding that the ovarian muscle tissue defects are present in *ths* mutants, we investigated whether the *Ths* FGF ligand activates *Htl* to support ovarian muscle sheath formation through regulation of ACs.

At the middle pupal stage, AC migration from the anterior to posterior region is apparent (Fig. 2D). Primordial ovarioles lose direct contact with each other, as the space between them is filled with ACs that migrate in between them; first at anterior regions (Fig. 5C) and, subsequently, at posterior regions as well (Fig. 5E). In contrast, many fewer ACs were observed between the primordial ovarioles of *ths* mutant ovaries (Fig. 5B, compared with Fig. 5A; and Fig. 5D, compared with Fig. 5C). This finding suggested that FGF signaling through *Htl* and *Ths* controls apical cell migration and/or proliferation.

Even in *ths* mutants, however, somatic cells were identified near basal stalks at the posterior suggesting AC migration was occurring, though the number of cells present in this domain was reduced compared with wildtype (Fig. 5F, compared with Fig. 5E). AC number and organization phenotypes in *ths* mutants were examined more closely by introducing the *htl-mcherry* transgene, which marks ACs, into this background. At the middle pupal stage when AC migration completes, we found that mCherry-expressing ACs do reach their final destination (the posterior of the developing gonad) in *ths* mutants (Fig. 5G, white arrows). However, *ths* mutant exhibited aberrant muscle tissues formation at later stages. The muscle tissues failed to completely encompass the whole ovary and appeared torn (Fig. 5I, compared with Fig. 5H). In *htl* KD ovaries, size reduction of the apical cell population was apparent (Fig. 6B, compared with Fig. 6A). In contrast, when *Ths* was ectopically expressed in somatic cells, ACs over-proliferated (Fig. 6C,

compared with Fig. 6A). These results demonstrated that FGF signaling is not required for AC migration but suggested instead that FGF signaling regulates AC number.

To test this idea directly, we examined the number of mitotic ACs by counting Phospho-Histone-H3 (PH3) positive ACs (e.g. Deng et al., 2001). In the *htl* KD ovaries upon expression of *htl*.RNAi40627 or *htl*.RNAi6692 in somatic cells, significant decreases in PH3 positive ACs were observed (Fig. 6F). Similarly, *thsmutant* ovaries also exhibited a reduction in the number of PH3 positive ACs compared with ovaries from *ths* heterozygotes, which served as control (Fig. 6F). In contrast, when *Ths* was ectopically expressed in somatic cells, the number of PH3 positive ACs was increased significantly (Fig. 6E,F, compared with Fig. 6D). Collectively, these results suggest that FGF signaling mediated by interaction between *Htl* FGFR and *Ths* FGF controls the AC population size by promoting cells' proliferation, to support AC migration and proper ovarian muscle generation.

### **FGF signaling is required for ovarian muscle tissue morphogenesis during ovarian development**

We further investigated whether these larval and pupal FGF signaling functions relate to phenotypes observed in the adult ovaries for *ths* mutants (e.g., lack of epithelial sheath; Fig. 1B). In the wildtype adult ovary, the peritoneal sheath is composed of thick muscle bundles encompassing each ovary that are aligned perpendicular to the anterior–posterior (AP) axis of the ovary. Additionally, thin muscle fibers are also present on the peritoneal sheath, which is organized in a parallel manner to its AP axis (Hudson et al., 2008). In *htl* KD ovaries, the muscle fibers on the peritoneal sheath appeared disorganized as they no longer aligned perpendicular to the AP axis

(Fig. 7B, arrowhead, compared with 7A) and failed to encompass the entire ovary (Fig. 7B, white arrow).

*htl* KD ovarioles exhibited additional structural abnormalities related to organization of egg chambers. In wildtype, regardless of age, germaria and egg chambers were aligned linearly (Fig. 7C and D). In contrast, when *htl* was knocked down by expressing *htl*.RNAi40627 or *htl*.RNAi6692 in somatic cells, the linear alignment of ovarioles was disrupted. Young egg chambers, isolated from two day-old females, appeared clumped beside the germarium (Fig. 7E and G, yellow arrow, compared with Fig. 7C); a phenotype described previously as a “flop-down” egg chamber (Cohen et al., 2002). In ovaries isolated from older flies, ten days in age, this phenotype was exacerbated (Fig. 7F and H, yellow arrow, compared with Fig. 7D).

This result suggested that FGF coordinates movement of egg chambers within ovarioles through regulation of ovarian muscles. To investigate, live imaging was used to observe contraction of ovaries when cultured in vitro. The wildtype ovary exhibited smooth and constant rhythmic contraction (Movie S1), as described previously (Middleton et al., 2006). However, in *htl* KD ovaries, in which *htl*.RNAi40627 or *htl*.RNAi6692 was induced in somatic cells by *c587*.GAL4 at earlier stages, muscle contractions were present but appeared uncoordinated. In some *htl* KD ovaries, though contractions appeared to extend throughout the length of the ovary, erratic movements resulted (Movie S2, compared with Movie S1). In other cases, muscle contractions were limited to the anterior region of the ovary and absent from the posterior (Movies S3, S4 compared with Movie S1). As *htl* RNAi also resulted in a decrease in egg-laying, collectively, these data suggest that proper organization of ovarian muscle tissues is crucial to maintain coordinate muscle contractions required to support fertility.

To determine whether these adult phenotypes relate to secondary roles for FGF at this adult ovarian stage or instead stem from earlier functions (e.g. proliferation of ACs at the larval/pupal

stages), we examined mutant phenotypes associated with *htl* KD, specifically, at the adult stage. A *htl*.RNAi40627 transgene was expressed via *htl*.GAL4<sup>GMR93H07</sup> together with GAL80<sup>ts</sup> to support temporal control of the KD within the epithelial sheath of adult ovaries (see Materials and methods). In *htl* KD at the adult stage, epithelial and peritoneal sheath muscle tissues were present (Fig. 7J) and the muscle tissues appeared morphologically normal (Fig. 7I). Ovarioles were aligned linearly regardless of age (Fig. 7J). Furthermore, muscle contractions (Movie S5) and egg-laying frequency (data not shown) were normal as well. Taken together, these data demonstrate that Htl function at earlier stages in the larval and pupal stages is required for proper ovarian muscle tissue morphogenesis to support fertility.

## DISCUSSION

The data presented demonstrate that FGF signaling acts in the gonadal somatic cells throughout larval and pupal development of the *Drosophila* ovary. Early, at the larval stage, FGF signaling also promotes TF cell differentiation through the Htl FGFR and its FGF ligand Pyr. An additional role for FGF signaling in supporting AC migration, at a slightly later stage, was uncovered that is required for normal ovary function. Rather than providing a directional cue to support AC migration, our results are consistent with the view that FGF acts to regulate AC number. FGF signaling regulates AC number through influence on cell proliferation. In this manner, FGF ensures sufficient numbers of cells are present along the length of the developing gonad so that the epithelial and peritoneal sheaths can encompass the entire ovary. The sterility phenotype of *ths Drosophila* females is likely explained by this earlier role for FGF, in supporting development of this muscle tissue, during gonadal development. Without proper muscle development, oogenesis is not supported.

As a function of FGF signaling pathway during ovarian development, we identified defects in gene expression within TF cells upon loss of FGF signaling. Most of *htl* KD mutants did not develop into adults. However, a few escapers were observed, possibly due to incomplete KD of *htl* levels by the RNAi approach. When these escapers were examined at the adult stage in terms of TF cell development and functionality, no apparent phenotypes were observed. Thus, it is unclear how the FGF-specific differentiation defects identified here at the larval stage affect ovary function. Furthermore, in the adult ovary, TF cells are contiguous with the epithelial sheath (Cohen et al., 2002). However, we show here that FGF signaling has distinct functions in the TF cells versus ACs (epithelial sheath), suggesting regulation of these cell types is separable.

During ovarian development and oogenesis, expression of genes *ths*, *pyr*, and *htl* is dynamic. For instance, the expression of the *htl* gene, encoding the FGFR, is not limited to TF cells and ACs. *Htl* expression through the *Htl.mCherry* reporter is also detected within the swarm cells at ML3, basal cells at the early pupal stage, and within the epithelial/peritoneal sheaths at the adult stage, suggesting additional roles during gonadal development and oogenesis. It is likely that FGF signaling, which is expressed in multiple domains of the developing ovary, supports various roles and possibly promotes differentiation or proliferation of additional cell types not examined here. Furthermore, why multiple FGF ligands are necessary to support FGF receptor activation is unclear, but the prevailing view is that ligands may influence different FGF response outputs. In the ovary, *Pyr* and *Ths* appear to encompass distinct functions as *Pyr* regulates TF cell differentiation at the larval stage, while *Ths* controls AC number at the early pupal stage. Thus, with FGF signaling components and primary function defined, the *Drosophila* ovary is an excellent system to study molecular mechanisms regulating FGF signaling activity.

We show here that control of FGF signaling is critical for the regulation of ovary growth as it acts as a mitogen, a role demonstrated previously in *Drosophila* only in the formation of air

sacs of the tracheal system (Sato and Kornberg, 2002). In ovaries ectopically expressing the FGF8-like ligand, vast over-proliferation was observed, which is one of the hallmarks of cancer. Many cell-based studies and mouse models have demonstrated that FGF signaling promotes tumor cell proliferation. Frequent amplification of the FGF1 gene in humans, resulting in increased gene expression, has also been reported in ovarian cancer (Birrer et al., 2007). In ovarian cancer states exhibiting increased FGF1 ligand expression, this factor functions to promote angiogenesis in a paracrine fashion but also may function in an autocrine manner to promote tumor cell proliferation (Birrer et al., 2007, Turner and Grose, 2010). Taken together, study of the development of this organ may serve as an excellent system to study the interplay between FGF signaling and cell proliferation with the possibility of far-reaching implications to the cancer biology field.

For example, as multiple signaling pathways have been implicated in the control of cell proliferation during ovary development in *Drosophila*, it is likely that mechanisms exist to co-regulate the activation of multiple signaling pathways to integrate growth of this tissue in a balanced manner. The Hippo pathway cooperates with EGFR signaling to maintain homeostasis of ICs and PGCs, and it also interacts with JAK/STAT pathways to control number of TF cells at the larval stage (review in Gilboa (2015); Sarikaya and Extavour (2015)). As we found that the FGF ligand is produced by TF cells and provides the cue that stimulates AC proliferation via FGF signaling pathway activation, it is possible that Hippo, which specifies number of TF cells, plays a role in controlling FGF signaling to coordinate the number of muscle precursors produced relative to size of the ovary. Each ovarian cell type appears to utilize a different signaling pathway to regulate proliferation of cells to support balanced growth of the gonad.

Our data also suggest that ovarian muscle contractions provide mechanical support to promote fertility. In *htl* KD mutants, the ovaries exhibited incomplete generation of ovarian muscle sheaths, and the organization of muscle sheaths was aberrant. These ovaries had

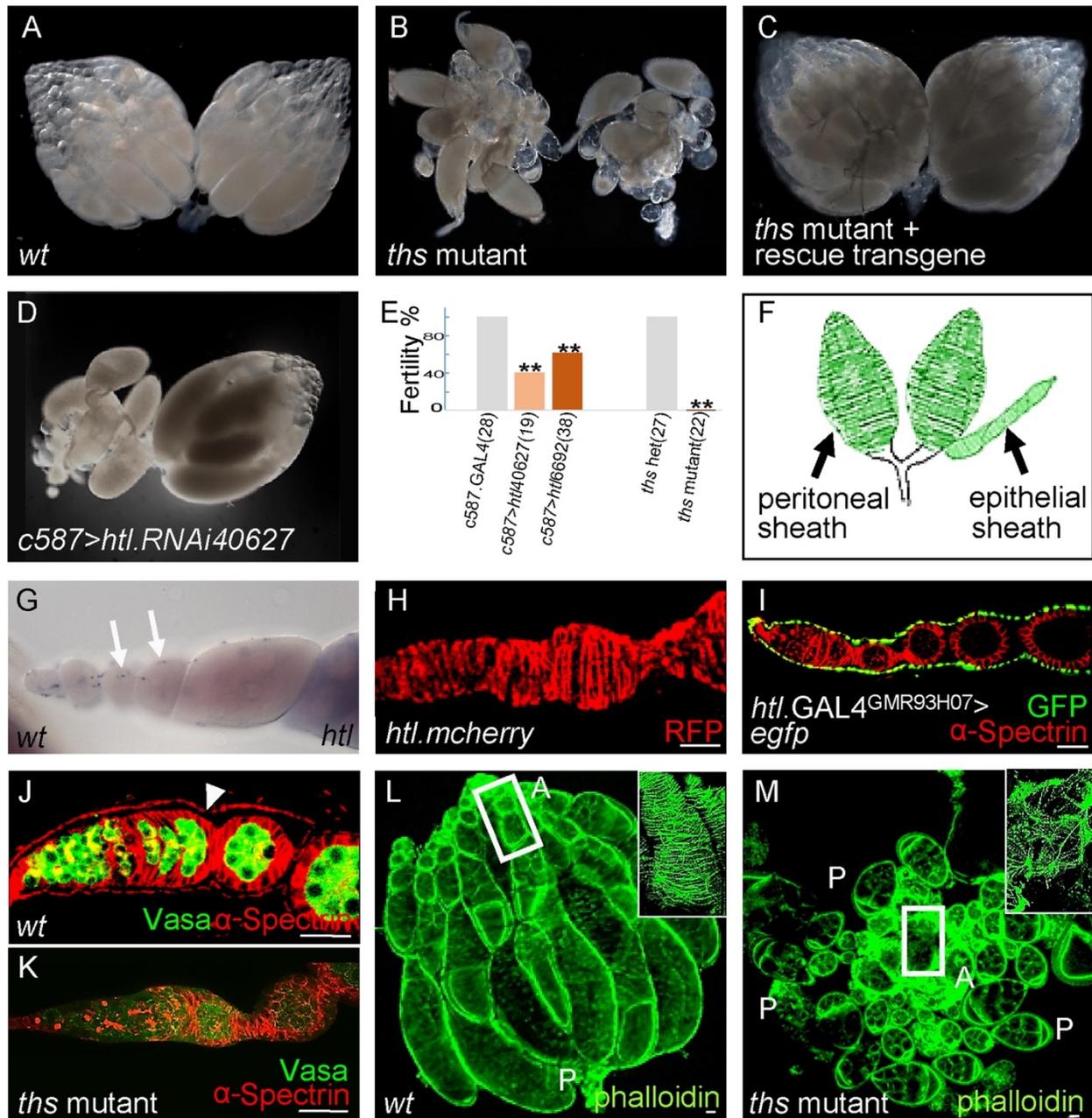
uncoordinate muscle contractions throughout the ovaries and exhibited “flop-down” egg chambers. Furthermore, their egg deposit frequencies were significantly decreased. In contrast, when *htl* was knocked down only at the adult stage, none of these phenotypes were observed, suggesting that phenotypes observed in *htl* KD adult ovaries, when *htl*.RNAi was activated early, are consequences of abnormal muscle sheaths generation at the larval/pupal stage. This viewpoint is also further supported by study of a *Drosophila* Wnt ligand, DWnt4, during ovarian development. During this stage, DWnt4 controls AC migration in a focal adhesion kinase dependent manner by activating the *Drosophila* Frizzled 2 receptor (DFz2) (Cohen et al., 2002). DWnt4 mutants fail to generate full-length epithelial sheaths. Also similar to the *thsmutant* phenotype, DFz2 mutants are viable and sterile (Chen and Struhl, 1999). Furthermore, like *htl* KD ovarioles, Dfz2 mutants and Dwnt4 mutants exhibit a “flop-down” ovariole phenotype (Cohen et al., 2002). Taken together, we propose that dysfunctional epithelial sheath formation, due to reduced FGF signaling (this study) and possibly also Wnt signaling, results in sterility.

## REFERENCES

- Bae, Y.K., N. Trisnadi, S. Kadam, and A. Stathopoulos. 2012. The role of FGF signaling in guiding coordinate movement of cell groups: guidance cue and cell adhesion regulator? *Cell Adh Migr.* 6:397-403.
- Basu, M., S. Mukhopadhyay, U. Chatterjee, and S.S. Roy. 2014. FGF16 promotes invasive behavior of SKOV-3 ovarian cancer cells through activation of mitogen-activated protein kinase (MAPK) signaling pathway. *J Biol Chem.* 289:1415-1428.
- Bodmer, R. 1995. Heart development in *Drosophila* and its relationship to vertebrates. *Trends Cardiovasc Med.* 5:21-28.
- Brown, E.H., and R.C. King. 1964. Studies on the Events Resulting in the Formation of an Egg Chamber in *Drosophila Melanogaster*. *Growth.* 28:41-81.
- Eliazer, S., and M. Buszczak. 2011. Finding a niche: studies from the *Drosophila* ovary. *Stem cell research & therapy.* 2:45.
- Gancz, D., and L. Gilboa. 2013. Insulin and Target of rapamycin signaling orchestrate the development of ovarian niche-stem cell units in *Drosophila*. *Development (Cambridge, England).* 140:4145-4154.
- Gilboa, L., and R. Lehmann. 2006. Soma-germline interactions coordinate homeostasis and growth in the *Drosophila* gonad. *Nature.* 443:97-100.
- Godt, D., and F.A. Laski. 1995. Mechanisms of cell rearrangement and cell recruitment in *Drosophila* ovary morphogenesis and the requirement of bric a brac. *Development (Cambridge, England).* 121:173-187.
- Kadam, S., A. McMahon, P. Tzou, and A. Stathopoulos. 2009. FGF ligands in *Drosophila* have distinct activities required to support cell migration and differentiation. *Development (Cambridge, England).* 136:739-747.
- King, R.C. 1970. The meiotic behavior of the *Drosophila* oocyte. *International review of cytology.* 28:125-168.
- Kirilly, D., and T. Xie. 2007. The *Drosophila* ovary: an active stem cell community. *Cell research.* 17:15-25.
- Ornitz, D.M., and N. Itoh. 2001. Fibroblast growth factors. *Genome Biol.* 2:REVIEWS3005.
- Ornitz, D.M., J. Xu, J.S. Colvin, D.G. McEwen, C.A. MacArthur, F. Coulier, G. Gao, and M. Goldfarb. 1996. Receptor specificity of the fibroblast growth factor family. *J Biol Chem.* 271:15292-15297.
- Pfeiffer, B.D., A. Jenett, A.S. Hammonds, T.T. Ngo, S. Misra, C. Murphy, A. Scully, J.W. Carlson, K.H. Wan, T.R. Laverty, C. Mungall, R. Svirskas, J.T. Kadonaga, C.Q. Doe, M.B. Eisen,

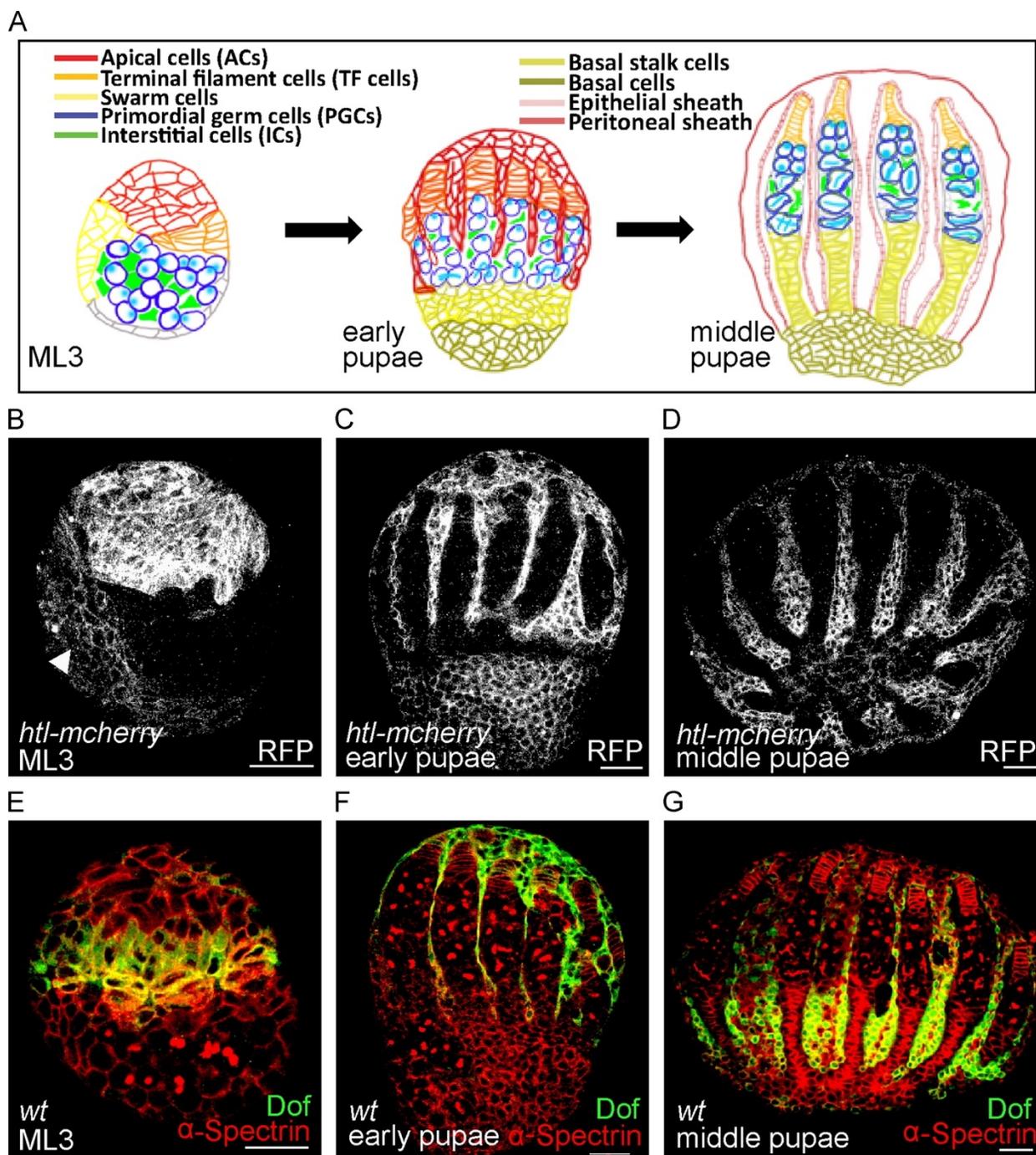
- S.E. Celniker, and G.M. Rubin. 2008. Tools for neuroanatomy and neurogenetics in *Drosophila*. *Proc Natl Acad Sci U S A*. 105:9715-9720.
- Ropiquet, F., D. Giri, B. Kwabi-Addo, A. Mansukhani, and M. Ittmann. 2000. Increased expression of fibroblast growth factor 6 in human prostatic intraepithelial neoplasia and prostate cancer. *Cancer Res*. 60:4245-4250.
- Shishido, E., N. Ono, T. Kojima, and K. Saigo. 1997. Requirements of DFR1/Heartless, a mesoderm-specific *Drosophila* FGF-receptor, for the formation of heart, visceral and somatic muscles, and ensheathing of longitudinal axon tracts in CNS. *Development*. 124:2119-2128.
- Song, X., G.B. Call, D. Kirilly, and T. Xie. 2007. Notch signaling controls germline stem cell niche formation in the *Drosophila* ovary. *Development (Cambridge, England)*. 134:1071-1080.
- Stathopoulos, A., B. Tam, M. Ronshaugen, M. Frasch, and M. Levine. 2004. pyramus and thisbe: FGF genes that pattern the mesoderm of *Drosophila* embryos. *Genes Dev*. 18:687-699.
- Tulin, S., and A. Stathopoulos. 2010. Extending the family table: Insights from beyond vertebrates into the regulation of embryonic development by FGFs. *Birth Defects Res C Embryo Today*. 90:214-227.
- Warming, S., N. Costantino, D.L. Court, N.A. Jenkins, and N.G. Copeland. 2005. Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res*. 33:e36.
- Wieschaus, E., and J. Szabad. 1979. The development and function of the female germ line in *Drosophila melanogaster*: a cell lineage study. *Dev Biol*. 68:29-46.
- Wilson, R., A. Battersby, A. Csiszar, E. Vogelsang, and M. Leptin. 2004. A functional domain of Dof that is required for fibroblast growth factor signaling. *Mol Cell Biol*. 24:2263-2276.
- Xie, T., and A.C. Spradling. 2000. A niche maintaining germ line stem cells in the *Drosophila* ovary. *Science*. 290:328-330.
- Yeh, J., and R. Osathanondh. 1993. Expression of messenger ribonucleic acids encoding for basic fibroblast growth factor (FGF) and alternatively spliced FGF receptor in human fetal ovary and uterus. *The Journal of clinical endocrinology and metabolism*. 77:1367-1371.

