Genome Activation and Regulation of Signaling in the Rapidly Dividing Drosophila Embryo

Thesis by

Jeremy Edward Sandler

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

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2017
(Defended May 24, 2017)

Jeremy Edward Sandler
ORCID: 0000-0001-8340-7583

ACKNOWLEDGEMENTS

It is hard to find the right words to express my gratitude to my advisor, Professor Angela Stathopoulos, for her support and guidance over the past five years. Angela has helped me grow as a scientist, and importantly, grow as an independent scientist. The positive lab environment fostered under Angela's leadership made it easy for me to come in every day, even when progress was slow. I will always be grateful to Angela for the opportunities she has given me.

To my wife Tammy: you have always been there for me, cheering me on and providing unlimited love and support. I know my long hours in the lab have left you picking up the slack and making sacrifices, especially after Avi was born. I appreciate everything you have done more than I can describe. As I sit here with you and finish the last edits on my thesis, we are in the hospital getting ready to welcome our second baby into our family, and I am reminded how strong you are.

To Avi: you are my motivation every day in the lab. I am amazed at how quickly you have grown up, and how your personality has already developed in just two years. You can accomplish anything you want if you set your mind to it. To Ellie: you are just a few days old as I submit my thesis and I'm already in love with you. I'm excited to watch you grow and see who you become. To my parents: your constant enthusiasm and encouragement mean a lot to me, and I appreciate everything you did to bring me to this point.

I feel very privileged to have spent my graduate career at Caltech, surrounded by experts in every field of biology, and with ample resources to support my studies. I am especially thankful for my committee members, Professor Paul Sternberg, Professor

Marianne Bronner, and Professor Ellen Rothenberg, for guiding my research and providing personal advice when I needed it.

I also want to thank and honor Professor Eric Davidson, who passed away in 2015. He was a committee member and much more. He was a scientific role model as a giant in the field of developmental biology, and I turned to his decades of work when thinking about my own experiments. I took every class that Eric taught at Caltech, and then went on to be his teaching assistant for two years. Eric always pushed me to ask the right questions and to learn more from my data, and I am a better scientist because of him.

I have really enjoyed working with the other members of the Stathopoulos lab; I have learned so much from all of you, and I will miss seeing you every day. Vince Stepanik is immensely knowledgeable and is always willing to share his skills and give his time to help. His excitement about my project means a lot to me. Leslie Dunipace keeps the lab going, both physically as the lab manager and emotionally as someone we can always turn to for advice or just to talk about things. Frank Macabenta has a positive attitude that lifts up the rest of the lab. He is a master of the English language and the best pun crafter I know. Nathanie Trisnadi and Mayra Garcia, both former graduate students, helped me when I first joined the lab and set a high standard for me to follow. I am also grateful to Jihyun Irizarry, Heather Curtis, Theodora Koromila, Jingjing Sun, Zsuzsa Akos, and James McGehee for being great labmates.

I have had institutional support from many individuals, all of who made my time at Caltech easier. Liz Ayala, the graduate coordinator for the division, had an answer to every question and always made sure that students had the support we needed. I would also like to thank the Graduate Dean Professor Doug Rees, Associate Dean Felicia Hunt, and

Assistant Dean Natalie Gilmore. I developed relationships with all three of them during my time at Caltech, and they all helped and supported me in different ways that I am very grateful for.

Lastly, I need to thank the staff at the Caltech Y, especially Greg Fletcher. The Y was like a second home for me at Caltech, providing a place for me to explore the outdoors and grow as a leader. I've gone on amazing backpacking trips all through the Sierras, summited Half Dome, Mt. Dana, and Mt. Whitney among many peaks, plunged into frigid alpine lakes, and even stared down a few bears with the Caltech Y. The memories and friendships I made while on Caltech Y trips will last a lifetime. I met Casey Handmer on my first Yosemite trip with the Y, and from our first hike together, a 17-mile day hike of the Panorama Trail, to the flying trips we are planning right now, I have enjoyed our adventures on foot and in the air. Your intelligence and wit amaze me, and your common sense, straightforward approach to problems, and general life advice are always welcome.