## Figures and Legends



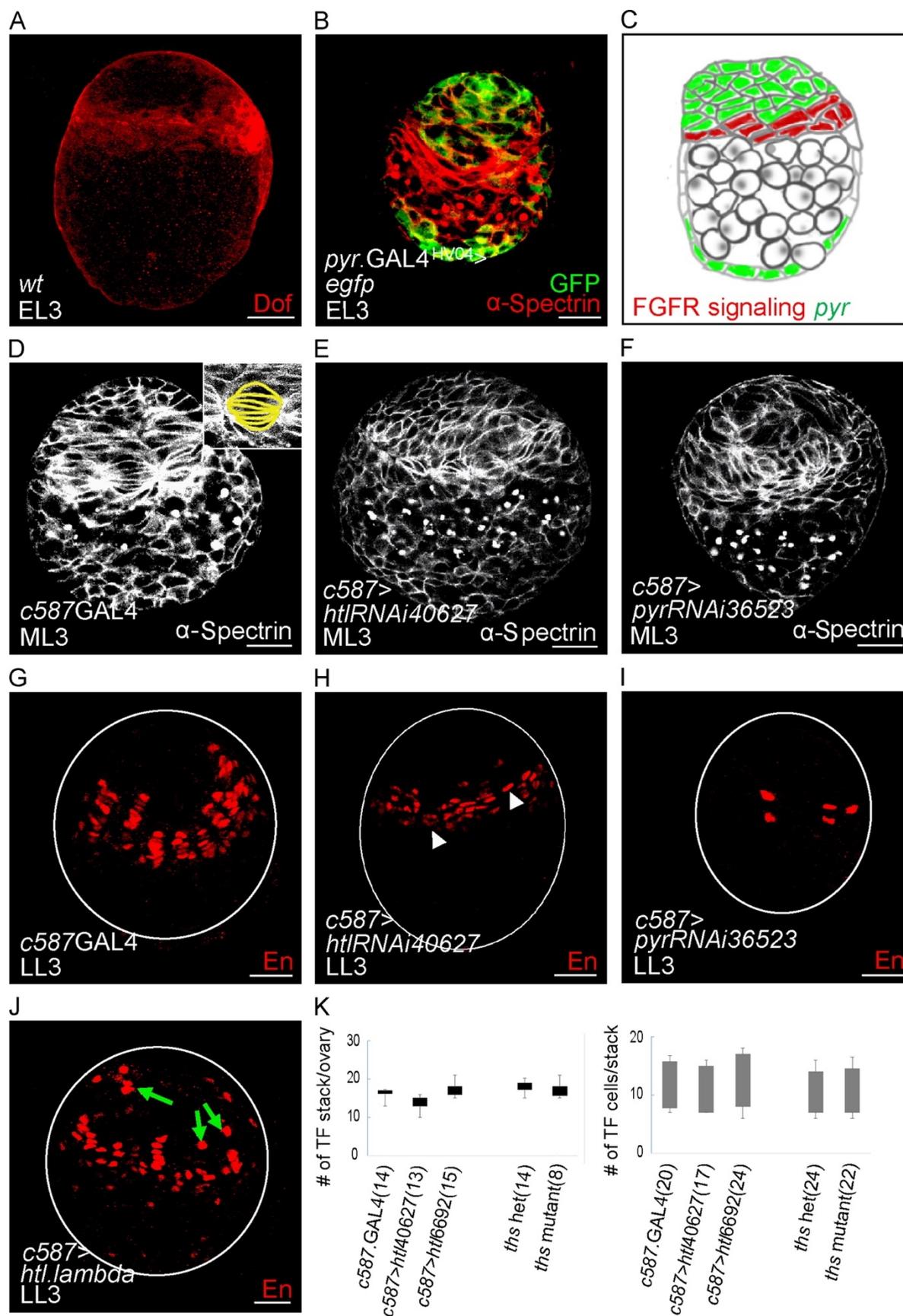
**Figure 1.** *ths* mutant females are viable but sterile, and exhibit ovarian muscle defects. (A–D) Images of adult ovaries isolated from the following genetic backgrounds: wildtype (A), *ths* mutant (B), *ths* mutant rescue (C) or *htl*-knockdown (D) using a light microscope. In this

and all other figures, “*ths* mutant” refers to the *ths*<sup>02026</sup>/Df(2R)*ths*238 transheterozygous combination of alleles, which is viable. (E) Graph of fertility assays comparing FGF mutants to controls. The number of deposited eggs was counted for mutants (i.e. two different *htl* KD RNAi constructs and *ths* mutant) and the values normalized relative to respective “wildtype” controls (i.e. *c587.GAL4* or *ths* heterozygote, respectively). Sample size (i.e. number of flies) indicated within bracket. In this and all other figures, “*ths* heterozygote” refer to *ths*<sup>02026</sup>/Cyo,actin-GFP or Df(2R)*ths*238/Cyo,actin-GFP. For statistical analysis, two-tailed Student's *t*-test was used, and  $p < 0.05$  was considered significant (\*\*). (F) Schematic showing two types of ovarian muscle tissues: peritoneal and epithelial sheaths. (G) In situ hybridization of wildtype adult ovarioles using an antisense riboprobe to detect *htl* transcripts. Arrows mark *htl* transcripts identified in the epithelial sheath. (H–K) Stainings to examine epithelial sheaths associated with ovarioles isolated from flies of the following genetic backgrounds: transgenic *htl-mcherry* line (H), UAS.GFP driven by *htl.GAL4*<sup>GMR93H07</sup>(I), wildtype (J), and *ths* mutant (K). Antibodies recognizing RFP (H, red), GFP (I, green), Vasa (J, K, green), and  $\alpha$ -Spectrin (I, J, K, red) were used. Arrow in (J) marks the epithelial sheath. (L–M) Views of wildtype (L) and *ths* mutant (M) ovaries stained with rhodamine phalloidin (green) used to detect the actin cytoskeleton. In panels (L) and (M), “A” denotes the apical region, whereas “P” denotes the posterior region. Boxed regions are shown within insets at magnification. In this and all other figures, scale bars denotes 20  $\mu$ m.

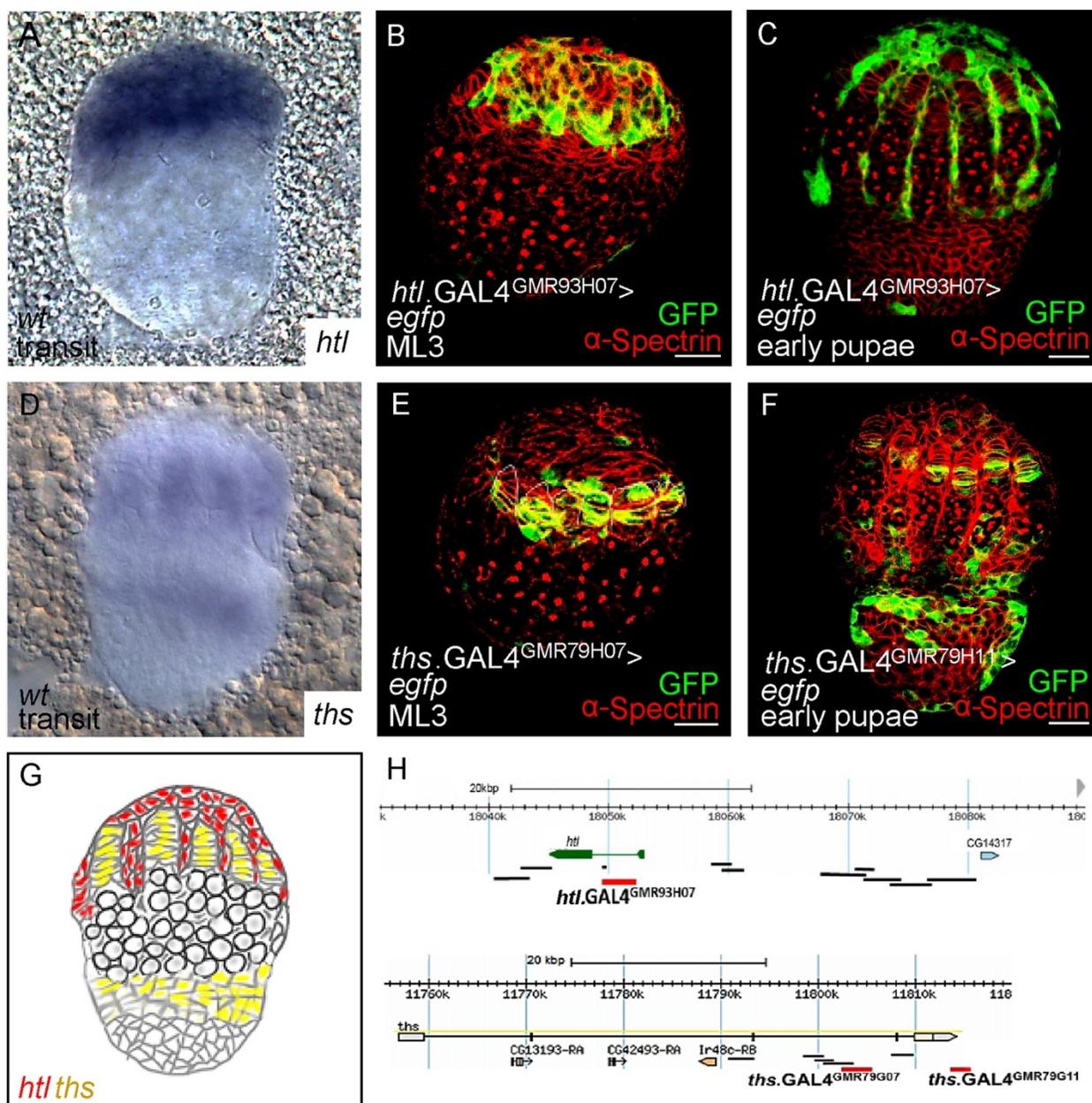


**Figure 2. FGF signaling acts in multiple somatic cell types within the developing gonad at the larval and pupal stages.** (A) Schematic showing ovarian morphogenesis at three stages: mid-larval third instar (ML3), early pupae, and middle pupae. In this and all other panels/figures, lateral views of the larval/pupal ovary are shown with anterior up and posterior down. (B–

D) Immunostainings of ovaries obtained from transgenic line *htl-mcherry* using anti-RFP antibody (white). Three different stages are shown: ML3 (B), early pupae (C), and middle pupae (D). Arrowhead in (B) marks migrating swarm cells at ML3. (E–G) Immunostainings of wildtype ovaries using anti-Dof (green) and anti- $\alpha$ -Spectrin (red) antibodies. Three different stages are shown: ML3 (E), early pupae (F), and middle pupae (G).

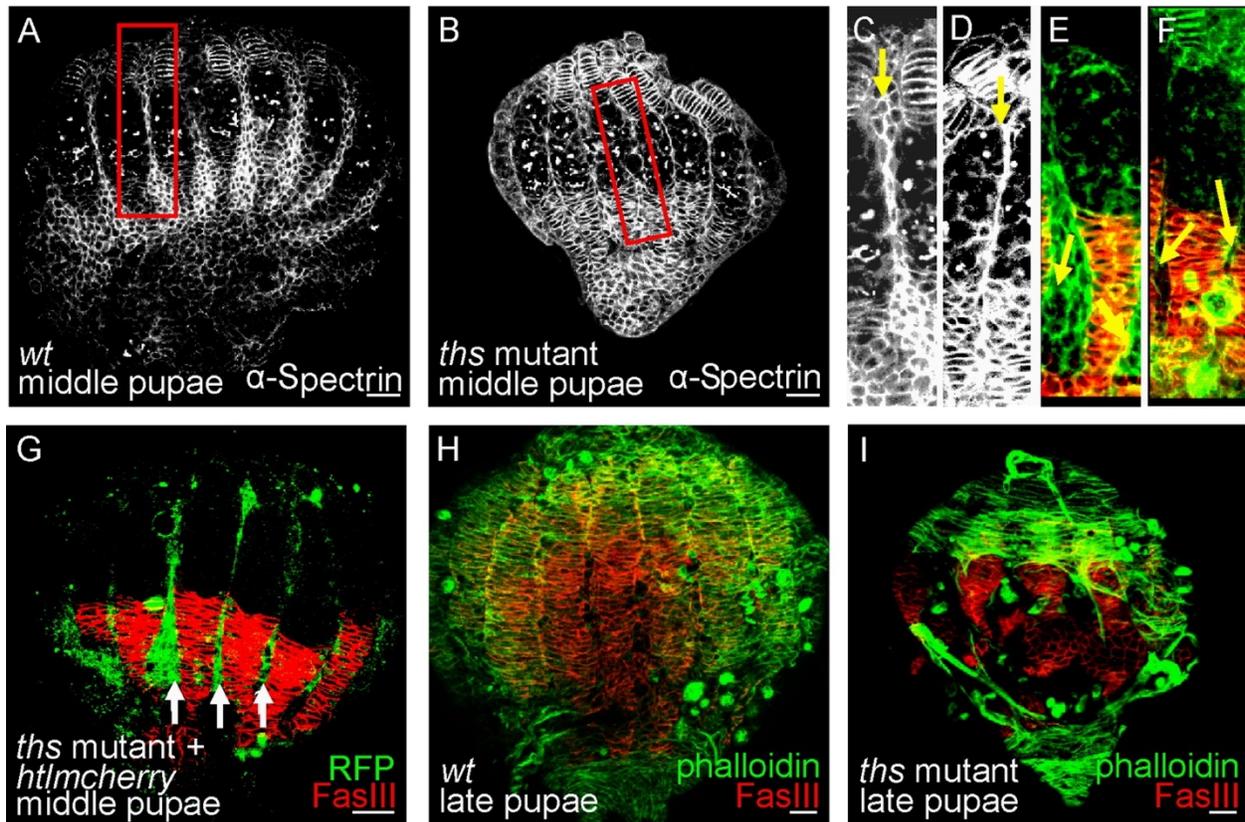


**Figure 3. FGF signaling activated by Htl FGF receptor supports TF cell differentiation at the larval stage.** (A) Immunostaining of wildtype ovary of early-larval third instar (EL3) stage using anti-Dof antibody (red). (B) Immunostaining of an ovary of EL3 stage using anti-GFP (green) and anti- $\alpha$ -Spectrin (red) antibodies to examine expression of UAS.GFP supported by *pyr.GAL4<sup>HA04</sup>*. (C) Schematic showing inferred activation domain of FGF signaling (red) relative to *pyr* expression domain (green) at EL3. (D–F) Immunostainings of ovaries at ML3 stage using anti- $\alpha$ -Spectrin antibody (white) to examine TF cell morphogenesis. A role for *htl* and *pyr* in somatic cells was examined using the *c587.GAL4* driver: control, driver alone (D) compared with *htl* KD (using UAS.*htl*.RNAi40627, E) and *pyr* KD (using UAS.*pyr*.RNAi36523, F). Within the magnified inset of (D), TF cells are outlined in yellow to demonstrate normal morphology. (G–J) Immunostainings of ovaries from late-larval third instar (LL3) stage using anti-En antibody (red) to detect differentiated TF cells. Effects on somatic cells were examined using the *c587.GAL4* driver: control, driver alone (G) compared with *htl* KD (using UAS.*htl*.RNAi40627, H), *pyr* KD (using UAS.*pyr*.RNAi36523, I), and constitutively-active-Htl (using UAS.*htl*.lambda, J). Representative single confocal sections are displayed; and ovary outlines are marked by white circles. In (H), TF stalks that contain fewer En<sup>+</sup> TF cells are marked by white arrowhead. In (J), ectopically expressed En<sup>+</sup> cells are marked by green arrows. (K) Graphs showing number of TF stalks per ovary (left) and TF cells per stack (right) in *c587.GAL4* control, two *htl* KDs (using UAS.*htl*.RNAi40627 and UAS.*htl*.RNAi6692), *ths* heterozygote control, and *ths* mutant.



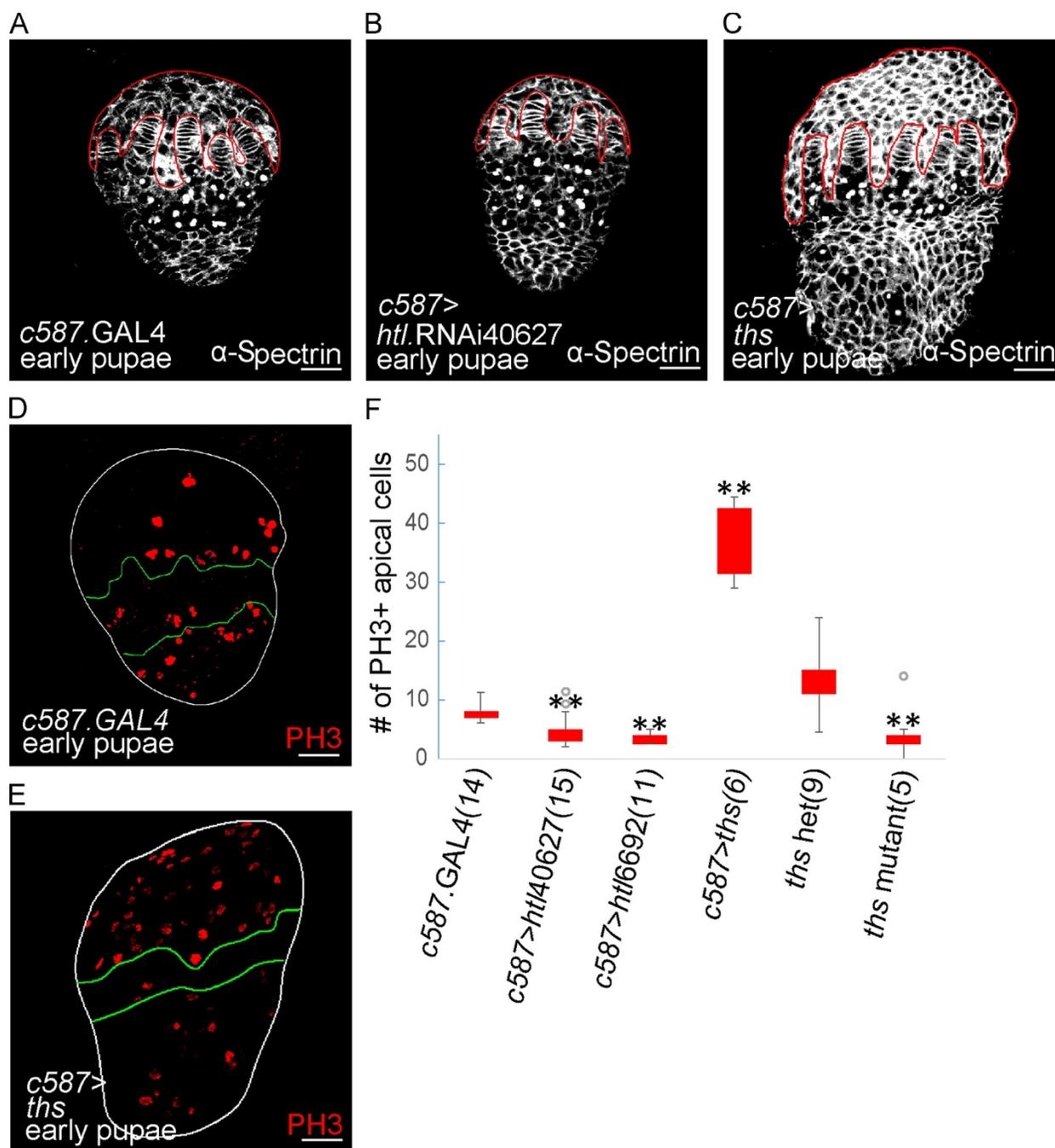
**Figure 4.** At the early pupal stage, *htl* is expressed in the ACs, whereas *ths* is expressed in the TF and presumptive basal stalk cells. (A, D) In situ hybridization using antisense riboprobes to detect *htl* (A) and *ths* (D) transcripts in wildtype ovaries at the stage of transition from larvae to pupae. (B, C, E, F) Immunostainings of ovaries expressing UAS.GFP driven by *htl.GAL4<sup>GMR93H07</sup>* (B, C), *ths.GAL4<sup>GMR79H07</sup>* (E), and *ths.GAL4<sup>GMR79H11</sup>* (F) detected using anti-GFP (green) and anti- $\alpha$ -Spectrin (red) antibodies. Two different stages are shown: ML3 (B, E) and

early pupal stage (C, F). (G) Schematic showing expression domains of *hhl* (red) and *ths*(yellow) at the transition stage. (H) Location of non-coding DNA regions used to make GAL4 drivers, which were assayed for expression in the developing ovary. Those that did or did not support expression are labeled red or black, respectively.