ABSTRACT

Embryonic development of the fruit fly *Drosophila melanogaster* is unique among model organisms and animals in general, as rapid and syncytial nuclear divisions characterize the early stages before cell membranes form. These nuclear divisions occur every eight to fifteen minutes, culminating with a 45-minute cell cycle where cell membranes form and the 6000 nuclei become 6000 cells before the embryo undergoes gastrulation. At the beginning of development, maternally deposited transcripts define the major axes of the embryo and control all processes that occur. As the syncytial nuclear cycles slow and nuclei migrate to the periphery of the embryo, maternal transcripts are degraded and the zygotic genome is first activated. The rapid pace of nuclear divisions concurrent with the activation of the zygotic genome presents unique challenges to the developing embryo, as the constraints imposed by mitosis limit the ability to transcribe new genes. This switch of control, the Maternal to Zygotic Transition, has been the subject of studies at the molecular and genetic level for almost 30 years. Here, we use new tools and approaches to study the developing embryo at a time scale not previously achieved. We show how the gene regulatory network (GRN) along the dorsal-ventral axis, including entire signaling pathways, is activated using time point intervals of 10 minutes. While GRNs compress a 4-Dimensional time course into a 2-Dimensional space to describe gene interactions, we use tools to preserve the 4-D information. Using mutants, we show the contribution of individual genes in the process of development and the resulting changes in expression levels for the entire network. Finally, we examine the transcription of long genes during the rapid syncytial nuclear cycles, when time constraints limit the ability to transcribe the entire gene. We show how an RNA binding protein regulates the truncation of the transcripts into

short isoforms with novel coding sequences, and how these short gene products code for functional proteins that regulate the spatiotemporal activation of key signaling pathways in the embryo.

PUBLISHED CONTENT AND CONTRIBUTIONS

Sandler, J. E. and Stathopoulos, A. (2016) "Quantitative Single-Embryo Profile of Drosophila Genome Activation and the Dorsal–Ventral Patterning Network". *Genetics*, 202.4, 1575-1584. doi: 10.1534/genetics.116.186783.

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

Sandler, J. E. and Stathopoulos, A. (2016) "Stepwise Progression of Embryonic Patterning". *Trends in Genetics*, 32.7, 432-443. doi: 10.1016/j.tig.2016.04.004.

J. E. S. and A. S. wrote the manuscript.

Sandler, J. E., Irizarry, J., Stepanik, V., Amrhein, H. and Stathopoulos, A. (2017) "A developmental program truncates long transcripts to temporally regulate cell signaling". In submission.

J. E. S., J. I., and V. S. 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

LIST OF FIGURES AND/OR TABLES

Number *Page*

CHAPTER 5

Chapter 1

STEPWISE PROGRESSION OF EMBRYO PATTERNING

ABSTRACT

It is long established that the graded distribution of Dorsal transcription factor influences spatial domains of gene expression along the dorsal-ventral axis of *Drosophila melanogaster* embryos. However, the more recent realization that Dorsal levels also change in time raises the question of whether these temporal dynamics are instructive. Here, an overview of dorsoventral axis patterning is provided focusing on new insights into this patterning process identified recently through careful, quantitative analysis of temporal changes in Dorsal target gene expression that result from one nuclear cycle to the next ('steps'). Possible roles for the step-wise progression of this gene expression program are discussed including (i) tight, temporal regulation of signaling pathway activation, (ii) control of gene expression cohorts, and (iii) to ensure irreversibility of the patterning and cell fate specification process.