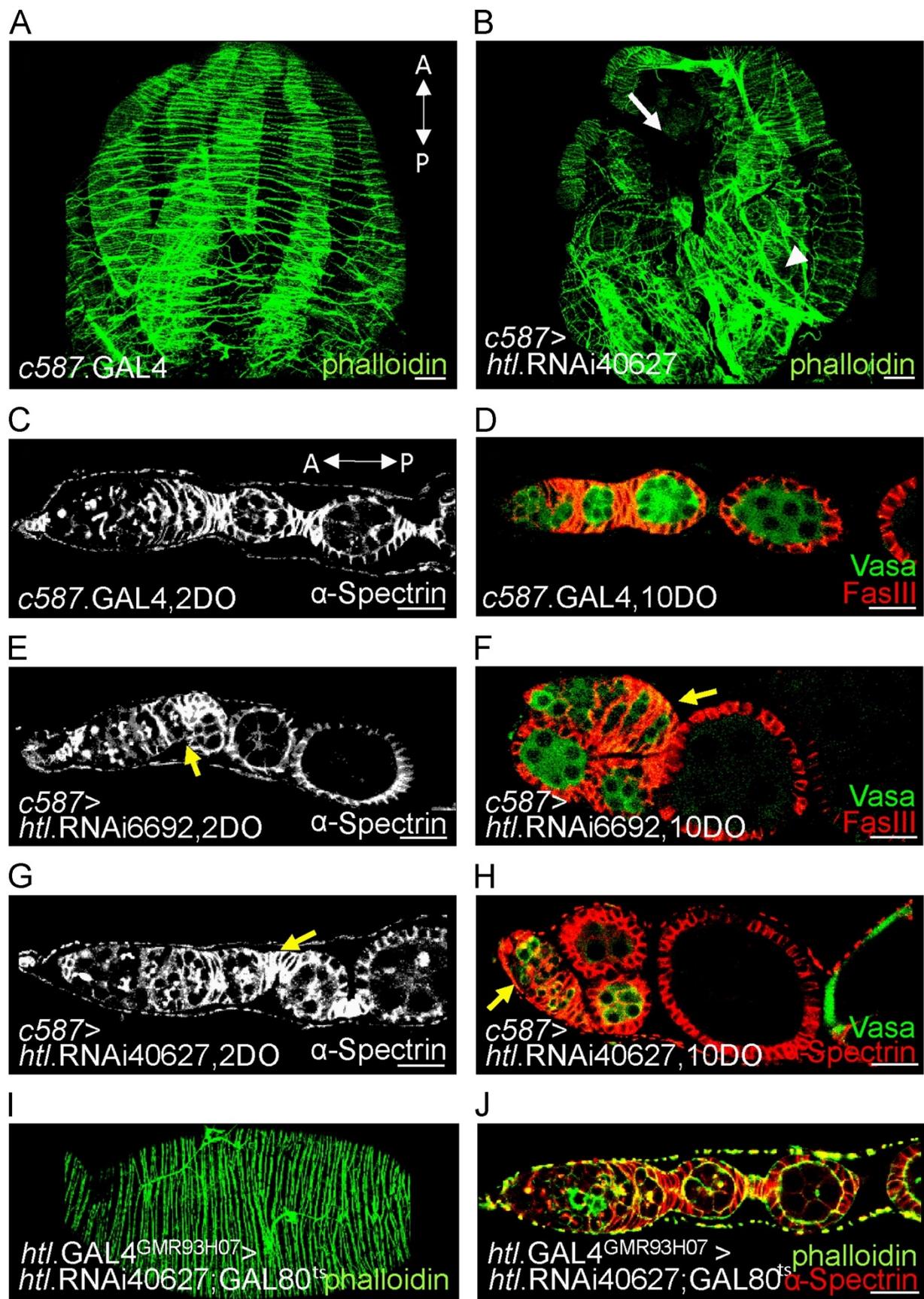
**Figure 5. FGF signaling is not necessary for AC migration during ovarian development.**

Stainings of ovaries using anti- $\alpha$ -Spectrin (A–D, white), anti-FasIII (E–I, red), and anti-RFP (G, green) antibodies or rhodamine phalloidin (E, F, H, I, green). (A–F) Ovaries at the middle pupal stage obtained from wildtype (A, C, E) and *ths* mutants (B, D, F). Migrating ACs located between presumptive ovarioles are marked with yellow arrows (C–F). (G) Htl.mCherry reporter introduced into *ths* mutant background stained with anti-RFP (green) and anti-FasIII (red) to detect ACs and basal stalks, respectively. (H, I) Staining of wildtype (H) or *ths* mutant (I) at the late pupal stage showing gaps in the muscle sheath.



**Figure 6. FGF signaling regulates the AC population size by controlling cells' proliferation at the early pupal stage.** (A–E) Immunostainings of ovaries at the early pupae stage using anti- $\alpha$ -Spectrin (A–C, white) and anti-Phospho-Histone H3 (PH3, D, E, red) antibodies. A role for *htl* and *ths* in somatic cells was examined using the *c587.GAL4* driver: control, driver alone (A,

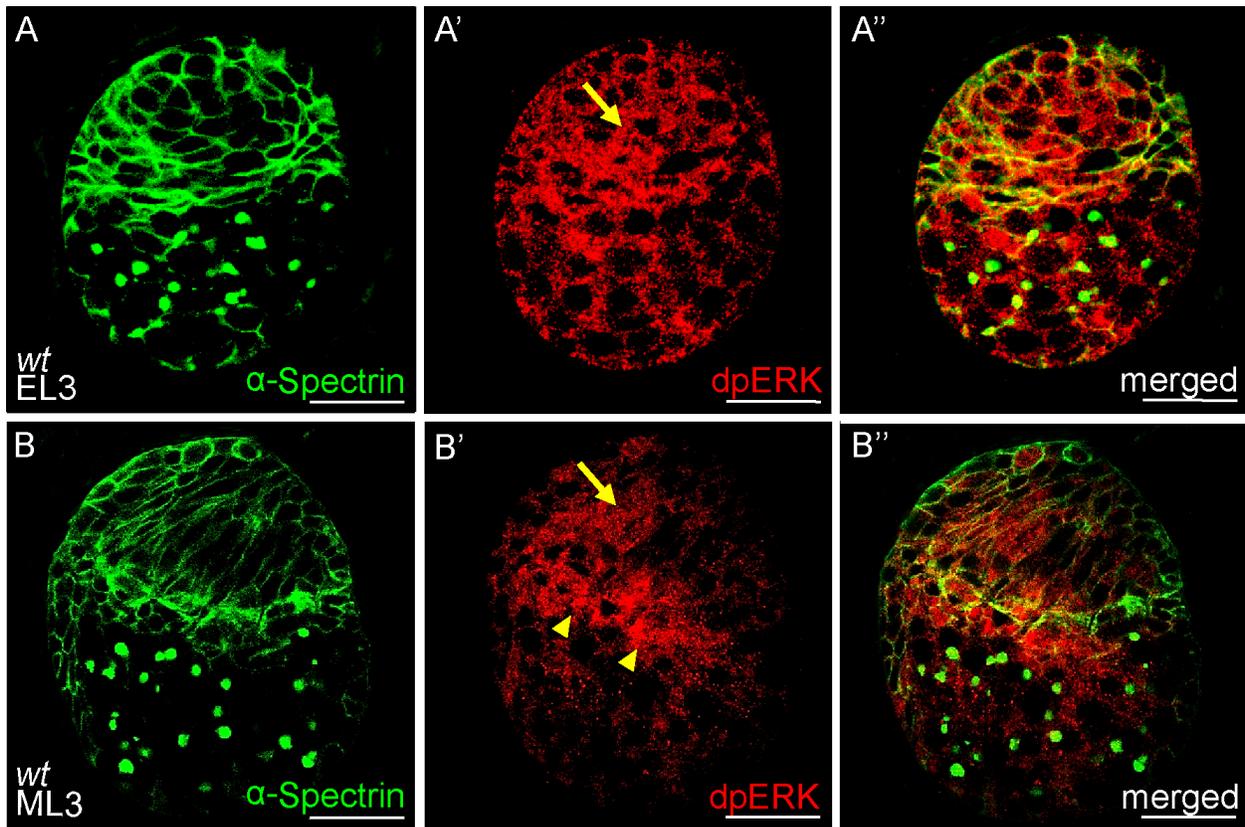
D) compared with *htl* KD (using UAS.*htl*.RNAi40627, B) and *ths*-ectopic expression (C, E). AC domain in (A–C) is outlined in red. In (D, E), ovaries are outlined in white, while the location of germ cell is bounded by green lines. (F) A graph showing number of PH3+ ACs in *c587*.GAL4 control, two *htl* KDs (using UAS.*htl*.RNAi40627 and UAS.*htl*.RNAi6692), *ths* heterozygote control, and *ths* mutant. The sample size is indicated within brackets. Gray circles represent outlier datapoints. For statistical analysis, two-tailed Student's *t*-test was used, and  $p < 0.05$  was considered significant (\*\*).



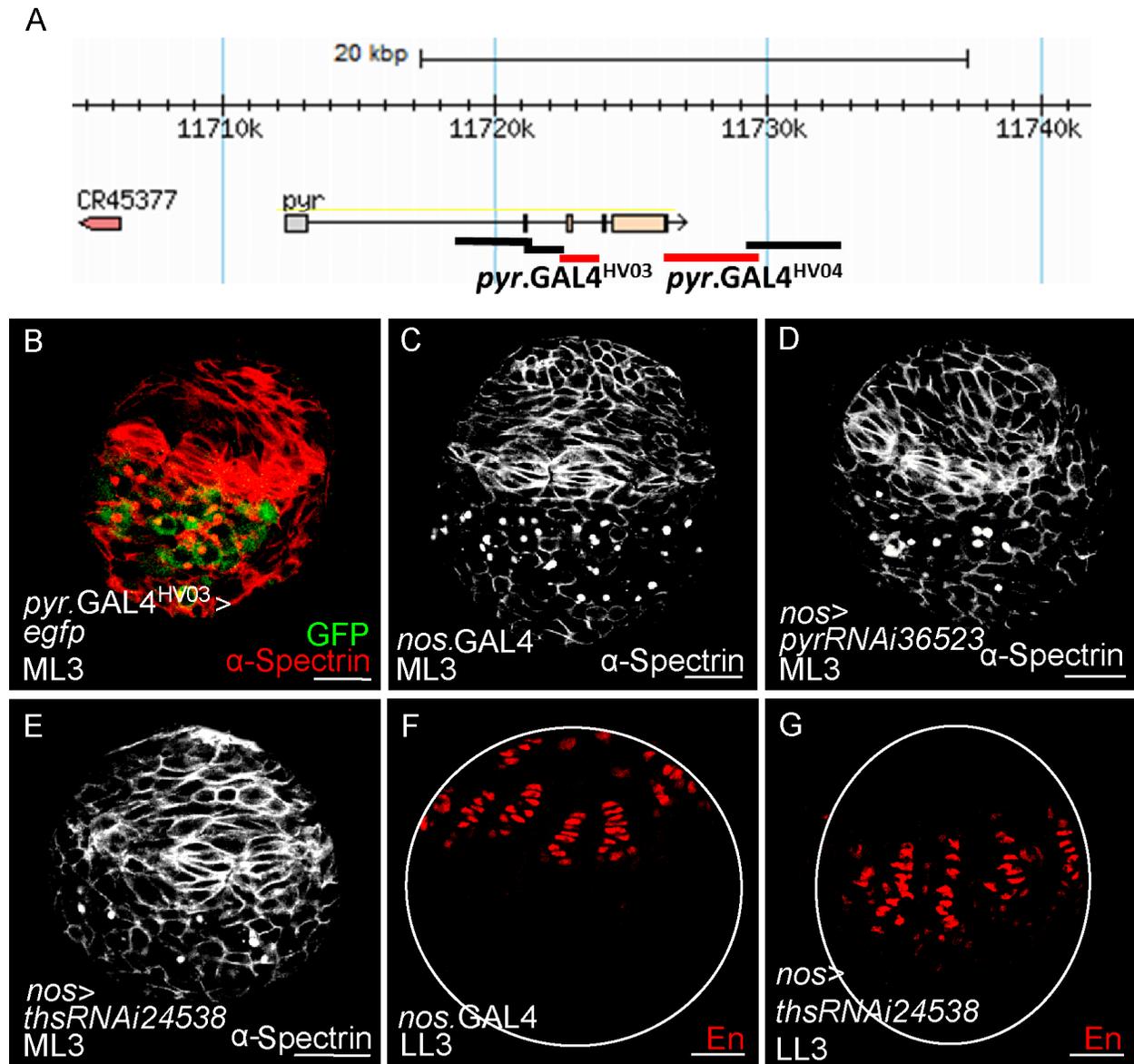
**Figure 7. Ovarian muscle sheaths provide structural support during oogenesis, which affects female fertility.** Stainings of adult ovaries by rhodamine phalloidin (A, B, I, J green) or anti- $\alpha$ -Spectrin (C, E, G, white; H, J red), anti-FasIII (D, F, red), and anti-Vasa (D, F, H, green) antibodies.

(A, B) The role of *htl* in morphogenesis of ovarian muscles was examined using *c587.GAL4* somatic cell driver: control, driver alone (A) compared with *htl* KD (using *UAS.htl.RNAi40627*, B). A white arrow and arrowhead in (B) mark absence and aberrant organization of ovarian muscle sheaths, respectively. (C–H) The role of *htl* in structural organization of adult ovarioles was investigated using the somatic cell *c587.GAL4* driver: control, driver alone (C, D) compared with *htl* KDs [using *UAS.htl.RNAi6692* (E, F) and using *UAS.htl.RNAi40627* (G, H)]. Yellow arrows in (E–H) mark “flop-down” egg chamber phenotypes. Phenotypes for two time points are shown: two days old (2OD; C, E, G) and ten days old (10OD: D, F, H). (I, J) Timing of *htl* action was examined by using  $GAL80^{ts}$  to limit *htl* KD to the adult stage. *htl.GAL4<sup>GMR93H07</sup>* driver together with  $GAL80^{ts}$  were used to support expression of *UAS.htl.RNAi40627* only in adults, by switch to growth at 29 °C. In panels (A) and (C), “A” denotes the anterior tip and “P” denotes the posterior tip.

## SUPPLEMENTARY FIGURES &amp; LEGENDS

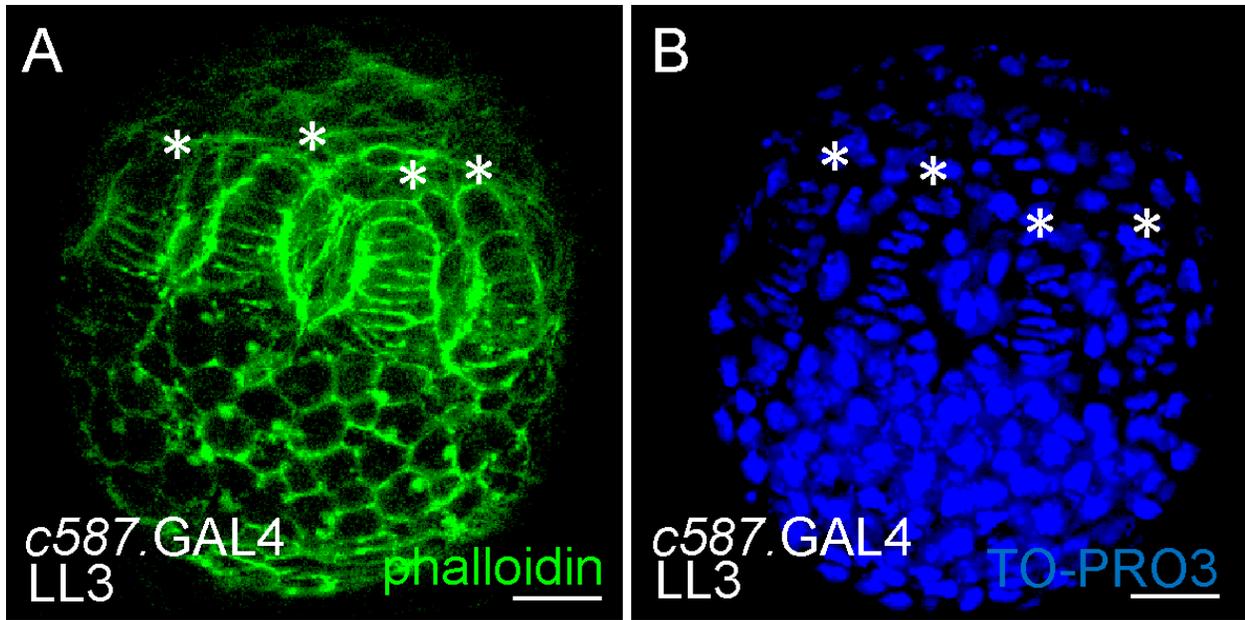


**Supplementary material Figure S1. MAP kinase is activated in somatic cells at the larval stage including the differentiating TF cells.** (A-B) Immunostainings of wildtype ovaries isolated from two different stages, EL3 (A-A'') and ML3 (B-B''), using anti- $\alpha$ -Spectrin (green) and anti-dpERK (red) antibodies to detect MAPK activation domains. Yellow arrows mark dpERK-positive ACs, whereas yellow arrowheads mark dpERK-positive differentiating TF cells.



**Supplementary material Figure S2. *pyr* or *ths* KD in the germline does not appear to affect TF cell development.** (A) Illustration of GAL4 drivers tested for the expression during ovary development. Segments of non-coding DNA from the vicinity of the *pyr* gene were assayed. Those supporting expression or not are denoted in red and black, respectively. (B-G) Immunostainings of ovaries isolated at ML3 (B-E) and LL3 (F,G) stages using anti- $\alpha$ -Spectrin (B, red; C-E, white), anti-Vasa (green) and anti-Engrailed (F,G, red) antibodies. Expression domain of *pyr* was tested by expressing *gfp* with *pyr.GAL4<sup>HV03</sup>* (B). A role for *pyr* or *ths* in germline cells was tested

using *nos.GAL4* driver: control, driver alone (C,F) compared with *pyr* KD (using UAS.*pyr*.RNAi36523, D), and *ths*KD (using UAS.*ths*.RNAi24538, E,G). Ovary outlines in (F) and (G) are marked by a white line. Two stages are shown: ML3 (B-E) and LL3 (F-G).



**Supplementary material Figure S3. TF cells can be identified by their distinctive cell and nuclear shapes.** (A,B) Staining of ovaries obtained from *c587.GAL4* at the LL3 stage using rhodamine phalloidin (green, A) and TO-PRO3 (blue, B), respectively, to identify TF cells, which are disc-shaped, and associated nuclei, which appear flattened. Asterisks indicate a TF stack.

**Supplemental Table 1. Primer sequences utilized in this study.**

	Forward primer	Reverse primer
<i>pyr.GAL4<sup>HV01</sup></i>	caccataaagagacccgctg	ggcatgaacttgtggaacct
<i>pyr.GAL4<sup>HV02</sup></i>	caccaagttcatgcccaatgt	gtatgttctcggcggtcac
<i>pyr.GAL4<sup>HV03</sup></i>	caccagcaaacacagcgaac	atccaggctgaacacatcgt
<i>pyr.GAL4<sup>HV04</sup></i>	cacctttcccatcatctttgg	acagttagcattccgtcagt
<i>pyr.GAL4<sup>HV05</sup></i>	cacctccgataattgcgagag	cggtccctctgattgactg
<i>htl-mcherry</i>	gaggaggaagatgaaacggacaacctg cagaagtgggtgtaa ttatggtggaggcgggtgggggatggt gagcaagggcgagga	accaggcttgaattgtcctgatc tcatactaaggatc ctgattcgttcgaagagctattc cagaagtagtga
<i>htl-la</i>	aggcgcgcccggctaaccaagtctctt gccatt	cgcggatcctggtgcaagttgcc tggcat
<i>htl-ra</i>	cgcggatccgaagctcgacagaatcc attcca	accttaattaagcaggttgctga tgtagttcggtt
<i>ths-la</i>	aggcgcgccatgtgatatgaccagtc gtcc	cgcggatccgctagttggccact cgaatttg
<i>ths-ra</i>	cgcggatccacgtcattatgccacgct gctcc	aaggaaaaagcgatcgccaatc tgaagtatgc

## Appendices

B.

A developmental program truncates long transcripts to temporally regulate cell signaling

In this study, Jihyun Irizarry established CRISPR fly stocks and performed antibody staining. It was published in *Developmental Cell* in 2018.

## SUMMARY

Rapid mitotic divisions and a fixed transcription rate limit the maximal length of transcripts in early *Drosophila* embryos. Previous studies suggested that transcription of long genes is initiated but aborted, as early nuclear divisions have short interphases. Here, we identify long genes that are expressed during short nuclear cycles as truncated transcripts. The RNA binding protein Sex-lethal physically associates with transcripts for these genes and is required to support early termination to specify shorter transcript isoforms in early embryos of both sexes. In addition, one truncated transcript for the gene *short-gastrulation* encodes a product in embryos that functionally relates to a previously characterized dominant-negative form, which maintains TGF- $\beta$  signaling in the off-state. In summary, our results reveal a developmental program of short transcripts functioning to help temporally regulate *Drosophila* embryonic development, keeping cell signaling at early stages to a minimum in order to support its proper initiation at cellularization.

## INTRODUCTION

Early embryonic development of the fruit fly *Drosophila melanogaster* is characterized by 14 rapid and syncytial mitotic nuclear cycles (NCs) as the fertilized egg divides into ~6000 nuclei before cell membranes form and gastrulation occurs (Foe and Alberts, 1983). These NCs occur within three hours of egg laying and vary in length from ~10 minutes to about an hour, gradually lengthening as the embryo nears gastrulation (Pritchard and Schubiger, 1996; Tadros and Lipshitz, 2009). This rapid pace of nuclear divisions leads to a dynamic transcriptional environment, where patterns and levels of gene expression change between and within NCs (Reeves et al., 2012; Sandler and Stathopoulos, 2016a). Transcription is aborted during mitosis between NCs, and nascent transcripts are degraded, with transcription restarting at interphase of the following NC (Shermoen and O'Farrell, 1991).

As the rate of transcription in *Drosophila* has been measured at ~1.1-1.5kb per minute of interphase (Ardehali and Lis, 2009; Garcia et al., 2013), transcription of zygotic genes during syncytial NCs is likely time constrained. In support of this view, early zygotic genes have an average length of 2.2kb, while the overall average length of coding genes in *Drosophila* is 6.1kb (Artieri and Fraser, 2014; Hoskins et al., 2011), suggesting a bias towards short genes during this time period. It was previously thought that long genes, those over 20kb, are either not transcribed before the longer and final syncytial NC14 or are aborted mid-transcript, and no protein products were present (O'Farrell, 1992; Rothe et al., 1992).

Activation of the zygotic genome and the maternal to zygotic transition (MZT) takes place during the syncytial nuclear period and cellularized blastoderm period before gastrulation, concurrent with time constraints on transcript length (Tadros and Lipshitz, 2009). This is also when the dorsal-ventral and anterior-posterior axes that pattern the embryo, and eventually the adult fly, are established by zygotically transcribed genes relying on a few key maternal signals. Lastly, components of virtually all signaling pathways are zygotically transcribed during this time (Lott et al., 2011; Sandler and Stathopoulos, 2016b) and these signaling pathways, such as TGF- $\beta$ , JAK/STAT, Notch, FGF, and EGFR, are active and essential during embryonic development (rev. in Stathopoulos and Levine, 2004). For all these reasons, it is essential that the necessary genes for these processes be transcribed at the correct time in development, yet the observations of the exclusion of long genes remain, along with the questions about consequences for development in the absence of these transcripts.

Recently, studies have produced evidence that some long genes are transcribed during early NCs (Ali-Murthy et al., 2013; Lott et al., 2011; Sandler and Stathopoulos, 2016b). To explore these observations that seemingly contradict previous research, we examined transcription of long genes during short syncytial NCs, specifically NC13, with an interphase of 15 minutes, and compared

the transcription of these same genes during the longer interphase associated with NC14, which is over 45 minutes.

## RESULTS AND DISCUSSION

### Long transcripts are truncated during short nuclear cycles

Using an available RNA-seq dataset of *Drosophila* early embryonic development, we selected four genes over 20kb with evidence of transcription during NC13: *short gastrulation (sog)*, *scabrous (sca)*, *Protein kinase cAMP-dependent catalytic subunit 3 (Pka-C3)*, and *Netrin-A (NetA)* (Figure 1A) (Lott et al., 2011). 5' and 3' rapid amplification of cDNA ends (RACE) was performed on RNA from embryos aged 1-3 hours, which includes NCs 13 and 14 (Figure 1B), to search for alternative transcript isoforms. Only the previously defined 5' transcription start sites were recovered (Graveley et al., 2011) suggesting that alternative start sites are not used for these genes, whereas 3' RACE products identified truncations in these four transcripts (Figure 1A). The short forms aligned to annotated transcripts at the beginning of the full-length genes, but ended with an alternative exon, including novel coding sequence and a 3' UTR in what is usually an intron (Figure 1A, red transcripts; and Figure S2). The RACE products were spliced and polyadenylated, with no poly-A in the genome at the locus of alignment, suggesting they are mature transcripts.

To distinguish between full-length transcripts and short forms, we designed fluorescence in situ hybridization (FISH) riboprobes to the 5' and 3' ends of *sog*, *sca*, *Pka-C3*, and *NetA* with 3' exonic probes downstream of mapped short RACE sequences and therefore recognizing full-length forms only (Figure 1A). In all cases, there was no observable nascent signal from the 3' exonic probes during NC13 while signal from the 5' exonic probes was present, indicating that transcription did not reach the 3' ends of genes assayed (Figure 1C,D,G,H). In contrast, full-length

transcripts were present in NC14 when interphase is longer, as indicated by equivalent levels of expression detected by both 5' and 3' probes (Figure 1E,F,I,J). The signals were quantified relative to ubiquitous histone staining and compared for NC13 and NC14, showing that at NC13 the signals associated with 5' versus 3' ends were significantly different while roughly equivalent at NC14 (Figure 1K-N).

*The RNA-binding protein Sex-lethal controls transcript truncation*

Since the short transcripts include intron-derived coding sequence (Fig. S2A-D, blue sequences), we reasoned it is likely that transcriptional regulation is a cause of truncation at NC13 as opposed to post-transcriptional cleavage of full-length, mature mRNAs, which after splicing would lack intron-derived coding sequence. The sequence within 1 kB downstream of the new exons was examined for all four transcripts found to be truncated. While there were binding sites for 20 temporally relevant RNA Binding Proteins (RBPs) in all four genes, we found that the sites for Sex-lethal (Sxl) (Figure 2A; Ray et al., 2013) were the only ones statistically enriched, with  $p < 0.001$  calculated using AME motif enrichment algorithm (see Methods; Figure S1) (McLeay and Bailey, 2010).

*Sxl* is a well characterized sex determination gene in *Drosophila* involved in splicing (Moschall et al., 2017; Salz and Erickson, 2010). Zygotic expression of functional Sxl protein only occurs in female embryos, while males express a non-functional form (Bell et al., 1991; Bopp et al., 1991). However, short transcripts of long genes (e.g. *sog*) were observed in all embryos examined at NC13 (not only females) demonstrating that the RBP fulfilling this role is not sex-specific. Notably, *Sxl* is also maternally expressed with transcripts deposited into eggs/early embryos; while based on activation of the *Sxl* associated Pe zygotic promoter and in situ hybridization using riboprobes designed to the 5' end of the gene (i.e. Ex1), female-specific,

zygotic transcription is thought to occur at NC11 (Cline, 1993; Keyes et al., 1992). It remains unclear, however, whether full-length transcripts of the long (>23kb) *Sxl* gene can be completely transcribed. Moreover, when we examined RNA-seq data from a fine time course of early *Drosophila* development, we found support for the view that zygotic *Sxl* transcripts are not upregulated in females until mid NC14 (Lott et al., 2011).

Since we observed short transcript production in embryos of both sexes, we investigated whether at NC13 maternal *Sxl* could support this role. Maternal and female-specific zygotic mRNA transcripts should support the production of proteins with shared sequences and thus be recognized by the same antibody. However, although we were able to detect female-specific *Sxl* protein by Western blot at NC14 and show specificity for the antibody via maternal RNAi knockdown that was also able to downregulate zygotic levels (Figure 2H and S3M,N), we were unable to unambiguously visualize *Sxl* in unfertilized and early stage embryos, as bands of similar (but not identical) size to female-specific *Sxl* identified by Western blot in these embryos and early stages were unaffected by the equivalent RNAi conditions, suggesting these are background bands that possibly mask true maternal *Sxl* (Figure S3M,N). As assays of maternal *Sxl* by Western proved inconclusive, immunostaining to examine the protein in individual embryos at NC13 did reveal *Sxl* present in both sexes, using an intronic probe to *sog* (on the X chromosome) to determine the sex of the embryos (Figures 2B-F). In both male and female embryos, we observe presence of *Sxl* protein at NC13 (Figures 2B',C' and S3O,P) and in earlier NCs as well (e.g. Figure S3K,L). *Sxl* levels are reduced by heat-shock induced, maternal RNAi, initiated during oogenesis but which perdures into the early embryo (Figures 2J-M; see Methods). The immunostaining of individual embryos is sensitive enough to detect low levels of *Sxl* and to identify that a 43% reduction occurs upon *Sxl* RNAi (Figure 2G). Furthermore, this fine time resolution analysis of *Sxl* protein levels demonstrates that female-specific, zygotic *Sxl* protein is not produced until late

NC14 and suggests that earlier signal detected by immunostaining relates to a low-abundance maternal isoform (Figures 2D',E' and S3Q-T). To support this, maternal Sxl at NC13 in both sexes (Figures 2B',C') is much lower level than zygotic Sxl in females at late stage NC14 (Figure 2E',S3O,P,T) but comparable to (or even higher than) levels retained at this stage in males (Figure 2D',S3P,Q,S) or in female *daughterless* (*da*) mutant embryos (Figure 2F'), where all *Sxl* zygotic transcription is eliminated (Cronmiller and Cline, 1987).

To characterize effects of maternal *Sxl* RNAi on transcript truncation, which showed sufficient knock-down of Sxl (Figures 2G, J-M, and S3N), riboprobes targeting transcription of intronic sequences 3' of the initially defined truncation sites of the four long genes, just 3' of the cluster of predicted Sxl binding sites (e.g. Intron 3 probe, Figure 2I), were used to test the hypothesis that Sxl binding to the transcript could act to influence termination. In embryos subject to maternal RNAi against *Sxl*, intronic FISH signal past the truncation point was observed during NC13 for all four long genes assayed, indicating that transcriptional read-through past the truncation point occurs in both sexes (Figures 2N,O and S3A,C,E, compare with Figure 2P and S3B,D,F,G,J). Furthermore, *da* mutants expressing only maternal Sxl were not able to support transcriptional read-through past the truncation point (Fig. 2Q). Collectively, these results support the view that maternal, not zygotic, Sxl is responsible for transcriptional truncation in early-stage embryos of both sexes. Since Sxl's role in supporting sex determination is not conserved outside of the *Drosophila* genus (Cline et al., 2010), it is possible that the role we have defined here resembles an ancestral one that evolved to balance fast development with proper activation of cell signaling.

#### *Sex-lethal is associated with truncated transcripts*

If truncation of long RNAs is mediated through direct binding of Sxl, then the clusters of

Sxl consensus binding sites (e.g. orange arrowheads, Figures 2I) must be transcribed for Sxl to bind and act. Using qPCR primer sets spaced along the *sog* locus (Figure S4A, blue markers) and individually staged, but not sex-selected, embryos, we found that during NC13, the *sog* intronic sequence including the Sxl binding site cluster was present, but abundance of *sog* transcript was drastically reduced in the downstream intronic sequence and the 3' coding exon (Fig. S4B,D), indicating that the truncated form, but not full-length transcript is transcribed. During NC14, the entire intron including the Sxl binding site cluster was spliced out, but the 3' coding exon was retained at high levels equivalent to the 5' exon (Figure S4C,E). At NC13 in embryos subject to maternal Sxl RNAi, more of the intron was retained, but the full transcript was still not present (Figure S4B). These results reinforce the idea that Sxl is needed for truncation of the *sog* transcript, and when Sxl is removed, truncation fails and the intron is retained.

To determine if Sxl physically associates with transcripts that exhibit truncation, we immunoprecipitated Sxl protein from a bulk collection of 2-4h embryos and performed qPCR on eluted RNA. We found that mRNAs of the genes *Sxl*, *msl-2*, and *tra*, which are known to be bound by Sxl for alternative splicing (Moschall et al., 2017), were enriched in the Sxl IP compared to a mock IP/negative control using an antibody to Ubx, a nuclear DNA binding protein without RNA binding function as were transcripts of *sog*, *NetA*, *sca*, and *Pka-C3* (Figure 2S). Surprisingly, there was no statistical difference between the enrichment of the canonical Sxl sex-determining targets and the short transcripts investigated here. On the other hand, the genes *twi* and *sna* (short genes under 5kb) and *sog* In3B (qPCR primer 3' of the cluster of Sxl binding sites; Figure S4A), were not significantly enriched (Figure 2S), indicating little to no Sxl binding to mRNA of short genes or past the truncation point of long genes, though the intronic nature of *sog* probe In3B could lower its measured enrichment due to splicing out. These results, in combination with the presence of Sxl binding sites in the transcripts for the short forms of *sog*, *NetA*, *sca*, and *Pka-C3* genes (e.g.

Figure S1A-D) strongly indicate that Sxl binds to all four mRNAs found to be truncated. As short *sog* is also produced at NC14 (Figure S4F,G), the binding detected is likely a mix of Sxl protein binding to short *sog* transcript at NC13 and NC14. It is also possible that Sxl protein also associates with full-length transcripts once they are produced.

Using CRISPR/Cas9, we deleted a ~1kb region of the *sog* intron containing the Sxl binding site cluster, which we term *sog*  $\Delta$ Sxl (Figure 3X). When we immunoprecipitated Sxl from embryos with this deletion and performed qPCR on associated mRNA, the association of Sxl with *sog* was greatly reduced compared to wild type, approaching the levels of negative control genes (Figure 2S). The association of Sxl with *sog* transcripts was not completely eliminated however, suggesting that while the 1kb Sxl cluster supports a significant amount of binding, other sites in the *sog* locus are likely still bound by Sxl (e.g. Fig. S1A). The association of Sxl with other mRNAs tested did not significantly change in the embryos lacking the binding site cluster in *sog*, indicating a specific interaction between Sxl and the binding sites in the *sog* intron (Figure 2S).

We also performed FISH on *sog*  $\Delta$ Sxl embryos using an intronic probe downstream of the deletion. In these embryos, transcriptional read-through past the truncation point was observed at NC13 with *sog* Intron 3 signal detection (Figure 2R), which does not occur in wild type embryos or in other controls (Figures 1C, 2Q, S3B,D,F,G,I,J), providing additional evidence that Sxl plays a key role in truncation. Furthermore, when the Sxl binding sites were mutated at the endogenous genomic locus using CRISPR-Cas9 (maintaining the spacing of the gene and FISH probes), transcriptional read-through past the truncation point into the intron was observed (Figure S3 H compare with I) suggesting Sxl directly controls transcriptional termination, noting this result exhibited partial penetrance (~70% of embryos, n=5 of 7) and the extension of the transcript was observable late in NC13.

Protein products of short transcripts are functional in signaling pathways

We investigated whether short products code for functional peptides in signaling pathways. Of particular interest, the short form of Sog contains the entire first cysteine-rich domain, which binds and sequesters TGF- $\beta$  ligands Decapentaplegic (Dpp) and Screw (Scw) (Figures 3A and B; Marqués et al., 1997). The short form predicted from the 3' RACE sequence closely resembles a Sog fragment known as Supersog, both in structure and function, which was hypothesized to arise from proteolytic cleavage of full-length Sog (Yu et al., 2000). However, the 3' RACE sequence recovered for *sog* includes the use of intronic sequence as coding RNA, which would be absent from full-length Sog after splicing. Full-length Sog is cleaved by the protease Tolloid (Tld) to release ligands for signaling, but short Sog protein predicted by 3' RACE does not contain Tld cleavage sites (Peluso et al., 2011) and may bind TGF- $\beta$  ligands Dpp and Scw irreversibly (Figure 3B).

To test the idea that short Sog inhibits Dpp-Scw action, we assayed the effect of ectopic expression of short Sog on the TGF- $\beta$  target genes *race*, *hnt*, and *ush*, expressed as stripes in the dorsal ectoderm at NC14, and commonly used to assay TGF- $\beta$  activity (Figures 5D,E and S5A,B; Ashe et al., 2000; Rusch and Levine, 1997). We placed the short *sog* cDNA under control of the *even-skipped* (*eve*) stripe 2 enhancer as previously done for full-length *sog* (Ashe and Levine, 1999), producing a stripe of expression along the anterior-posterior axis in addition to endogenous expression in a broad lateral domain (Figure 3C and S5F,L). In these embryos, expression of *race* is lost within the trunk and retained only in a small patch at the anterior end of the dorsal ectoderm (Figures 3H,I, compare with D,E), similar to embryos lacking functional Sog, since only the trunk expression, but not anterior domain, is Sog-dependent (Ashe and Levine, 1999; Xu et al., 2005). The expression pattern of *hnt* is also much weaker in these embryos, with the onset of expression slightly delayed, a gap in the stripe near the posterior, and a posterior retraction from the middle

of the embryo (Figure S5Q and V, compare to M and R). The expression of *ush* is weaker and slightly retracted at the anterior (Figure S5F and L, compare to A and G). These results indicate that the product of the truncated *sog* transcript, likely a short Sog peptide, acts as a dominant negative repressor of TGF- $\beta$  signaling.

We also expressed *eve stripe 2-short sog* in a *gastrulation-defective (gd)* background, which lacks endogenous *sog* expression due to defective Toll signaling and has expanded Dpp expression throughout the embryo (Konrad et al., 1998). Concomitantly, the TGF- $\beta$  pathway is activated along the entire DV axis, and *race* is ubiquitously expressed in the anterior two-thirds of the embryo (Figure 3K). In these embryos, and as shown previously (Ashe and Levine, 1999), when full-length cleavable Sog is expressed in the *eve stripe 2* domain, robust *race* expression is observed in the anterior and mid-trunk regions, but excluded from cells expressing Sog, as it represses locally and is cleaved at a distance to activate signaling (Figure 3J). In the case of *eve stripe 2-short sog* in *gd* embryos, *race* expression is limited to a band at the anterior pole of the embryo but is absent from the trunk (Figures 3 F,G compare with J,K). This result shows that short Sog does repress *race* but likely does not eliminate all signaling as expression in the head, supported by lower levels of signaling, is retained. Tld cleavage of full-length Sog is concomitant with release of ligands at a distance from the source of Sog expression and is the source of *race* expression in the trunk stripe (Ashe and Levine, 1999). In contrast, the local inhibition and lack of *race* activation at a distance in *eve stripe 2-short sog* embryos (Figure 3H,I,F) suggests that the predicted short Sog product cannot be cleaved by Tld to support activation of signaling and that binding of short Sog to Dpp and Scw is irreversible.

A further examination of signal transduction in the TGF- $\beta$  pathway provides more evidence that short Sog sequesters the ligands and modulates signaling. The signal transducer and transcription factor Mothers against dpp (Mad) is responsible for activating transcription of TGF-

$\beta$  targets, and phosphorylated Mad (pMad) indicates active TGF- $\beta$  signaling (Raftery and Sutherland, 1999). In wild type embryos, pMad is found in a narrow, robust, stripe along the entire dorsal ectoderm (Figures 3L,M), but in *eve stripe 2-short sog* embryos, pMad is diminished, ranging from decreased levels overall to mostly absent except for small patches at the anterior and posterior poles of the embryo (Figures 3N,O S5F,L). This change indicates that short Sog prevents Mad from being phosphorylated, shutting down TGF- $\beta$  signaling as the retraction of the gene *race* closely matches the gap in pMad, and these changes in *race* match those observed in flies with decreased pMad (Deignan et al., 2016).

To expand our study of short Sog, we used two new and one existing mutant lines which either remove or preferentially express the short form of *sog*. Specific regions of the *sog* locus were deleted using CRISPR with the intention of disrupting or decreasing short Sog (see Methods). One deletion removed the novel short Sog 3' UTR sequence in the *sog* intron possibly decreasing protein levels or mRNA stability (*sog*  $\Delta$  *New 3' UTR*), and a second deletion removed the ~1kb Sxl binding site cluster in the *sog* intron possibly leading to lack of product or longer mRNA due to defects in Sxl-mediated truncation (*sog*  $\Delta$  *Sxl*; Figure 3X). In both mutants, we observed precocious and sporadic activation of *race* throughout the embryo not present in wild type embryos of the same stage (Figure 3Q and R, compare to P). *hnt* expression in the trunk is observed earlier than in wild type (Figure S5M-O). The changes to *ush* patterns include weak early expression in the normal domain with some spots of ectopic expression that co-localize with ectopic *race* (Figure S5B,C compare with A; data not shown). This ectopic expression of *race* and *ush* and early activation of *hnt* suggest that short Sog is a dominant negative version of the protein that is important to keep cell signaling in check before cellularization, when TGF- $\beta$  ligands are widely expressed throughout the embryo. Our data indicate that when levels of short Sog are altered, possibly reduced, early sequestration of the ligands fails, and TGF- $\beta$  signaling is activated

ectopically in the mutant.

Changes in pMad observed in the *sog*  $\Delta$ New 3' UTR and *sog*  $\Delta$ Sxl lines help explain changes in TGF- $\beta$  target genes, which are dependent on pMad for their expression. In wild type embryos, pMad is localized in a narrow band of cells in the dorsal ectoderm (Figure 3M and S5A,G), but in the two CRISPR lines, pMad is weaker in dorsal regions (Figure S5B,C). This weaker expression is likely due to a lack of Dpp concentrated at the dorsal ectoderm, spread wider throughout the embryo instead, and responsible for the precocious, ectopic *race* and *ush* expression observed in these mutant embryos (Figures 3Q,R and S5B,C). In both of the CRISPR manipulated lines, full-length *sog* is eventually transcribed later in NC14, and its activity presumably restores *race*, *ush*, and *hnt* to their usual expression domains in late NC14 (Figures 3U,V and S5H,I,S,T).

We also identified a mutant in which the *sog* locus is interrupted by a P-element insertion ~3.5kb downstream of the Sxl truncation point (Figure 3X), which causes a ~7-fold decrease in transcription of full-length Sog but allows transcription of short Sog (Figure S4H). In this genetic background, short Sog is likely intact and functional at NC13 but a deficit in long Sog occurs at NC14. In embryos with this insertion, at NC14B, when full *sog* is normally first transcribed, *race* expression is retracted to a somewhat wider anterior patch compared to wild type embryos (Figure S5D), and later in NC14, *race* is not expressed at full strength in the trunk region (Figure 3W, compare with 3T). At NC14B, *ush* is weak and slightly expanded laterally, and *hnt* expression is difficult to detect (Figure S5D,P). These results suggest that when full *sog* is available in wild type embryos to establish TGF- $\beta$  signaling, the P-element line, which reduces full *sog* but allows dominant negative short *sog*, shows overall weaker expression from target genes. The phenotypes associated with the P-element at NC14B somewhat resemble those of embryos of *sog*<sup>Y506</sup> background (Ray et al., 1991), an RNA null mutant (e.g. Figure 3Z, S5E); with *race* and *ush* in *sog*<sup>Y506</sup> weaker, somewhat retracted along the AP axis, and expanded laterally. Ectopic expression

of *race* at early NC14 is observed in *sog*<sup>Y506</sup> mutant embryos (Fig. 3Y) but is not present in the *sog* P-element mutant embryos (Figure 3S). These data support the view that short Sog keeps signaling off at early stages, as short *sog* is present in the P-element line, with little to no early ectopic expression, but absent in *sog*<sup>Y506</sup>, where ectopic expression is observed. The similarities between the P-element and *sog*<sup>Y506</sup> diverge at late NC14, when full *sog* is available in the P-element line but not in *sog*<sup>Y506</sup>, and TGF- $\beta$  targets appear as normal (Figs S5 J,K,U).

The truncations we found were not limited to *sog*, and when the short peptides predicted by *NetA*, *sca*, and *Pka-C3* short transcripts (Figures 1A and 4A,D) were compared with full-length forms, a subset of functional domains were encoded, suggesting the short forms of these genes could correspond to functional truncated proteins. By qPCR, we determined that these transcripts are truncated at NC13 with novel coding sequence retained, but fully transcribed with novel coding sequence spliced out at NC14 (Figures 4B,C). Hydrophobicity plots of the short forms demonstrate that the novel amino acids likely maintain the structure and function of the short proteins (Figure S2A-D). Previous research involving either random or targeted mutagenesis of these genes, or mammalian orthologs, has uncovered evidence of dominant negative activity in all cases at later stages of development (Hu et al., 1995; Miloudi et al., 2016; Schneiders et al., 2007).

To provide insight into the role of these other short products in embryos, we expressed short *sca* in the *eve* stripe 2 domain and looked for phenotypes in early embryos. Specifically, as Sca has been shown to form a complex with Notch and modulate its activity (Powell et al., 2001), we assayed effects on one Notch target gene *single-minded* (*sim*), expressed in a thin stripe on the border of the mesoderm and neurogenic ectoderm (Figures 4E,F; rev. in Reeves and Stathopoulos, 2009). In embryos expressing *eve* stripe 2-*short sca*, *sim* is expressed early and expanded late only in the *eve* stripe 2 region, which is consistent with membrane-bound Sca protein affecting Notch locally (Figures 4G,H). In a previous study, the *sca* locus was subject to random mutagenesis, and

one allele was found to have a dominant negative phenotype that affected Notch signaling (Hu et al., 1995). This allele is a truncation of the *sca* transcript just after the Rab binding domain and resembles the *short sca* truncation we recovered using 3' RACE. It is possible that changes to *sca* shift the balance of Notch in the membrane vs. endosomes, which is mediated by Rab proteins (Hu et al., 1995).

Collectively, these data demonstrate that the long genes we observed and manipulated are truly truncated, and the full-length forms are not transcribed during NC13. Still, a recent publication has described a faster rate for RNA Pol II in *Drosophila* embryos of ~2.4 kb/min, using an analysis of heterologous engineered reporter genes of ~5kb in length (Fukaya et al., 2017). In this situation, transcription and subsequent translation of genes longer than 35kB during NC13 within 15 min would be challenging, while expression of genes less than 15kb would be achievable. Our qPCR quantification suggests long forms, if present from nascent transcription or maternal contribution, are present at ~600-fold lower levels than the short forms at NC13 (Fig. S4A,B,D Ex1:Ex5). Furthermore, we detect short transcripts present at NC14 (Fig. S4F,G) when full-length transcripts are also present suggesting that the balance of short and long forms is important for proper regulation of cell signaling.

#### Global 3' RNA-seq identifies additional truncated transcripts

To provide insight into the global or programmatic nature of transcript truncation, RNA-seq was performed on *Drosophila* embryos from NC13 and NC14 separately, targeting the 100bp at the 3' end of transcripts (i.e. 3' RNA-seq; Lianoglou et al., 2013). While there is little difference in 3' transcript ends of short genes between NCs 13 and 14 (Figure 5D), long genes show large differences in 3' transcript abundance (Figure 5A,B). We analyzed the dataset looking for additional short forms in NC13 examining long genes, greater than 15kb, as well as a shorter set

of genes 8-15kb that are longer than average but theoretically could be transcribed within the time window available at NC13. In addition, we narrowed the search to include only genes with mapped reads in both NC13 and NC14. Using these criteria, we manually annotated 450 genes greater than 15kb, and 354 genes 8-15kb, searching for additional short forms (Table S1). Among the 450 long genes, we found 27 putative short forms, such as the gene *grh* (Figure 5C), in addition to the four found by the original 3' RACE experiments, for a total of 31 truncated genes enriched for Gene Ontology (GO) terms Developmental Protein and Differentiation Gene (Table S2; see Methods). These two enriched GO functions point to a short transcript program specifically involving key developmental genes functioning in signaling and transcription in the early embryo.

In addition, many of these genes have clusters of Sxl binding sites within 1kb of their truncation points (Figure 5D). We did not find any clearly truncated genes in the 8-15kb group. This 3' RNA-seq experiment identifies global differences in truncated transcripts for both short and long genes. Moreover, our previous study using NanoString to quantify transcripts in the early embryo (Sandler and Stathopoulos, 2016b), including *sog* and *NetA*, also showed a difference in 5' vs 3' transcript abundance before NC14, confirming the results from 3' RNA-seq (Figure S4I-K).

The 3' RNA seq data also provided information on previously annotated 3' UTR usage, as a large number of genes had different 3' UTR usage between maternal and zygotic isoforms (Figure 5E). 125 of 450 long genes (i.e. >15kB) and 50 of 354 8-15kb genes had 3' peaks that were different between NC13 and NC14 (Figure 5F). All of these genes are both maternal and zygotic and using previously generated RNAseq data from staged embryos (Lott et. al., 2011), suggesting that the different 3' UTR peaks we observed corresponded with the switch from maternal to zygotic transcript in the early embryo. Moreover, the switch to zygotic 3' UTR usage, especially for long genes, occurs at NC14, when the time permissive length of the NC allows the

full transcription of the zygotic form. Although likely unrelated to Sxl-mediated truncation, this observation emphasizes both the time constraints early in development and the rapid switch in transcriptional program between NC13 and NC14 during the maternal to zygotic transition.

In closing, the need to temporally regulate the quiescence and rapid initiation of signaling pathways in the embryo is critical for proper development (Ashe et al., 2000; Noordermeer et al., 1992; Queenan et al., 1997). Rapid nuclear divisions limit transcript length of key signaling pathway members (Rothe et al., 1992), but we have shown that the truncation of these long transcripts to produce short products is a mechanism used to resolve this temporal challenge to ensure the proper timing for activation and/or maintenance of signaling. In a sense, the truncation of long transcripts can be thought of as a “rescue” whereby long transcripts that would usually be degraded and lost during rapid mitotic cycles are made mature and stable by truncation, and survive to produce functional proteins. Short forms may either act as dominant negatives, like short Sog, or be constitutively active, such as short Sca. Furthermore, the shortening of transcripts and 3' UTRs has been implicated in the activation of oncogenes and the progression of cancer, in the activation of immune cells, and regulation of axon guidance (Flavell et al., 2008; Mayr and Bartel, 2009; Sandberg et al., 2008). Short transcript programs may be more widespread and important during normal development than currently appreciated.

## **METHODS & MATERIALS**

### *Fly Stocks and Husbandry*

All flies were reared under standard conditions at 23°C. *yw* background was used as wild type unless otherwise noted. Fly stocks used in this study are: P{His2Av-mRFP1}III.1 [Bloomington Drosophila Stock Center (BDSC)#23650], *Sxl* RNAi P{TRiP.GL00634}attP40

(BDSC #38195), *sog* P-element disruption  $w^{67c23P}\{GSV2\}$ GS51273 (Kyoto Stock Center#207284), *gd<sup>7</sup>* (BDSC #3109), *sog<sup>Y506</sup>/FM7 ftz-lacZ* (Ferguson and Anderson, 1992), *da1/SM5* (BDSC #273), *da<sup>k08611</sup>/CyO*(BDSC #12385), and *eve Stripe 2-sog* a gift from Hilary Ashe (Ashe and Levine, 1999). Short *sog* and short *sca* cDNA fragments were PCR amplified from cDNA reverse transcribed from embryos aged 1-3 hours using primers (see Table S3 that also introduced *AscI* sites on 5' and 3' ends) and subsequently cloned into the *AscI* site of 2s2FPE (Kosman and Small, 1997), as similarly done for full *sog* (Ashe and Levine, 1999).

*Fly embryos were staged as follows for NC14:*

NC14A: 5–15 min into interphase, with a 1:1 ratio of nuclear length to width, before the start of cellularization.

NC14B: 20–30 min with a nuclear elongation ratio of 2:1 and cellularization progressed <33%.

NC14C: 35–45 min with a nuclear elongation ratio of 3:1 and cellularization progressed <66%.

NC14D: 50–60 min with a nuclear elongation ratio >3:1 and cellularization progressed >66%.

For CRISPR-Cas9 mediated genome editing flies are described in the sections below.

#### *RNA Extraction from Embryos*

All RNA used for RACE, NanoString, qPCR, and 3' RNA-seq was extracted from either a 2-3 hour timed collection of embryos (for RACE) or individually collected and staged embryos (for NanoString, qPCR, 3' RNA-seq) using Trizolreagent (Ambion). Timed pools of embryos were collected from apple juice plates and washed into a 1.5 ml microcentrifuge tube, excess water

removed, and crushed in 1ml of Trizol Reagent (ThermoFisher). The standard Trizol protocol was followed, with the addition of a second chloroform extraction and second 70% EtOH wash. A Histone H2Av-RFP fusion was used to stage individual embryos by nuclear cycle using an epifluorescence microscope (Sandler and Stathopoulos, 2016b). Individual embryos were imaged to confirm correct nuclear cycle, snap-frozen in Trizol using liquid nitrogen, and stored at -80° C until RNA extraction.

#### *Generation of cDNA Libraries to Map Transcripts*

Rapid amplification of cDNA ends (RACE) libraries were created using the GeneRacer kit (ThermoFisher) for the purpose of mapping 3' ends of transcripts. Standard protocol was followed, consisting of RNA extraction as described above, dephosphorylating mRNA using Calf Intestinal Alkaline Phosphatase (CIP), decapping mRNA using Tobacco Acid Pyrophosphatase (TAP), serial ligations of a 5' RNA oligo adapter and a 3' oligo dT adapter, and reverse transcription using Protoscript II (NEB). Extracted RNA was treated with DNase I (NEB) prior to library construction. Nested 5' and 3' RACE primers were designed to capture alternative start sites or truncations of the genes *sog*, *NetA*, *sca*, *Pka-C3*, and *vn*. Both 5' and 3' primers were designed to multiple exons of each gene to capture as much diversity as possible. RACE experiments were performed on RNA extracted from embryos aged 2-3 hours, which includes both NC13 and NC14. We recovered a single short isoform for each of the genes, using two separately prepared RACE libraries and sequencing eight individual RACE products per gene for both libraries. This repeated validation recovering the same short sequences for all four genes further verifies that the RACE products recovered were mature transcripts.

### *NanoString Assay to Quantify Levels of 5' and 3' Ends of sog and NetA Transcripts*

We used NanoString technology, which directly counts mRNA transcripts using gene-specific fluorescent barcodes, without reverse transcription, fragmentation or amplification, to observe the expression of 5' and 3' ends of the genes *sog* and *NetA* (Geiss et al., 2008, Sandler and Stathopoulos, 2016b). Once extracted from individually staged embryos, total RNA was hybridized with NanoString probes at 65°C for 18 hours and then loaded onto the NanoString nCounter instrument for automated imaging and barcode counting. To normalize between embryos and allow for absolute quantification, 1µl of Affymetrix GeneChip Poly-A RNA Control was spiked into Trizol with each embryo at a dilution of 1:10000 before RNA extraction. A linear regression was made for RNA spike-in input versus counted transcripts, and all other genes were fit to the regression and quantified.

### *Fluorescence In Situ Hybridization Staining and Signal Quantification*

Embryos aged 1-4 hours were collected and fixed using standard protocols, and Fluorescence In Situ Hybridization (FISH) was performed in order to identify transcripts in situ using labelled riboprobes following published methods (Kosman et al., 2004) but omitting Proteinase K treatment, briefly described below, To start, timed embryos were collected from apple juice plates, washed to remove yeast and debris, bleached to dechorionate, and fixed in 1:1 formaldehyde:heptane. Embryos were devitellinized and stored in MeOH at -20°C. To perform in situ hybridization, embryos were transferred to EtOH, cleared using xylenes, rehydrated and fixed in PBS, and equilibrated in hybridization solution at 55°C. Probe hybridization was done in an Eppendorf ThermoMixer C instrument at 55°C for 18 hours, gently agitating every 30 minutes. Riboprobes were synthesized using T7 RNA

Polymerase and digoxigenin or biotin labeled NTP nucleotides (Roche) and a primary antibody to Histone H3 (Rabbit anti-H3, 1:10000; Abcam) was used to label histones for precise embryo staging by nuclear cycle. Embryos were sectioned along the anterior-posterior axis manually using a razor blade, and cylindrical mid-embryo sections were imaged face-on. FISH signal was quantified by normalizing signal intensity from probes to 5' and 3' ends of genes compared to signal intensity from histones in individual embryos.

#### *Preparation of Extracts and Sxl Western Blots*

Extract equivalent to 16 embryos was loaded for all samples, except the 0-4hr wild-type (WT), which was loaded with 20 embryos. For unfertilized eggs and specific nuclear cycles, samples were pooled and lysed directly into 2X SDS sample buffer. Embryos from specific nuclear cycles were identified, added to the lysate pool in 2X SDS sample buffer, and snap-frozen in liquid nitrogen to prevent further development until all embryos were collected for each stage. For 0-4hr WT embryos, a large collection of embryos was taken, counted, and lysed in PBS pH7.4 with 6M urea and 1% CHAPS, incubated 10 minutes on ice, homogenized, and spun for 20 minutes to pellet debris, followed by addition of SDS lysis buffer to a 1X concentration. For the Sxl RNAi NC10-13 sample (Figure S2M), the bands in the vicinity of Sxl are somewhat warped due to a local deformation of this particular gel, but the background bands are still visible.

Extracts were separated by discontinuous denaturing 9% SDS-PAGE with AccuRuler RGB Plus/Bluestain molecular weight marker (Gold Biotechnology), and transferred to PVDF (Immobilon-P, Millipore) for Figure 2H, or BA85 Whatman Protran nitrocellulose (Figure S2M and S2N) in Towbin buffer (25mM Tris, 192mM Glycine) with 5% (v/v) methanol. The membrane was rinsed extensively with dH<sub>2</sub>O, equilibrated for several minutes in TBS-T (pH 7.5 with 0.05%

Tween-20), and blocked with 0.2% BSA (w/v) in TBS-T for five minutes, followed by a 10 minute TBS-T wash. The membrane was incubated overnight at 4°C with antibodies diluted in 4ml TBS-T. Mouse  $\alpha$ -Sxl M114 (Bopp et al., 1991) was diluted 1:50, as was mouse  $\alpha$ -BicD 1B11 (Suter and Steward, 1991). Membranes were washed 5x10 minutes, incubated with HRP-conjugated goat- $\alpha$ -mouse (Millipore 12-349) at 1:10,000 in TBS-T for one hour, washed as above, and rinsed extensively with TBS. The blot was developed with ProSignal Dura (Genesee Scientific) diluted 1:7 in TBS for each component, and detected with HyBlot CL film (Denville Scientific). Blots were stripped with 0.1M glycine pH 2.3 with 2% Tween-20 (v/v), and 5% SDS (w/v), washed extensively with TBS-T, reblocked as above, and reprobbed.

#### *Immunostaining of Drosophila Embryos*

Concurrent immunostaining was done with in situ hybridizations using the same methods of fixation and probe hybridization as described above. Embryos were incubated in a 1:10 dilution of primary antibody supernatant ( $\alpha$ -Sxl M114 or M18, or  $\alpha$ -PhosphoSmad1/5) overnight at 4°C, then the antibody was washed off and embryos were incubated in a fluorescent secondary (Alexa Fluor 647 donkey  $\alpha$ -Mouse, 1:500) for one hour at room temperature. Embryos were then washed and mounted for imaging.

#### *RNAi Experiments Using a Heat-Shock Gal4 Approach to Knock-Down Maternal Transcripts Midway through Oogenesis*

In most cases, the use of RNAi against or mutation of the selected RPBs causes sterility or is lethal (Johnson et al., 2010, Staller et al., 2013, Yan et al., 2014). Therefore, we employed combined

heat-shock Gal4 driver with UAS-RNAi lines to generate female flies primed for RNAi (Staller et al., 2013) using an empirically devised heat-shock approach to allow the early stages of oogenesis to proceed normally and to support RNAi later in oogenesis so that maternal product in the egg would be depleted. We crossed Hsp70-GAL4 flies (BDSC #2077) to UAS-RNAi line for *Sxl* (BDSC #38195). Once a stock with both components was generated, virgin females were collected and crossed back to males of the original RNAi stock. Flies were heat-shocked three days in a row at 37°C for 1.5 hours, and embryos collected on the three subsequent days. Flies from the same cross were kept without heat shock and embryos collected in parallel, as a control to confirm any phenotypes seen were due to RNAi and not non-specific effects of the constructs.

#### *CRISPR-Cas9 Mediated Genome Modification*

To target a deletion of the new exon or *Sxl* binding sites located downstream of the *sog* truncated transcript 3' end, a transgenic line was generated expressing two guide RNAs (gRNAs) targeting the region that includes the new exon or *Sxl* binding sites at *sog* locus. First, the unique PAM recognition sites were identified flanking this region using the flyCRISPR optimal target finder (<http://tools.flycrispr.molbio.wisc.edu/targetFinder>). Subsequently, these two sites were cloned into the plasmid pCFD4-U6:1\_U6:3tandemgRNAs (Addgene plasmid#49411). The plasmid including these two PAM sites was injected into *y<sup>2</sup>cho<sup>2</sup>v<sup>l</sup>; P {nos-phiC31\int.NLS}6X; attP2 (III)* (NIG-Fly #TBX-0003), resulting in phiC31-mediated site-integrated transgenesis at landing site attP2 (Chr. III) (Kondo and Ueda, 2013). Integration in the genome at this position was confirmed by PCR/sequencing.

We attempted to delete the new coding exon of short *sog*, but no PAM sequences were available, so to delete the new 3' UTR, non-homologous end joining (NHEJ) mediated by the CRISPR-Cas9

genome editing system was utilized (Kondo and Ueda, 2013). *y<sup>2</sup>cho<sup>2</sup>v<sup>l</sup>;sp/CyO;P {nos-Cas9,y<sup>+</sup>,v<sup>+</sup>} 2A* (NIG-Fly #Cas-0004) virgin flies were collected and crossed with gRNA transgenic male flies. The individual progeny were screened by PCR and sequencing for the deletion ( $\Delta$ New Exon, see below). The end result is a deletion of the short Sog 3'UTR sequence that destabilizes the transcript.

> $\Delta$ New3'UTR (black=genomic sequence, blue=introduced sequence/junction)

agtcctagcataaccattcatagcagctgccacacagaaca

To delete the region including Sxl binding sites at the *sog* locus, homology directed repair (HDR) mediated CRISPR-Cas9 system was utilized (Gratz et al., 2014). A donor construct was generated using pHD-DsRed vector (Addgene plasmid #51434). An ~1kb 5' or 3' homology arm to the regions either upstream or downstream of the Sxl binding sites at the *sog* locus was cloned with SmaI/NheI or AscI/XhoI, respectively (creating HDR.del.sxl).

*y<sup>2</sup>cho<sup>2</sup>v<sup>l</sup>;sp/CyO;P {nos-Cas9,y<sup>+</sup>,v<sup>+</sup>} 2A* (NIG-Fly #Cas-0004) virgin flies were collected and crossed with gRNA transgenic male flies. Embryos were collected and injected with 300 ng/ $\mu$ l of the donor vector. By HDR mediated CRISPR-Cas9, an ~1.1kb region including four Sxl binding sites was replaced by a ~1.3kb fragment, which induces RFP expression in eyes (3xP3-DsRed); essentially retaining similar organization at the locus save presence of Sxl binding sites/associated sequence. The deletion of the region including Sxl binding sites was confirmed by expression of RFP in adult fly eyes and by sequencing. The RFP marker was subsequently removed by crossing the line to a Cre expressing fly line (*y[1] w[67c23] P{y[+mDint2]=Crey} 1b; D[\*]/TM3, Sb[1], BDSC #851*). Excision of the marker was confirmed by PCR ( $\Delta$ Sxl, see below).

> $\Delta$ Sxl (black=genomic sequence, blue=introduced sequence/junction, purple=loxP remnant sequence after Cre-mediated excision)



ctcgtttttctttactttacttagcaatttgttgAAAAAAAAAAAAACAAAAAgttttgcaccgcttccaaaaagaaaactccca  
 acgcaactcgttgccataaatagttagaaggcacggcatatgcacacctgcatcgtagtgccccaactggggtaacctttgagttcttca  
 gttgggggcgtagataacttcgtataatgtatgctatacgaagttatagaagagcactagtaaagatctccatgcataaggcgcgcccgcgcg  
 gctttccagcgagac

This sequence was added to the left homology arm of the HDR donor construct used to generate the Sxl deletion (see above) following NotI/NdeI digestion, and used as donor construct in order to mutate the 4 Sxl binding sites. CRISPR-Cas9 screening to identify changed genomic sequence as well as DsRed RFP marker removal, leaving behind a loxP footprint, were conducted as described (Gratz et al., 2014). To confirm mutated sequence, genomic DNA was extracted, PCR amplified, and sequenced (mutSxl, see above).

$\Delta$ New3'UTR,  $\Delta$ Sxl, and mutSxl fly stocks are viable and fertile.

#### *RNA IP and qPCR to Assay Sxl Association with Transcripts*

Nuclear extract preparation was based on a previously described method (Kamakaka et al., 1991). Approximately 0.4g of 2-4 hour O-R embryos were collected at 25°C and dechorionated for 3 minutes according to standard protocols in 50% bleach, washed with water, followed by a Triton-NaCl embryo wash, then rinsed with water. All following steps were performed on ice or at 4°C. Embryos were homogenized in a 2ml dounce (10 passes with pestle A, 3 passes with pestle B) in NE I (15mM HEPES pH 7.4, 10mM KCl, 5mM MgCl<sub>2</sub>, 0.2mM EDTA, and 350mM sucrose supplemented with 1x Complete protease inhibitors and PhosStop (Roche)), at a ratio of 2 ml buffer to 1g embryos. Extract was filtered through miracloth to remove debris. Nuclei were collected at 3000 x g for 10 minutes, then washed twice with NE I with gentle resuspension of nuclei, while

avoiding yolk and other embryonic debris with each wash. Nuclei were then resuspended and disrupted in 150ul of NE II (50mM HEPES pH 7.4, 300mM NaCl, 0.1% Tween-20, 10% glycerol, and 0.1mM EDTA supplemented with inhibitors as in NE I) and incubated on ice for 12 minutes. The extract was spun in a microfuge at top speed for 30 minutes to remove debris.

For IP, the extract was diluted 1:1 with binding buffer (25mM HEPES pH 7.4, 10% glycerol, 1mM EDTA, 5mM KCl, and 1mg/ml BSA), using 150ul of diluted extract for each IP. Antibody-Protein G complexes were prepared by incubating 50ul of supernatants of  $\alpha$ -Sxl (DSHB M114) or  $\alpha$ -Ubx (DSHB Ubx/ABD-A FP6.87) in binding buffer with 30ul of Protein G beads for 1.5 hours in a total volume of 400ul, washed 2X with binding buffer, 2X with wash buffer (40mM HEPES pH 7.4, 300mM NaCl, 10% glycerol, and 0.2% NP-40), then 2X with binding buffer. Diluted nuclear extract was incubated with prepared beads with agitation for 1.5 hours, and washed 4X with wash buffer. Immunoprecipitated material was eluted with 100ul of 50mM HEPES pH 7.4, 2% Sarkosyl, and 10mM DTT for 30 minutes at 50°C. Proteinase K was added to the eluted material to a final concentration of 1mg/ml and incubated at 50°C for 30 minutes.

RNA was extracted from eluate using acid phenol:chloroform, pH 4.5 (Ambion), followed by chloroform extraction, isopropanol precipitation, and wash in 70% EtOH. RNA was treated with DNase I (NEB) and reverse transcribed using Protoscript II (NEB). qPCR was performed on cDNA using SYBR Green I Master Mix (Roche) on a StepOnePlus Real-Time PCR System (Applied Biosciences) using primers listed in Table S3. For long genes *sog*, *NetA*, *Pka-C3*, and *sca*, primers used all amplified the 5' exons of the genes expressed as part of the short forms. Relative quantification performed using the  $2^{-\Delta\Delta Ct}$  method (Livak and Schmittgen, 2001).

### *3' RNA-Seq to Detect Global 3' Ends of Genes in the Embryo*

RNA from pools of 50 embryos each from NCs 13 and 14 was extracted as described above. A sequencing library was created using a previously described method (Lianoglou et al., 2013) with modifications. Biotinylated oligo dT adapters with an rU residue in the dT section are were conjugated to M-280 Streptavidin Dynabeads (Invitrogen), and first and second strand cDNA synthesis were subsequently performed with Superscript III (Invitrogen) and DNA pol I (NEB). A single strand nick was introduced at the rU residue using Rnase HII (NEB), and translated using *E. coli* DNA Pol (Neb) for eight minutes at 8°C, approximately 100 bases from the original site of the nick. DNA fragments were cleaved and blunted at the site of the translated nick with T7 Exonuclease (NEB), Mung Bean Nuclease (NEB), and Klenow DNA Pol I (NEB). Illumina TruSeq adapters were ligated onto the DNA fragments at two-fold lower concentration than the original protocol in order to reduce unincorporated adapters. The library was PCR amplified through 15 cycles, and final library was size-selected at 150-210 bp. The concentration of ligated sequencing adapters was lowered two-fold to decrease unincorporated adapters sequenced, and final library was size-selected from a 2% Ultra Pure LMP Agarose (Invitrogen), extracted from gel slices using  $\beta$ -Agarase I (NEB), and purified with a phenol:chloroform extraction as described above. Libraries were sequenced on an Illumina HiSeq2500 and sequenced aligned to the FlyBase (April, 2006) annotation using Tophat version 2.0.13 and Bowtie 1.1.1 as the aligner (Kim et al., 2013, Langmead et al., 2009).

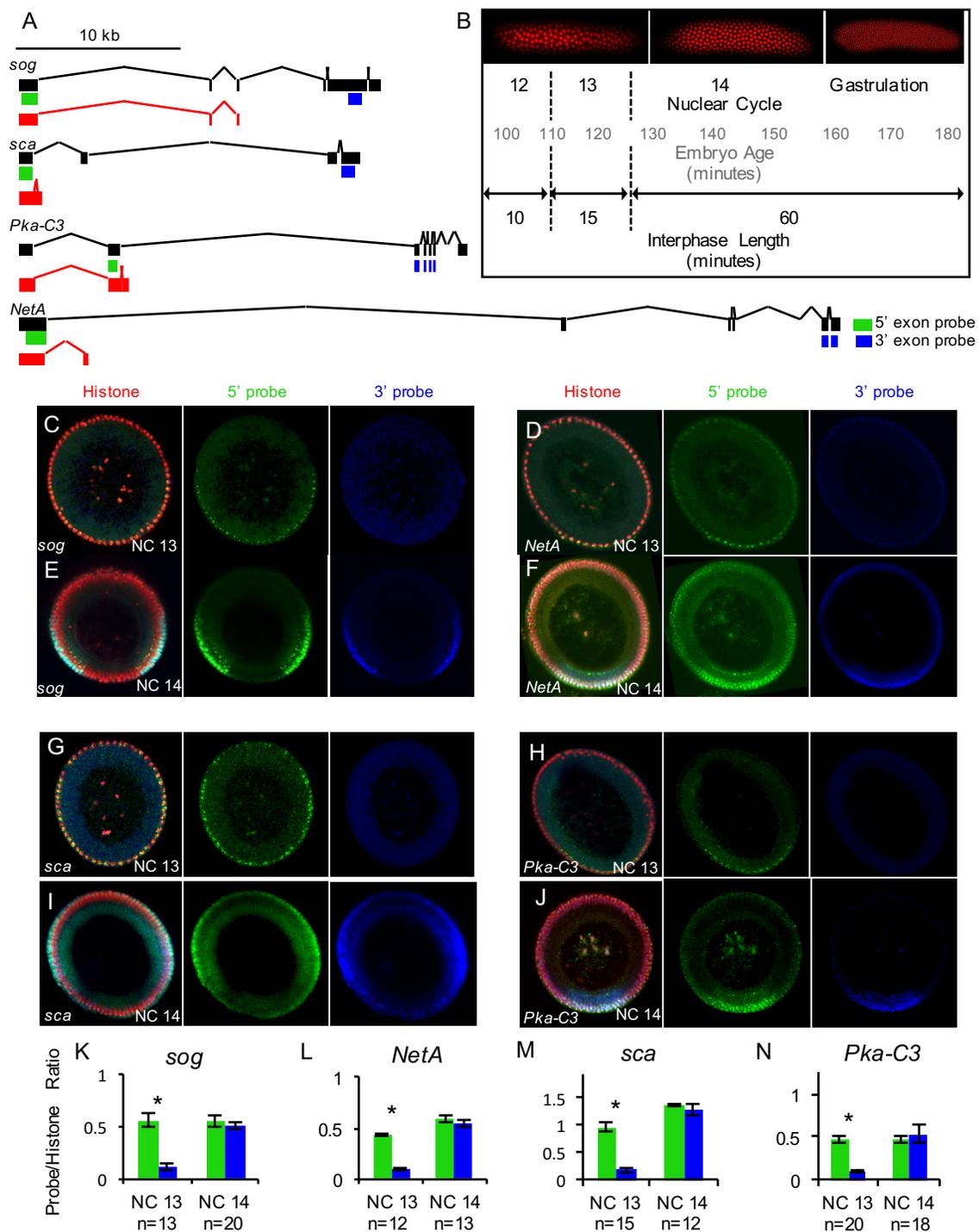
RNA-seq libraries from two separate biological replicates for each nuclear cycle were prepared and sequenced independently. The first replicate was sequenced to a depth of ~25 million reads, and the second replicate was sequenced to a depth of ~150 million reads. Internally primed reads were filtered out of the aligned reads using python to build a BED file of Poly-A and Poly-T islands

of at least eight bases in length, depending on sequence orientation. BEDTools was then used to intersect the BED file with the aligned reads to filter the reads within 10 bases of a Poly-A or Poly-T island (Quinlan, 2014). Internally primed reads greater than 10 bases away from a polyA stretch were not filtered out. Sequences were split based on strand orientation and separate browser tracks created to display stranded reads, relevant to orientation of genes on positive or negative strand. All sequence data has been uploaded to the NCBI GEO database under accession number GSE108152.

#### *Curation of 3'Seq Reads and GO Analysis*

All 450 genes >15kb and 354 genes 8-15kb were manually inspected, searching for signatures of short forms in the 3' RNA seq data, as seen with *sog*. Genes must have mapped reads in both NCs 13 and 14 to be included in the manual curation. 3' reads must be within 16.5kb of a transcription start site, and not within 10 bases of a poly-A stretch in the genome to be considered valid signatures of short forms. Using the DAVID Bioinformatics Gene Ontology clustering tool (Huang et. al, 2009), we found that the most enriched Gene Ontology (GO) term in 31 short forms was Developmental Protein ( $p=4.8E^{-7}$ ), followed by Differentiation Gene ( $p=2.8E^{-4}$ ).

## FIGURES & LEGENDS



**Figure 1. Long genes are transcribed as short forms in NC13.**

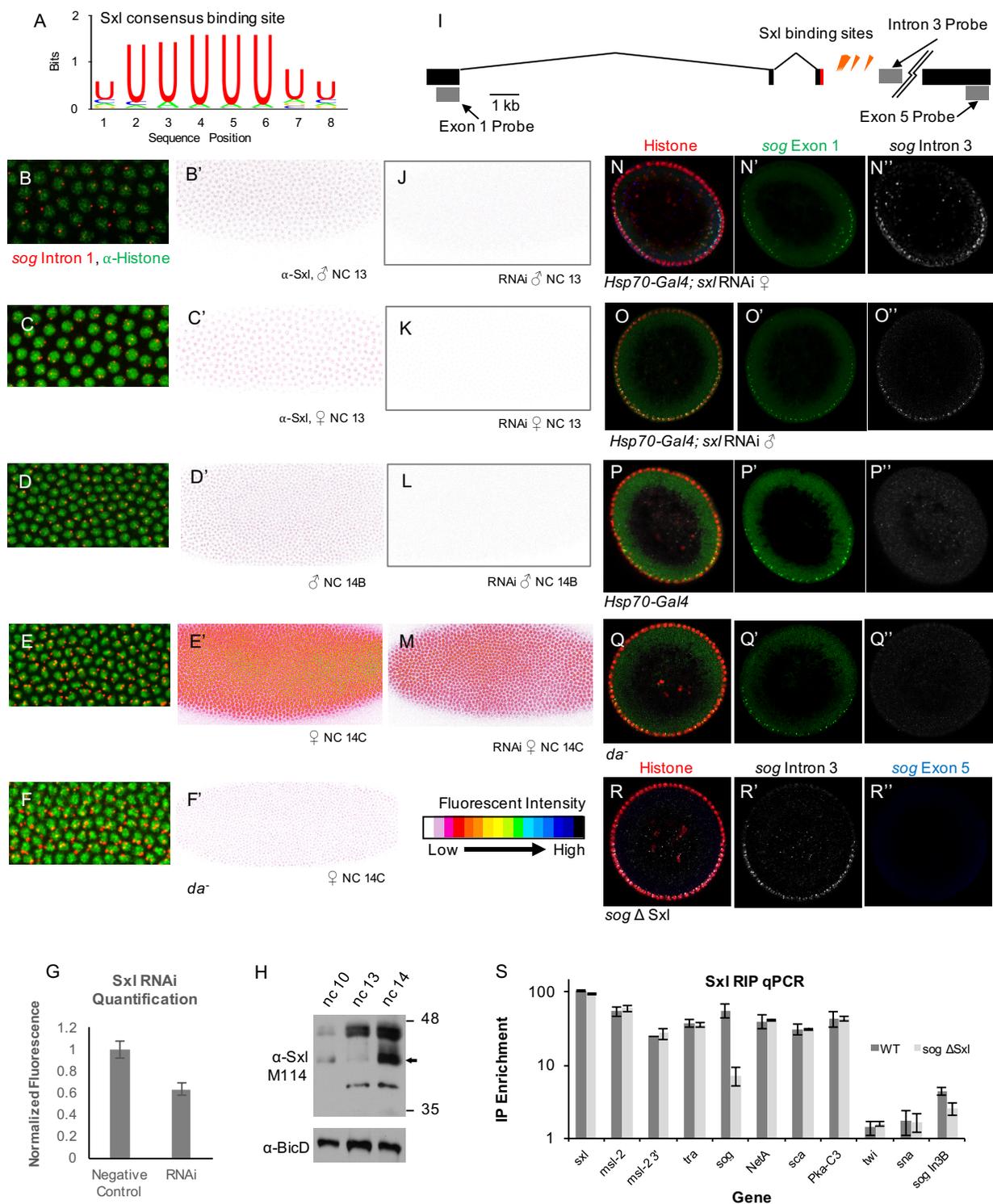
(A) Full-length transcripts (black) and mapped 3' RACE identified shorter transcripts (red) for genes investigated. Locations of 5' and 3' FISH riboprobes shown in green and blue, respectively.

(B) A timeline of the syncytial blastoderm development, showing age of embryo, nuclear cycle, and interphase length for the last three syncytial nuclear cycles. Embryo images illustrate rapid division of nuclei using a Histone H2Av-RFP fusion line.

(C-J) FISH using 5' and 3' riboprobes for the genes *sog*, *NetA*, *sca*, and *Pka-C3* and anti-Histone H3 antibody shown for embryos of stage NC13 and NC14. Images depict manually chopped embryo cross-sections, as described in the methods, stained to show histones (red), 5' probes (green), and 3' probes (blue).

(K-N) Normalization of 5' and 3' FISH riboprobe stainings of genes *sog*, *sca*, *Pka-C3*, and *NetA* to immunostained Histone H3 to compare signal intensity, with number of embryos analyzed for each gene and nuclear cycle below the charts. Differences are present for all genes in NC13 Data are presented as means  $\pm$  SEM. Asterisks specify  $p < 0.0001$ , two-tailed Student's t-test.

(See also Figure S1)



**Figure 2. Early embryonic Sxl is present in early embryos and its depletion leads to defective truncation of long transcripts.**

(A) A position weight matrix for the Sxl consensus binding site on mRNA (Ray et al., 2013).

(B-F) Immunostaining coupled with in situ hybridization to detect both protein and nascent transcripts in wild type (B-E) or *da<sup>-</sup>* (F) embryos. Embryos were stained with anti-histone H3 (green) and FISH using *sog* Intron 1 riboprobe (red) in order to determine nuclear cycle and sex based on nuclear density and number of nuclear dots (I-M), as well as monoclonal anti-Sxl antibody M114 to determine levels of this protein (I'-M', see also fluorescent heat map key lower right). Panels in I-M are magnified views of same embryos shown in I'-M'. (B-E) FISH in NC13 embryos of *Sxl* RNAi (B,C), negative control (D), or *da* mutant (*da<sup>1</sup>/da<sup>k08611</sup>*) (E) backgrounds showing signals for *sog* Exon 1 riboprobe in green, *sog* Intron 3 riboprobe in white. Number of nascent nuclear spots for *sog* gene (on X) also show embryo in C is female and that in D is male. Over 100 individual Sxl immunostained embryos were imaged for these experiments.

(G) Quantification of Sxl RNAi in immunostained embryos in which heat-shock mediated Gal4 expression was used to induce maternal Sxl RNAi or to drive LacZ expression, which served as negative control. Sxl RNAi (n=5) or LacZ positive (n=5) embryo images were analyzed using Fiji software. Embryos were outlined and average fluorescence intensity was calculated. Error bars represent SEM, Student's t-test p=0.0005.

(H) Western blot using M114 monoclonal Sxl antibody to probe extracts of 12 embryos each of NC10, NC13, and NC14. Blot was stripped and reprobbed with anti-BicD as a loading control.

(I) A diagram of the *sog* locus with in situ probe locations (grey boxes), novel short form coding sequence (red box), and Sxl binding site locations (orange arrowheads).

(J-M) Immunostaining for Sxl in *sxl* RNAi embryos, showing a large reduction of maternal Sxl in both male and female NC13 embryos (N and O) and male NC14 (P) embryos, and moderate reduction of Sxl in female NC14 embryos (M), as zygotic expression counteracts RNAi effect. Number of embryos imaged as follows: yw NC13: 47, yw NC14: 110, RNAi NC13: 14, RNAi

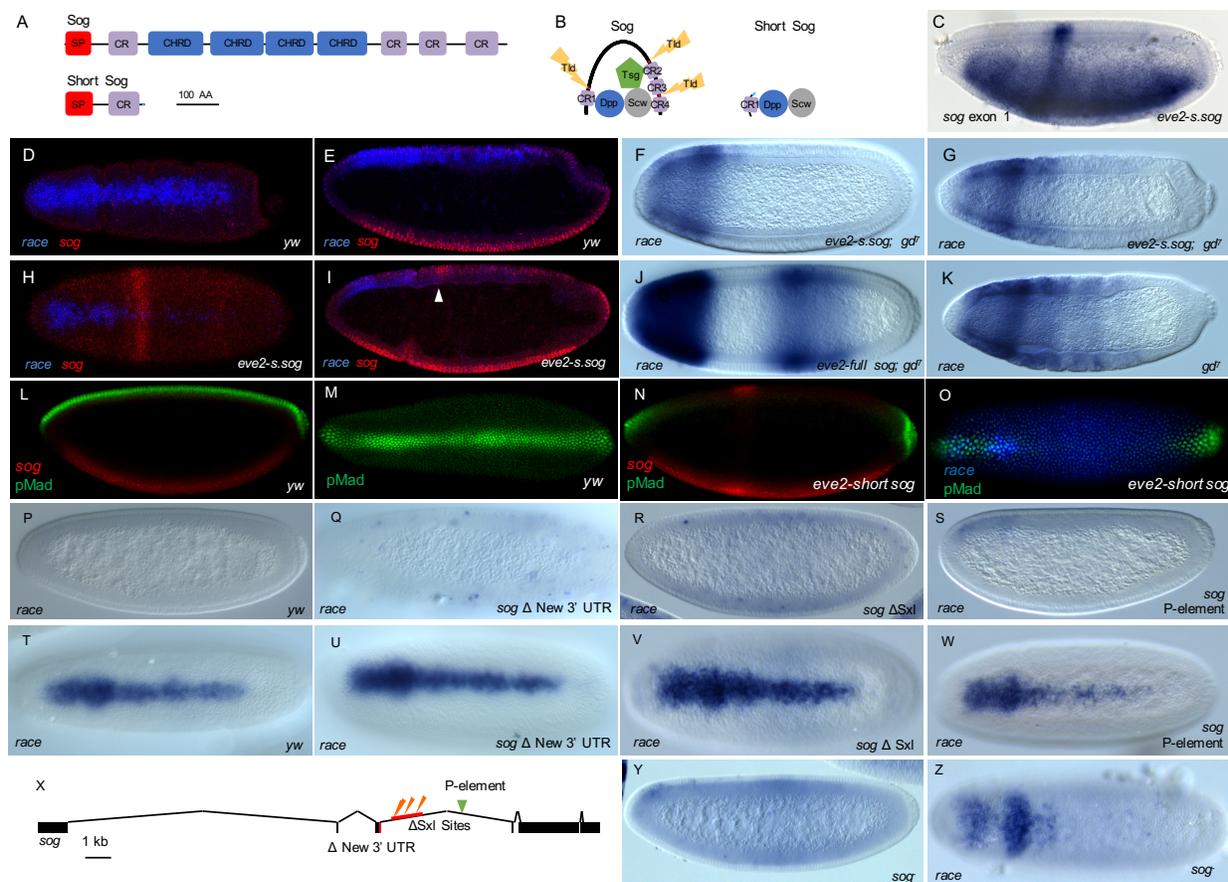
NC14: 25, da- NC13: 7, da- NC14: 14.

(R) FISH detection of *sog* Intron 3 (white) and Intron 5 (blue) with Histone H3 (red) in a *sog*  $\Delta$ *Sxl* mutant embryo of stage late NC13.

(S) Sxl RNA Immunoprecipitation (RIP)-qPCR (see Methods) on RNA eluted from wild type embryos (dark grey) and *sog* $\Delta$ *Sxl* embryos (light grey) using either antibodies to Sxl or Ubx (unrelated factor, mock sample) to immunoprecipitate. Transcripts assayed by qPCR include Sxl sex target genes (*Sxl*, *msl-2*, and *tra*), four truncated genes identified in this study (*sog*, *NetA*, *sca*, *Pka-C3*), short genes (*twi*, *sna*) as well as *sog* transcript past the short-isoform truncation point (assayed by *sog* In3B probe downstream of Sxl binding sites). Data are presented as means  $\pm$  SEM.

Asterisk indicates  $p < 0.0001$ , two-tailed Student's t-test.

(See also Figure S1, S3 and S4)



**Figure 3. Short Sog protein acts in a dominant negative manner to regulate onset of TGF- $\beta$  signaling in the embryo.**

(A) Full Sog protein and predicted Short Sog protein product based on 3' RACE sequence, with functional domains shown. SP=signal peptide, CR=Cystine repeat domain, CHRD=Chordin domain.

(B) Full Sog and Short Sog bound to TGF- $\beta$  ligands Dpp and Scw, along with binding partner Twisted gastrulation (Tsg) and protease Tolloid (Tld) (Shimmi and O'Connor, 2003). Short Sog lacks Tld cleavage sites (red in Sog), and instead contains novel amino acids from intron-derived sequence (blue; see also Figure S2A).

(C) In situ hybridization in *eve* stripe 2-*short sog* embryo, mid NC14, using riboprobe to first exon of *sog*, which recognizes both short and long forms of *sog*, and demonstrates ectopic expression

of short *sog* transcript in the *eve* stripe 2 domain.

(D,E,H,I) FISH colocalization using riboprobes to the TGF- $\beta$  target gene *race* (blue) as well as *sog* (red) at NC14D in wild type embryos (D,E) or *eve2-short sog* background (H,I). Dorsal (D,H; surface plane) and lateral (E,I; sagittal plane) views show single stripe of *race* gene expression present dorsally in wildtype embryos (D,E), while in comparison *race* expression in *eve2-short sog* background is diminished and excluded from the trunk (H,I). *sog* expression is only apparent in lateral view of wildtype embryos (E), but *eve2-short sog* construct supports additional expression in a stripe at the anterior (H and I, see arrowhead).

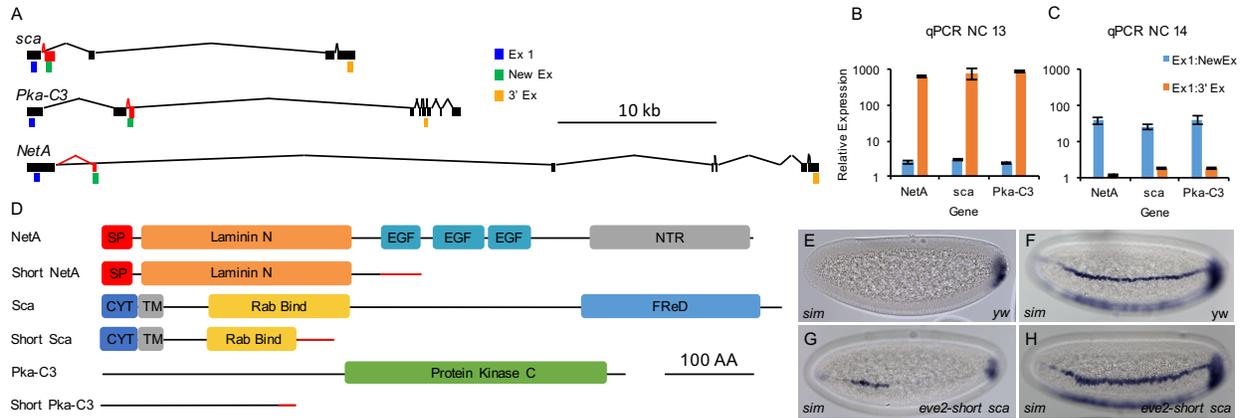
(F,G,J,K) Expression of *race* in *gd<sup>7</sup>* NC14C (F,J) and or gastrulating (G,K) embryos from *eve* stripe 2-*short sog* background (F,G), *eve* stripe 2- *sog* background (J), or native background, not containing any transgene (K).

(L-O) FISH detection using riboprobes to *sog* (red) or *race* (blue) coupled with anti-pMad immunostaining (green) in wild type (L, M) or *eve* stripe 2-*short sog* (N,O) embryos. Lateral (L,N) and dorsal (M, O) views are shown. Images are confocal single scans of the embryos' surface at NC14C-D.

(P-W,Y,Z) *race* expression in *yw* (P,T), *sog  $\Delta$ New 3' UTR* embryos (Q,U), *sog* (R,V), *sog* P-element (S,W), and *sog<sup>Y506</sup>* mutant (Y,Z) embryos. Shown are lateral views of NC14 early stage (P,Q,R,S,Y) or dorsal views of NC14 late stage (T,U,V,W,Z) embryos. Precocious expression of *race* (Q,R,Y) correlates with loss of short *sog* activity.

(X) A diagram of the *sog* locus with sites of genomic manipulation noted.  *$\Delta$ New 3' UTR* is the CRISPR-Cas9 deletion of the novel *short sog* coding sequence derived from the intron.  $\Delta$ Sxl sites is the CRISPR-Cas9 deletion of ~1kb of genomic DNA containing the Sxl binding sites. Green arrowhead indicates the insertion location of *sog* P-element: P{GSV2}GS51273.

(See also Figures S5)



**Figure 4. Differential expression and ectopic expression of short forms for additional genes.**

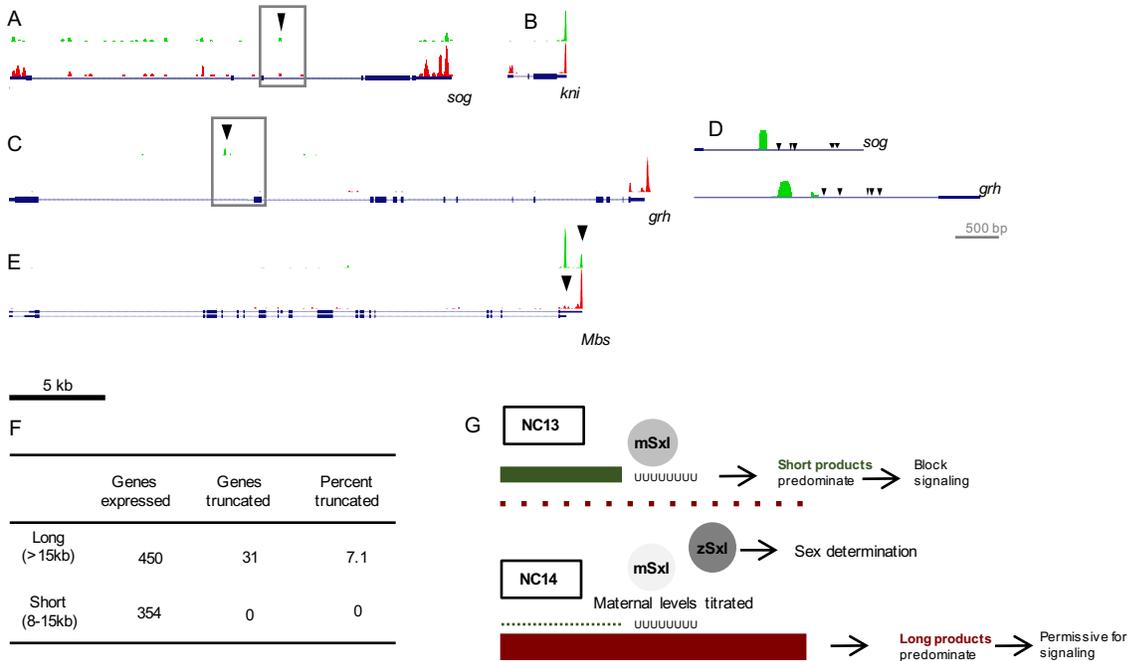
(A) Diagrams of the full-length forms (black) and short forms (red) of the genes *sca*, *Pka-C3*, and *NetA*. qPCR primer locations are also indicated in blue for common 5' exon (Ex1), green for short form specific exon (New Ex), and orange for long form specific 3' exon (3' Ex).

(B,C) qPCR of eight individually staged NC13 (B) or NC14 (C) embryos, comparing expression of 5' exons (Ex1) to both short form specific exons (NewEx) and 3' exons (3' Ex) for *NetA*, *sca*, and *Pka-C3*.

(D) Diagrams of functional domains for full length and short forms of proteins for *Sca*, *NetA*, and *Pka-C3*. SP=signal peptide; NTR=netrin domain; CYT=cytoplasmic domain; TM=transmembrane domain; Rab Bind=Rab protein binding domain; FReD=fibrinogen C-terminal domain; and other domains as marked.

(E-H) In situ hybridization using a riboprobe to detect *sim* expression in either wildtype embryos (E,F) or embryos expressing *eve2-short sca* at early (E,G) and late (F,H) NC14.

(See also Figures S2)



**Figure 5. Global 3' RNA-seq identifies genes with different transcripts lengths in NC13 versus NC14.**

(A-C) 3' RNA-seq browser tracks for the genes *sog* (A), *kni* (B), and *grh* (C) with sequencing reads for NC13 samples displayed in green and those for NC14 samples in red. Black arrows in *sog* and *grh* indicate locations of short form reads.

(D) Zoomed region of short form truncation points for *sog* and *grh*, with black arrowheads indicating Sxl binding sites.

(E) 3' RNA-seq browser track for the gene *Mbs*, showing maternal and zygotic isoforms and the switch in 3' UTR from NC13 to NC14.

(F) Analysis of the 3' RNA-seq displayed in a table displaying number of medium genes (8-15kb) and long genes (>15kb) expressed in both NC13 and NC14, in comparison with number of genes for each class truncated to a shorter form in NC13 compared to length in NC14 (“Genes truncated NC13 vs. NC14”), followed by calculation of percent of total number of genes of Short or Long

class designation exhibiting such a truncation (“Percent truncated”).

(G) Summary of mechanism by which maternal Sxl supports the generation of short transcripts and their requirement to temporally block signaling.

(See also Figure S4 and Tables S1,S2)

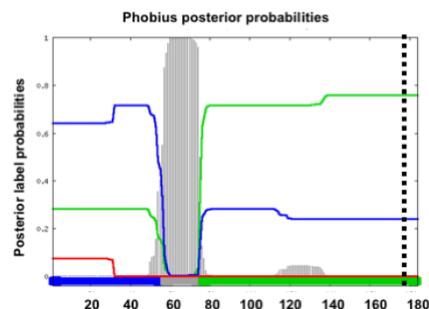
## SUPPLEMENTARY FIGURES &amp; LEGENDS

## SUPPLEMENTAL INFORMATION

A  
*short sog*

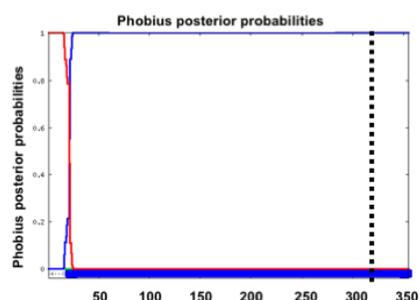
ATGGCCAACAAGCTGAGGAAATCGAACGCCATCGAATGGGCAACGG  
 CCACCGGCACAGTACCGCTCTGGAAAGGAGCTGTCGCCACAGCGA  
 GGACGCCGACTGGAGCCCAAGCGAGCAAAACCGCATAGAGAA  
 CAAGCCCCATCTGCGCCACCTGAGCCAACTGAGCCACCTGTCTCA  
 TCATCGCCGACTGCTGATGCTGCTTGGCGGGCGTGACGGAGGG  
 CCGCCGATGCGCCGCTCATGTTCCGAGGAGTCCGACACGGGCGAG  
 CGGGTCAACCGACAGCGGTACCGAATGCCAGTTTGGCAAGTT  
 TTGGCGAATTGGGTCACCTGGTATGCGGATTTGGGTCACCCCT  
 CGGAGTTATGACTGCATCAAGTGAATGTGGCGATACCAAGA  
 AGCGGCGCATCGTTCCAGCGTCCAGTGTCCGAAATCAAAACGAG  
 TGCCCGCCGGCAATGCGATGATCCATCTCGTTGCCCGGAAATG  
 CTGCAAGACCTGTCCGGCGATCGAAACGTAAGTCCATAGCATAAC  
 CATTCCATATCGATACCAACAATGCTGTACACACTAGAT

MANKLRKSNIAEWATATGTVPLLRSCCH  
 SEDAALPEQASKTSHREQAPILRHSLS  
 HLLIAGLLVLAGVTEGRRHAPLMFEES  
 DTGRRSNRPAVECOFGKVLRELSTWY  
 ADLGPFGVMYCIKCECAVAPKRRIVAR  
 VQCRNKNIECPAPKDDPISLPGKCKCTC  
 PGRNGKSIA\*

B  
*short NetA*

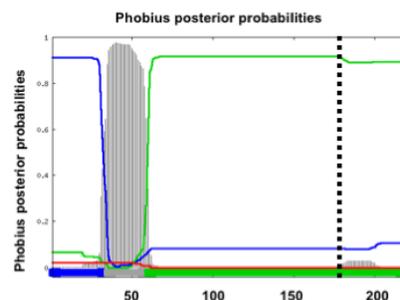
ATGATCCGTGGAATCTTGCTCCTGCTGCTGGGAACACCAGGTTCCAG  
 CCCCATCCAGTGCATCTCCAATGACGTGACTTCAAGATGTTTTCCGA  
 ACAGGCGCGCGGAGGATCCGTGCTACAATAAAGCGCACAGGCA  
 CGCCGCTGCATCCGGACTTTGTGAACGCCGCTACGATGCTCCAGT  
 GGTGGCCAGTTCCAGTGGCGATGTCGGGTGCGCAGCGATATGCG  
 GAGTACCAGGATCAGAGCGCTCTGCCACACCTGCGCATGACGG  
 ATCCGTTGGCAGCTTCCAGCCCGCTCGTACCATTGAATAAC  
 TCGAATAATGACCTGGCGCAGTGAGCCGGTACCAGGAAGTG  
 GGGACAACGTAACCTGACCTCTCGTGGGCAAGAAGTTGAGCT  
 GACTATGTGATCCTACAGCTGTGCCGCATGCTCCGCTCCCGATT  
 CGATGGTATCTACAAGAGCACCGATCACGGGCTCAGTTGGCAGCC  
 ATTCAGTTCTTAGCTCCAGTGGCGCGGCTTTTCGGACGTCCCG  
 CCGCCCAATCCACTGGAGCCCAATGAGCACGAGGCCCGCTGCGAG  
 CGATGTACCAGGCCACTGTCTCCAGAATTGCCCTCAGCACACTGG  
 AGGGACGCCATGTCAGGGGATGCGACTCATCGCCGCTGCTGCA  
 GGACTGGGTGACGGCCACCGATATCCGGGTGGTGTCCATCGCCT  
 CAACGCCCGATCCGACGGCCCTGCTCTATTGGAAGCCGGAGGAG  
 CAACGGATTTGGCCAGCGGCAAGTACAGTGTGCCCTGGCAATGG  
 ACCCGCTGGCAATAATATTGAGGGAATCTCGGTGGGGATGTGGCAA  
 CATCCGCGAGTGGATTGCACTATGCCATCTCTGACTTTTCGGTGGCG  
 GGACGGTGAAGTGAATGGACACGCCCTCAACGCCCTCTCGTCCG  
 CCCCCTCGTTTATTAGGGTTGCAAAATGCAACCGACCTTCTCGCTCTC  
 ACCCCCTCGGATGGCAGCTCAAAGGACTCGCGCAGCATATGATGAT  
 GTTTATGGACAAACGAAATTAACGTTGCCATCGCCGCTTTTACGTTT  
 AATGTGACCTCAAAACAGGCTACAAGGCAACAAAACACCGCAGA  
 CAACAACCAACAGGCAAGCAAT

MRDWQTFPDQLQKKVSRDLNCPA  
 TMAGSNVLWPILLAVLLQISVAFVS  
 GAASGGVVLSDVNNMLRDAKVVTS  
 EKPVVHSHKQETEAPSSVLLRFVD  
 DDEDESIDISSIERDQGRTEMESKMM  
 ADOVRLLTQKLNALMLRRREDEYML  
 EHNLRKSLRLTNNANSVDAMRSEL  
 NQLRHRPAPLYNVSIAICGPRNLALT  
 FPQSRRTCRPTLANKLMAL\*

C  
*short sca*

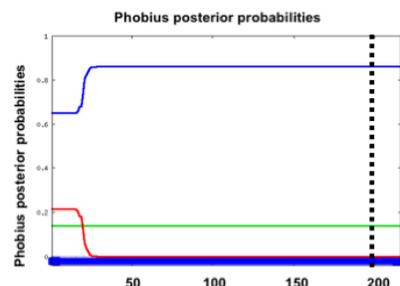
ATGAGAGATTGGCAAACTTCCCGGACCTCCAAAAAAGAAAGTTTC  
 GCGTGACCATTTAAATGGCCCTGCAACAATGGCAGGTTCAAACGTTTT  
 GTGGCAATACCTTGGCCGTGGTCTGCTCCAAATATCCGTGGCAT  
 TCGTGAGTGGAGCGCCAGTGGTGGGTAGTCTTAGCCAGCTGAA  
 AACATGTTGCGCGATGCCAAGGTGGTACCTCGGAGAAACCCGTT  
 GTGCACTAAAACAGGAAACGGAAGCGCCGGAATCCAGCGTGGAG  
 TGCTCCGCTTCTGATGATGACGAGGATAGCGAGGACATCAGCTCC  
 ATGAAACGGCAGGATGGCAGGCAATGGAGAGCAAAAAATGGCCG  
 ATCAGTGGCCCTGTTGACCAAGCAGCTCAACGCCCTGATGCTGCG  
 GCGCCGCGAGGATTACGAGATGCTGGAGCACAATCTGCGCAATCC  
 CTGCGGCTCACCAAGATGCGAAGCAGCTGGACGCCGACATGCGCA  
 CGAACTGAACCACTCAGACACAGACCCGCTCCATTGGTCAATGTC  
 TCGCTGGCAATTTGTGGCCCTAGAACTTGGCGCTAATCTTCCGCA  
 GTCAGTGAACCTGTGCGCCAACTTTGGCCAACTTGTGATGGCC  
 TCTAAGCGGGCAAGTTTAAACAGGGTGAATAATATATCTTTCT  
 TATTTAAGAGTCTAGTTTGTAGTGTAGATTAAATCAATTAAGTT  
 GGACATCAAAATTTGATACGTATGATAAGTAACTCGGGGATGATA  
 CGAATCATGTTAGTTGCTATATTTGCTAGCACTCCTTCACTTTA  
 AGATAGCT

MIRGILLLLGTRFSPQICISNDVYKMF5  
 QQAPPEPCYNKAHEPRACIPDFVNAAY  
 DAPVVASSTCGSSGAQRYCEYQDHRS  
 CHTDMDTDLRSPFARSLDLNNSNNVT  
 CWRSEPVTSQGDVNTLLSLGKFFELTY  
 VILQLPHAPRPDSMVIYKSTDHLSWQ  
 PFQFSSQCRRLFGRPARQSTGRHNEH  
 EARCSVDVTRPLVSRIFSTLEGRPSSRDL  
 DSSPVLQDVRVDIRVVFHRLQRPDPQ  
 ALLLEAGGATDLASGKYSVPLANGPAG  
 NNIEANLGGDVATSGSLHYAISDFVGG  
 RCKCNHGASNASSSAHRFIRVAKCNRP5  
 ASHPLGWQLKGLAQHM\*

D  
*short Pka-C3*

ATGTCCACGGGCACTGTGCCCGCTTTCGACGCCCTTGTGATCCGG  
 CACGGCAGGATCCACATCAAAATGACAACCGGAAACGGAAGCGGC  
 AACAGATGACGCTGGCGTACAATAATCCATCAAAACAGCACAAC  
 AGCAAAATGACAGCAGCAACACAGAGACAACATTCACATCAAAATGG  
 GCCGCTCGAAGCGCCGAAGTAGCAGCAATGTAGCCTCAAGCGAGTC  
 ATCCGATCCACTGGAGTCCGATTACAGCGAGGAGGATCCGGAGCAG  
 GAGCAGCAGCGACCGGATCCGGCGCAACAGCGGAGCAGCAGC  
 ACCGCCACCAACACCAAGTTCCTGCTGACCATGACAATGACGT  
 CGACGAGGAGGATGAGGAGGACGAGCAACGAGGGGAGGGCAA  
 CGGGCGGATCGGATGACGCCACTCAGGATTCACGCGAGAGCATC  
 GAGGAGGACGATGCCAATGAACCGATGACGAGGAGGACGACGAC  
 GAGTGGAGGAGAGCAGCAGCTCCAGACCGCGAAGGAGTGGCG  
 AAATACCACTTGGACACTACCAAGATAATCAAAACAGTGGGTGAGATT  
 GAAAGTCATGGTTGTGATCAGGTCAAGGCTGATTTATGGATGTTGTT  
 TAACCAAAATAGCGCAGTTTTGGTTTTGCTTAAATACCTTCTAAATG  
 TAGTAAATCGTTGGACTATTATAAATGTTTGGACTTAAAAATGTGA  
 TTAACCTCTCTTAAATAAATCATGTTTACG

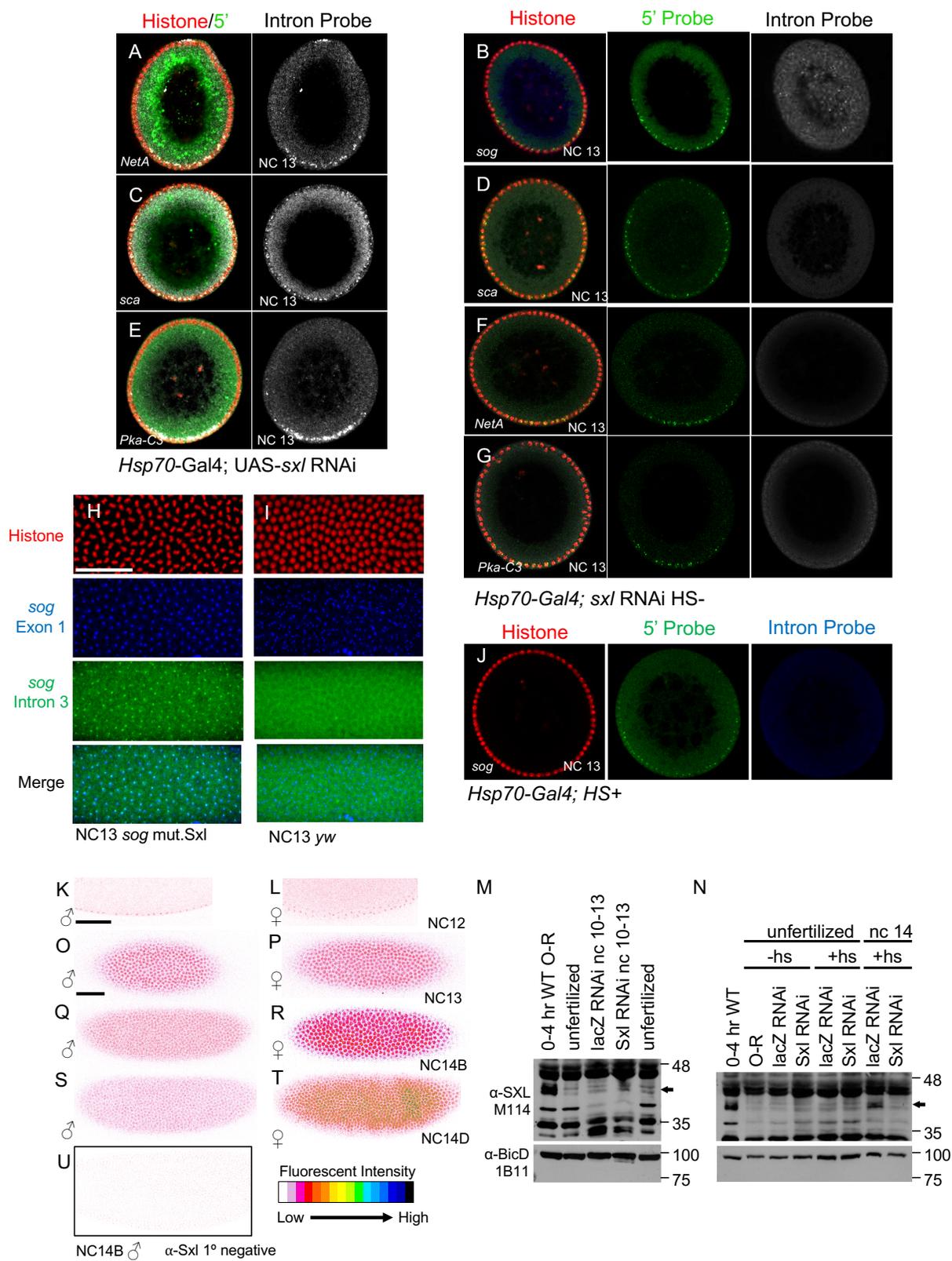
MSTATCARFCTPLSSGTAGSTSKLTTGNG  
 SGNMTMSAYKIPSNSTTANDSNTEFTF  
 TFKLGRSNRSSSNVASESSDPLESDY  
 SEEDPEQEQRDPATNSRSTATTTT  
 TSSADHNDVDEEEDDENEAGEGNR  
 DADDATHDSSEIEEDDNETDDEEDDD  
 ESEESSVQAKGVRYHLDDIYKIVTG  
 QIESHLIRSKADLWMLF\*



transmembrane =      cytoplasmic =  
 non-cytoplasmic =      signal peptide =

Figure S1. Short form sequences. Relates to Figures 1 and 4.

**(A-D)** Coding sequences, including new exons and 3' UTRs, of the four short forms (*sog*, *NetA*, *sca*, and *Pka-C3*) generated using 3' RACE. Black sequence is shared by both short and long forms, blue is new intron-derived coding sequence, and green is new 3' UTR. Amino acid sequences are also shown, with common sequence shared by full length and short form in black, and intron-derived amino acids predicted by 3' RACE in red. Plots show amino acid properties for short form proteins Sog, NetA, *sca*, and Pka-C3 respectively, with novel amino acids specific to the short forms after the dashed line. Blue indicates extracellular domain, green indicates cytoplasmic, red indicated signal peptide, and grey indicates transmembrane domain. Plots were generated using the Phobius tool for amino acid property prediction (Käll et al., 2004). Novel amino acids from the short forms retain the same or highly similar properties of preceding canonical sequence.



**Figure S2. Transcriptional read-through in Sxl RNAi and temporal dynamics of Sxl protein expression in early embryos. Relates to Figure 2.**

**(A-C)** Chopped embryo sections at NC13 showing RNAi against Sxl with FISH detecting the intron region downstream of the 3' RACE truncation point (white) for the genes *NetA* (A), *sca* (C), and *Pka-C3* (E). Histone H3 stained in red to stage embryos. RNAi leads to transcriptional read-through past truncation point.

**(B,D,F,G)** Chopped embryo sections at NC13 showing heat shock negative RNAi control for *sog* (B), *NetA* (D), *sca* (F), and *Pka-C3* (G) using 5' exon and intron probes.

**(H,I)** Multiplex FISH detection of *sog* Exon 1 (blue) and Intron 3 (green) with Histone H3 (red) for staging in wild type (i.e. *yw*) (H) or *sog mut.Sxl* (I) embryos of stage late NC13. When the four Sxl binding site matches in *sog* intron 3 are mutated, transcriptional read-through occurs past the site of truncation exhibited in wild type embryos at NC13. Scale bar is 50  $\mu$ m.

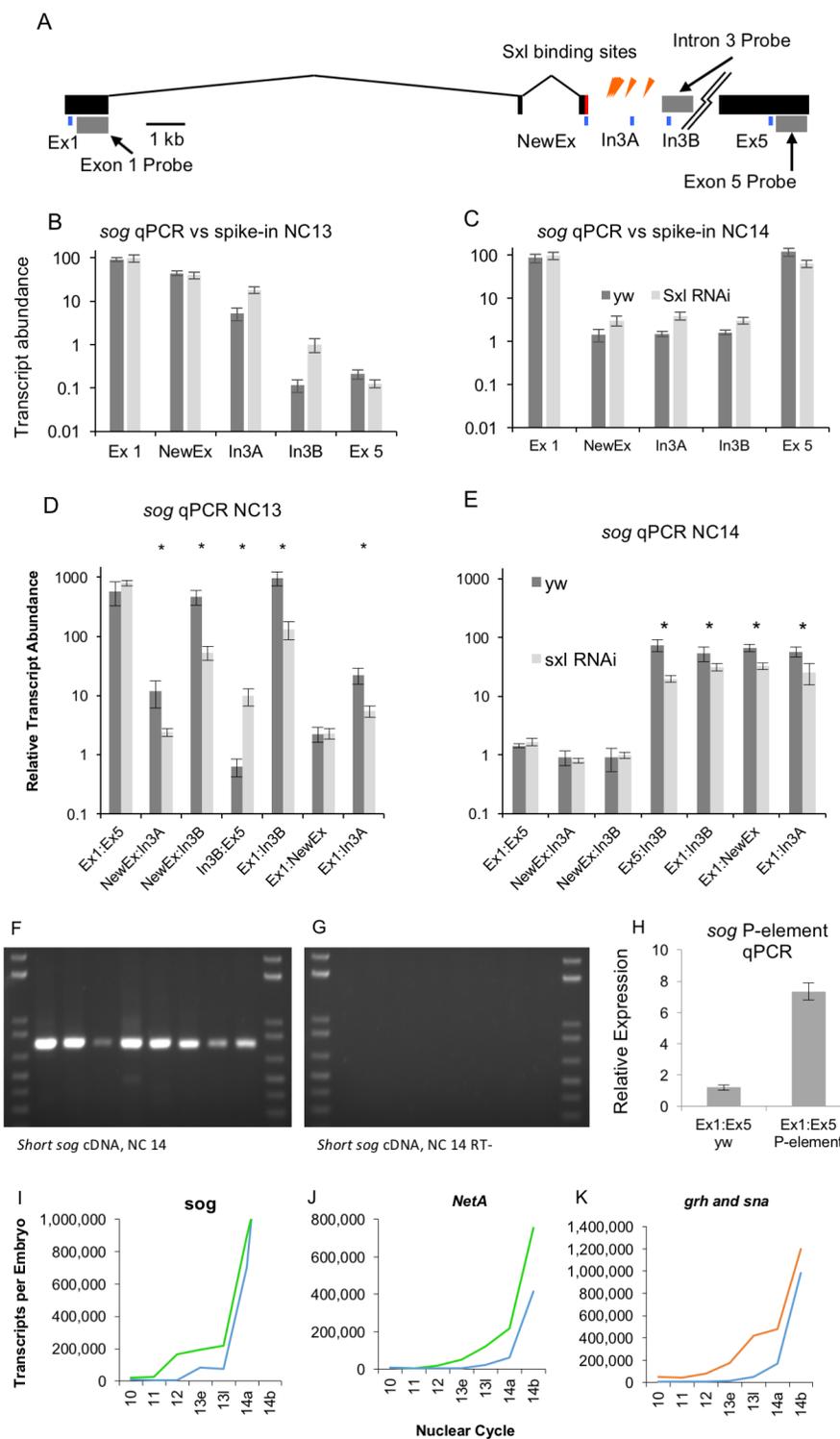
**(J)** Heat shock positive embryo containing only Hsp70-Gal4 driver. Read through into the intron of *sog* does not occur under heat shock only conditions without UAS-RNAi.

**(K,L,O,P,Q,R,S,T)** Embryos of indicated stages immunostained with  $\alpha$ -Sxl antibody M114 assayed as male (K,O,Q,S) or female (L,P,R,T) sex using *sog* intronic riboprobe by in situ hybridization (as described for Fig. 2). Stainings show clear nuclear localization in both males and females, with intensity decreasing from NC14B to NC14D males while it increases during the same time period in female embryos. Number of embryos assayed as follows: K: 3, L:3, O:3, P:5, Q:6, R:2, S:4, T:8, and U:4. Scale bar is 50  $\mu$ m. NC12 embryos in K and L are from a different staining experiment than embryos O-T.

**(M,N)** Western analysis of extracts from staged embryos subjected to RNAi for Sxl using maternal heat-shock protocol (see Methods), or the indicated controls at specific stages. No loss of the bands

seen in the MW range of zygotic Sxl is detectable in samples NC10-13 (left panel) or unfertilized eggs (right panel), but loss of zygotic Sxl in the NC14 sample is apparent (right panel). Extract equivalent to 16 embryos was loaded for all samples, except the 0-4hr wild-type (WT), which was loaded with 20 embryos.

**(S)** 2° antibody only control immunostaining without 1°  $\alpha$ -Sxl, showing lack of nuclear background signal without  $\alpha$ -Sxl antibodies.



**Figure S3. Experiments to assay levels of short and long transcripts. Relates to Figures 2 and 5.**

(A) A diagram of the *sog* locus with in situ probe locations (grey boxes), qPCR primer sets (blue

boxes), novel short form coding sequence (red box), and Sxl binding site locations (orange arrowheads).

**(B,C)** qPCR of RNA obtained from six individually assayed NC13 (B) or NC14 (C) embryos comparing expression levels using primers shown in (A, blue markers) to in vitro transcribed RNA spike-ins (GeneChip Eukaryotic Poly-A RNA Control Kit, Affymetrix, see Methods). Wild type embryos in dark grey, *Sxl* RNAi embryos in light grey.

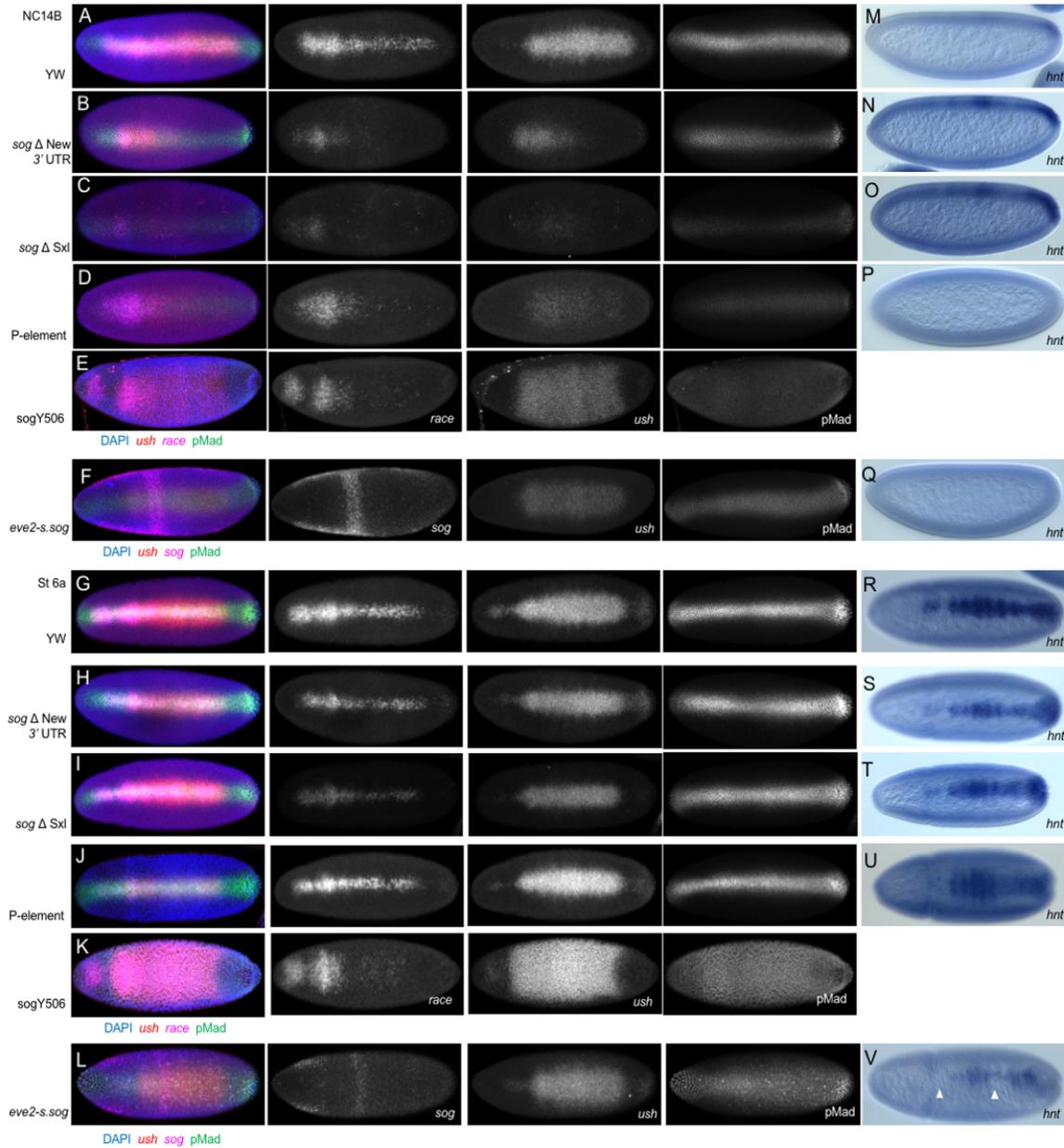
**(D,E)** qPCR of RNA obtained from six individually assayed NC13 (D) or NC14 (E) embryos comparing expression levels between primer sets within the *sog* locus shown in (A, blue markers). Wild type embryos in dark grey, *Sxl* RNAi embryos in light grey. RNAi against *Sxl* leads to a read-through of intronic sequence past the truncation point, but still no full-length transcript. Data are presented as means  $\pm$  SEM. Asterisks signify  $p < 0.0001$ , by two-tailed Student's t-test. (D) During NC13, Ex1 is expressed  $\sim 600$ -fold higher than Ex5 (Ex1:Ex5), indicating that Ex5 is not transcribed, consistent with FISH observations (see Figure 1C). Also during NC13, the novel coding region of truncated *sog* (NewEx) and the Sxl binding site cluster (In3A) are expressed  $\sim 2$ -fold and  $\sim 20$ -fold lower than Ex1 respectively, compared to a  $\sim 65$ -fold decrease for the same sequences during NC14 (Ex1:NewEx and Ex1:In3A, Figures 3D and E), indicating this section of the intron is retained during NC13. There is a marked difference between the three intronic probes during NC13; probe In3B decreased  $\sim 500$ -fold compared to NewEx, though they are equivalent during NC14 (NewEx:In3B, compare D and E). This decrease is similar to the difference between Ex1 and Ex5 (Ex1:Ex5), suggesting a truncation or lack of active transcription after the Sxl binding sites. In NC13, RNAi embryos retain intronic sequence past the Sxl binding site cluster, specifically primer set In3B, expressed  $\sim 10$ -fold more than wild type (Ex1:In3B, D *Sxl* RNAi). During NC13, Exon 5 was transcribed 600-fold less than the Exon 1 in *Sxl* RNAi embryos, indicating that the gene is not completely transcribed even when truncation is defective (Ex1:Ex5,

D *Sxl* RNAi). (E) Within *sog*, the 5' and 3' exons expressed at approximately equivalent levels in NC14 (see Ex1:Ex5), while introns are expressed an average of ~65-fold lower, representing their excision (Ex5:In3B, etc.). *Sxl* RNAi also had an effect in NC14, with the intron spliced out at a slightly lower rate, but no differences between the primer sets within the intron (e.g. see Ex5:In3B and NewEx:In3A, Figure 3E *Sxl* RNAi).

(F,G) *short sog* transcript is detected by RT-PCR using cDNA from eight individual NC14 embryos. cDNA was transcribed using oligo-dT and then amplified using a reverse primer located in novel coding region of *short sog* (F). Reverse Transcriptase negative control on RNA from the same embryos, demonstrating that signal is cDNA-dependent (G). DNA Ladder is 1kb Plus Ladder (ThermoFisher).

(H) qPCR on *sog* comparing the relative expression of Exons 1 and 5 in wild type and the P-element insertion line. In *yw* embryos, Exons 1 and 5 are equivalently transcribed, but in the P-element insertion line, Exon 5 is transcribed 7-fold less than Exon 1.

(I-K) Quantification of transcripts (y-axis individual transcript counts per embryo) within single *Drosophila* embryos of the indicated stages (x-axis) using NanoString (Sandler and Stathopoulos, 2016b). Levels of expression for long genes *sog* (I) and *NetA* (J) were probed at the 5' (green) and 3' (blue) end of genes. *grh* (K) was identified as truncated only after 3' RNA-seq data was obtained, and therefore had only been probed by NanoString using a 3' probe (blue), and is compared to short gene *sna* (orange). Nevertheless, all three long genes show vast upregulation of expression of 3' ends from NC 13 late (13l) to early-mid NC 14 (14a and 14b, respectively). The trajectories of short genes, 5' probes for *sog* and *NetA*, as well as *sna* are more similar. (See also Star Methods)



**Figure S4. Short Sog supports proper spatiotemporal TGF- $\beta$  signaling. Relates to Figure 3.**

(A-L) *race* and *ush* gene expression patterns as well as pMad are affected by partial *sog* mutants in early nc14 embryos but partially recover by stage 6. Dorsal views of embryos at NC14B (A-F) and stage 6a (G-L) stained for *race* or *sog*, *ush*, and pMad. In *sog*  $\Delta$  *New 3' UTR* (B) and *sog*  $\Delta$  *Sxl* mutants (C), *race* and *ush* are retracted and weak with dots of ectopic expression throughout the

embryo, and pMad is also weak compared to *yw* embryos (A). Expression patterns return to the normal domain by late NC14 (H and I), but are somewhat weaker than *yw* embryos (G).

**(M-V)** *hnt* gene expression is also affected by *sog* mutants. Expression of the TGF- $\beta$  target gene *hnt* in early NC14 (M) and late NC14 (R) embryos in *yw* embryos. Early expression in wild type is limited to a patch in the posterior dorsal ectoderm near the pole cells, and expression expands along the posterior  $\sim 3/4$  of the dorsal ectoderm by late NC14. In *sog*  $\Delta$ New 3' UTR embryos (N,S) and *sog* $\Delta$ *Sxl* embryos (O,T), *hnt* expression along the dorsal ectoderm is observed precociously in NC14 embryos (N,O) before it normally appears in wild type, at the same time *race* and *ush* have ectopic expression. *hnt* returns to mostly normal by late NC14 (S,T). In *sog* P-element embryos (P,U), *hnt* expression is not present at early NC14 (P), when the posterior patch first appears wild type (M), but expression is expanded laterally, as is seen in TGF- $\beta$  target genes in *sog* hypomorphic mutants (U)(Ashe and Levine, 1999; Wharton et al., 2004). In embryos expressing *eve 2*-short *sog* (Q,V), *hnt* expression is not observed in early NC14 (Q), and is also decreased in late NC14 (V) with gaps in the expression pattern, as observed for *race* (Fig. 3H,I,O).

**TABLE S2**

<b>Ac78C</b>	<b>CG7029</b>	<b>Mkp3</b>	<b>sdt</b>
<b>boi</b>	<b>dlg1</b>	<b>NetA</b>	<b>sog</b>
<b>btsz</b>	<b>eIF4G</b>	<b>oc</b>	<b>ssp3</b>
<b>CG17514</b>	<b>Fps85D</b>	<b>pan</b>	<b>Stlk</b>
<b>CG32556</b>	<b>grh</b>	<b>Pka-C3</b>	<b>toy</b>
<b>CG42314</b>	<b>L</b>	<b>sca</b>	<b>vn</b>
<b>CG42342</b>	<b>lt</b>	<b>scrib</b>	<b>zip</b>
<b>CG5594</b>	<b>Mbs</b>	<b>sd</b>	

**Table S2. Genes with putative short forms in 3' RNA-seq. Related to Figure 5.**

A list of 31 manually annotated long genes expressed in 3' RNA seq data, in both NCs 13 and 14, with short forms.

## REFERENCES

- Ali-Murthy, Z., Lott, S.E., Eisen, M.B., and Kornberg, T.B. (2013). An essential role for zygotic expression in the pre-cellular *Drosophila* embryo. *PLoS Genet.* *9*, e1003428.
- Ardehali, M.B., and Lis, J.T. (2009). Tracking rates of transcription and splicing in vivo. *Nat. Struct. Mol. Biol.* *16*, 1123–1124.
- Artieri, C.G., and Fraser, H.B. (2014). Transcript length mediates developmental timing of gene expression across *Drosophila*. *Mol. Biol. Evol.* *31*, 2879–2889.
- Ashe, H.L., and Levine, M. (1999). Local inhibition and long-range enhancement of Dpp signal transduction by Sog. *Nature* *398*, 427–431.
- Ashe, H.L., Mannervik, M., and Levine, M. (2000). Dpp signaling thresholds in the dorsal ectoderm of the *Drosophila* embryo. *Development* *127*, 3305–3312.
- Bell, L.R., Horabin, J.I., Schedl, P., and Cline, T.W. (1991). Positive autoregulation of sex-lethal by alternative splicing maintains the female determined state in *Drosophila*. *Cell* *65*, 229–239.
- Bopp, D., Bell, L.R., Cline, T.W., and Schedl, P. (1991). Developmental distribution of female-specific Sex-lethal proteins in *Drosophila melanogaster*. *Genes Dev.* *5*, 403–415.
- Cline, T.W. (1993). The *Drosophila* sex determination signal: how do flies count to two? *Trends Genet.* *9*, 385–390.
- Cline, T.W., Dorsett, M., Sun, S., Harrison, M.M., Dines, J., Sefton, L., and Megna, L. (2010). Evolution of the *Drosophila* feminizing switch gene Sex-lethal. *Genetics* *186*, 1321–1336.
- Cronmiller, C., and Cline, T.W. (1987). The *Drosophila* sex determination gene daughterless has different functions in the germ line versus the soma. *Cell* *48*, 479–487.
- Deignan, L., Pinheiro, M.T., Sutcliffe, C., Saunders, A., Wilcockson, S.G., Zeef, L.A.H., Donaldson, I.J., and Ashe, H.L. (2016). Regulation of the BMP Signaling-Responsive Transcriptional Network in the *Drosophila* Embryo. *PLoS Genet.* *12*, e1006164.
- Flavell, S.W., Kim, T.-K., Gray, J.M., Harmin, D.A., Hemberg, M., Hong, E.J., Markenscoff-Papadimitriou, E., Bear, D.M., and Greenberg, M.E. (2008). Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. *Neuron* *60*, 1022–1038.
- Foe, V.E., and Alberts, B.M. (1983). Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in *Drosophila* embryogenesis. *J. Cell Sci.* *61*, 31–70.
- Fukaya, T., Lim, B., and Levine, M. (2017). Rapid Rates of Pol II Elongation in the *Drosophila* Embryo. *Curr. Biol.* *27*, 1387–1391.
- Garcia, H.G., Tikhonov, M., Lin, A., and Gregor, T. (2013). Quantitative imaging of transcription in living *Drosophila* embryos links polymerase activity to patterning. *Curr. Biol.* *23*,

2140–2145.

Geiss, G.K., Bumgarner, R.E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D.L., Fell, H.P., Ferree, S., George, R.D., Grogan, T., et al. (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* *26*, 317–325.

Gratz, S.J., Ukken, F.P., Rubinstein, C.D., Thiede, G., Donohue, L.K., Cummings, A.M., and O'Connor-Giles, K.M. (2014). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. *Genetics* *196*, 961–971.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., et al. (2011). The developmental transcriptome of *Drosophila melanogaster*. *Nature* *471*, 473–479.

Hoskins, R.A., Landolin, J.M., Brown, J.B., Sandler, J.E., Takahashi, H., Lassmann, T., Yu, C., Booth, B.W., Zhang, D., Wan, K.H., et al. (2011). Genome-wide analysis of promoter architecture in *Drosophila melanogaster*. *Genome Res.* *21*, 182–192.

Hu, X., Lee, E.C., and Baker, N.E. (1995). Molecular analysis of scabrous mutant alleles from *Drosophila melanogaster* indicates a secreted protein with two functional domains. *Genetics* *141*, 607–617.

Johnson, M.L., Nagengast, A.A., and Salz, H.K. (2010). PPS, a large multidomain protein, functions with sex-lethal to regulate alternative splicing in *Drosophila*. *PLoS Genet.* *6*, e1000872.

Käll, L., Krogh, A., and Sonnhammer, E.L.L. (2004). A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* *338*, 1027–1036.

Kamakaka, R.T., Tyree, C.M., and Kadonaga, J.T. (1991). Accurate and efficient RNA polymerase II transcription with a soluble nuclear fraction derived from *Drosophila* embryos. *Proc. Natl. Acad. Sci. U. S. A.* *88*, 1024–1028.

Keyes, L.N., Cline, T.W., and Schedl, P. (1992). The primary sex determination signal of *Drosophila* acts at the level of transcription. *Cell* *68*, 933–943.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* *14*, R36.

Kondo, S., and Ueda, R. (2013). Highly improved gene targeting by germline-specific Cas9 expression in *Drosophila*. *Genetics* *195*, 715–721.

Konrad, K.D., Goralski, T.J., Mahowald, A.P., and Marsh, J.L. (1998). The gastrulation defective gene of *Drosophila melanogaster* is a member of the serine protease superfamily. *Proc. Natl. Acad. Sci. U. S. A.* *95*, 6819–6824.

Kosman, D., and Small, S. (1997). Concentration-dependent patterning by an ectopic expression domain of the *Drosophila* gap gene *knirps*. *Development* *124*, 1343–1354.

- Kosman, D., Mizutani, C.M., Lemons, D., Cox, W.G., McGinnis, W., and Bier, E. (2004). Multiplex detection of RNA expression in *Drosophila* embryos. *Science* *305*, 846.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* *10*, R25.
- Lianoglou, S., Garg, V., Yang, J.L., Leslie, C.S., and Mayr, C. (2013). Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev.* *27*, 2380–2396.
- Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C(T)}$  Method. *Methods* *25*, 402–408.
- Lott, S.E., Villalta, J.E., Schroth, G.P., Luo, S., Tonkin, L.A., and Eisen, M.B. (2011). Noncanonical compensation of zygotic X transcription in early *Drosophila melanogaster* development revealed through single-embryo RNA-seq. *PLoS Biol.* *9*, e1000590.
- Marqués, G., Musacchio, M., Shimell, M.J., Wünnenberg-Stapleton, K., Cho, K.W., and O'Connor, M.B. (1997). Production of a DPP activity gradient in the early *Drosophila* embryo through the opposing actions of the SOG and TLD proteins. *Cell* *91*, 417–426.
- Mayr, C., and Bartel, D.P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* *138*, 673–684.
- McLeay, R.C., and Bailey, T.L. (2010). Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* *11*, 165.
- Miloudi, K., Binet, F., Wilson, A., Cerani, A., Oubaha, M., Menard, C., Henriques, S., Mawambo, G., Dejda, A., Nguyen, P.T., et al. (2016). Truncated netrin-1 contributes to pathological vascular permeability in diabetic retinopathy. *J. Clin. Invest.* *126*, 3006–3022.
- Moschall, R., Gaik, M., and Medenbach, J. (2017). Promiscuity in post-transcriptional control of gene expression: *Drosophila* sex-lethal and its regulatory partnerships. *FEBS Lett.* *591*, 1471–1488.
- Noordermeer, J., Johnston, P., Rijsewijk, F., Nusse, R., and Lawrence, P.A. (1992). The consequences of ubiquitous expression of the wingless gene in the *Drosophila* embryo. *Development* *116*, 711–719.
- O'Farrell, P.H. (1992). Big genes and little genes and deadlines for transcription. *Nature* *359*, 366–367.
- Paz, I., Kosti, I., Ares, M., Jr, Cline, M., and Mandel-Gutfreund, Y. (2014). RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res.* *42*, W361–W367.
- Peluso, C.E., Umulis, D., Kim, Y.-J., O'Connor, M.B., and Serpe, M. (2011). Shaping BMP morphogen gradients through enzyme-substrate interactions. *Dev. Cell* *21*, 375–383.
- Powell, P.A., Wesley, C., Spencer, S., and Cagan, R.L. (2001). Scabrous complexes with Notch to mediate boundary formation. *Nature* *409*, 626–630.

- Pritchard, D.K., and Schubiger, G. (1996). Activation of transcription in *Drosophila* embryos is a gradual process mediated by the nucleocytoplasmic ratio. *Genes Dev.* *10*, 1131–1142.
- Queenan, A.M., Ghabrial, A., and Schüpbach, T. (1997). Ectopic activation of torpedo/Egfr, a *Drosophila* receptor tyrosine kinase, dorsalizes both the eggshell and the embryo. *Development* *124*, 3871–3880.
- Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinformatics* *47*, 11.12.1–34.
- Raftery, L.A., and Sutherland, D.J. (1999). TGF-beta family signal transduction in *Drosophila* development: from Mad to Smads. *Dev. Biol.* *210*, 251–268.
- Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation. *Nature* *499*, 172–177.
- Ray, R.P., Arora, K., Nüsslein-Volhard, C., and Gelbart, W.M. (1991). The control of cell fate along the dorsal-ventral axis of the *Drosophila* embryo. *Development* *113*, 35–54.
- Reeves, G.T., and Stathopoulos, A. (2009). Graded dorsal and differential gene regulation in the *Drosophila* embryo. *Cold Spring Harb. Perspect. Biol.* *1*, a000836.
- Reeves, G.T., Trisnadi, N., Truong, T.V., Nahmad, M., Katz, S., and Stathopoulos, A. (2012). Dorsal-ventral gene expression in the *Drosophila* embryo reflects the dynamics and precision of the dorsal nuclear gradient. *Dev. Cell* *22*, 544–557.
- Rothe, M., Pehl, M., Taubert, H., and Jäckle, H. (1992). Loss of gene function through rapid mitotic cycles in the *Drosophila* embryo. *Nature* *359*, 156–159.
- Rusch, J., and Levine, M. (1997). Regulation of a dpp target gene in the *Drosophila* embryo. *Development* *124*, 303–311.
- Salz, H.K., and Erickson, J.W. (2010). Sex determination in *Drosophila*: The view from the top. *Fly* *4*, 60–70.
- Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A., and Burge, C.B. (2008). Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* *320*, 1643–1647.
- Sandler, J.E., and Stathopoulos, A. (2016a). Stepwise Progression of Embryonic Patterning. *Trends Genet.* *32*, 432–443.
- Sandler, J.E., and Stathopoulos, A. (2016b). Quantitative Single-Embryo Profile of *Drosophila* Genome Activation and the Dorsal-Ventral Patterning Network. *Genetics* *202*, 1575–1584.
- Schneiders, F.I., Maertens, B., Böse, K., Li, Y., Brunken, W.J., Paulsson, M., Smyth, N., and Koch, M. (2007). Binding of Netrin-4 to Laminin Short Arms Regulates Basement Membrane Assembly. *J. Biol. Chem.* *282*, 23750–23758.

- Shermoen, A.W., and O'Farrell, P.H. (1991). Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. *Cell* 67, 303–310.
- Shimmi, O., and O'Connor, M.B. (2003). Physical properties of Tld, Sog, Tsg and Dpp protein interactions are predicted to help create a sharp boundary in Bmp signals during dorsoventral patterning of the *Drosophila* embryo. *Development* 130, 4673–4682.
- Staller, M.V., Yan, D., Randklev, S., Bragdon, M.D., Wunderlich, Z.B., Tao, R., Perkins, L.A., Depace, A.H., and Perrimon, N. (2013). Depleting gene activities in early *Drosophila* embryos with the “maternal-Gal4-shRNA” system. *Genetics* 193, 51–61.
- Stathopoulos, A., and Levine, M. (2004). Whole-genome analysis of *Drosophila* gastrulation. *Curr. Opin. Genet. Dev.* 14, 477–484.
- Suter, B., and Steward, R. (1991). Requirement for phosphorylation and localization of the Bicaudal-D protein in *Drosophila* oocyte differentiation. *Cell* 67, 917–926.
- Tadros, W., and Lipshitz, H.D. (2009). The maternal-to-zygotic transition: a play in two acts. *Development* 136, 3033–3042.
- Wharton, S.J., Basu, S.P., and Ashe, H.L. (2004). Smad affinity can direct distinct readouts of the embryonic extracellular Dpp gradient in *Drosophila*. *Curr. Biol.* 14, 1550–1558.
- Xu, M., Kirov, N., and Rushlow, C. (2005). Peak levels of BMP in the *Drosophila* embryo control target genes by a feed-forward mechanism. *Development* 132, 1637–1647.
- Yan, D., Neumüller, R.A., Buckner, M., Ayers, K., Li, H., Hu, Y., Yang-Zhou, D., Pan, L., Wang, X., Kelley, C., et al. (2014). A Regulatory Network of *Drosophila* Germline Stem Cell Self-Renewal. *Dev. Cell* 28, 459–473.
- Yu, K., Srinivasan, S., Shimmi, O., Biehs, B., Rashka, K.E., Kimelman, D., O'Connor, M.B., and Bier, E. (2000). Processing of the *Drosophila* Sog protein creates a novel BMP inhibitory activity. *Development* 127, 2143–2154.