DISCUSSION

Transcription factor dynamics regulate target gene expression

Subdividing the embryo into distinct domains of gene expression by combinatorial control of transcription factors is an important function of regulatory networks acting in early embryos including those of *Drosophila* (CARMENA *et al.* 1998); STATHOPOULOS AND LEVINE (2005); (DAVIDSON AND LEVINE 2008; ZINZEN *et al.* 2009);

BRISCOE AND SMALL (2015). These early patterning events influence the activation of signaling pathways to support tissue differentiation and also control cell movements required for the generation of a multilayered embryo; the developmental actions that encompass gastrulation (STATHOPOULOS AND LEVINE 2004; LEE *et al.* 2006). To study these events at the transcriptional level in *Drosophila* embryos, previous studies of early zygotic gene expression have considered one or two time-points spanning the first four hours of early embryo development (ZEITLINGER *et al.* 2007; MACARTHUR *et al.* 2009; GRAVELEY *et al.* 2011; OZDEMIR *et al.* 2011), and yet recent studies suggest gene expression patterns change on the order of minutes rather than hours (e.g. LOTT *et al.* 2011; REEVES *et al.* 2012; ALI-MURTHY *et al.* 2013). Furthermore, only recently has it come to light that transcription factors in the early embryo exhibit changes in levels over time (GREGOR *et al.* 2007; SHVARTSMAN *et al.* 2008; KANODIA *et al.* 2009; LIBERMAN *et al.* 2009). At least in part these dynamics relate to the fast nuclear divisions that encompass *Drosophila* early embryonic development and result in oscillatory inputs to target genes. Transcription factor dynamics appear to be a general mechanism of regulating gene expression (LEVINE *et al.* 2013; PURVIS AND LAHAV 2013) and highlight the need to study temporal regulation of developmental gene expression as a complement to previous studies of embryonic patterning in *Drosophila,* which have focused on spatial control of gene expression (STATHOPOULOS AND LEVINE 2002; CHEN *et al.* 2012; WOLPERT 2016).

The Dorsal transcription factor is dynamic as are its target genes

In the *Drosophila* embryo, the pivotal transcription factor, Dorsal, is present in a nuclear-cytoplasmic gradient along the dorsoventral (DV) axis that instructs differential gene expression, yet the establishment of this morphogen gradient is atypical (MOUSSIAN AND ROTH 2005; ROGERS AND SCHIER 2011; STEIN AND STEVENS 2014). *dl* transcripts are maternally deposited and uniformly distributed (SIMPSON 1983; ANDERSON AND NUSSLEIN-VOLHARD 1984). The protein, however, is present in a nuclear gradient through differential activation of the upstream receptor, Toll (ANDERSON *et al.* 1985). Thereby, this gradient does not result from localized expression of Dorsal protein but, instead, involves a nuclear-cytoplasmic shift in levels of this factor along the DV axis as regulated by Toll receptor signaling (ROTH *et al.* 1989; RUSHLOW *et al.* 1989; STEWARD 1989). Dorsal acts as activator of transcription to support the expression of target genes in ventral and lateral regions of the embryo as well as repressor of transcription to limit the expression of a subset of target genes to dorsal regions (RAY *et al.* 1991; JIANG *et al.* 1992; STATHOPOULOS AND LEVINE 2002). In this manner, more than fifty genes are differentially expressed along the DV axis (STATHOPOULOS AND LEVINE 2002; BIEMAR *et al.* 2006). High levels of nuclear-localized Dorsal in ventral regions specify the mesoderm, whereas lower levels of nuclear Dorsal in lateral regions specify the neurogenic ectoderm (CHOPRA AND LEVINE 2009; REEVES AND STATHOPOULOS 2009). The prevailing model in the field had been that the changes in levels of Dorsal in space, along the DV axis, is important for establishing different domains of gene expression.

However, more recent studies have identified that Dorsal levels also change in time (LIBERMAN *et al.* 2009; RUSHLOW AND SHVARTSMAN 2012), raising the question of

whether and how temporal changes of this factor impact gene expression. How the nuclear distribution of Dorsal gives rise to precise gene expression patterns was recently investigated using live *in vivo* imaging and quantitative analysis. It was revealed that the Dorsal transcription factor gradient is highly dynamic, increasing in levels over time, and not achieving steady state until Dorsal levels plummet at gastrulation (REEVES *et al.* 2012). Up to this point during the first three hours of development, levels of this factor build within nuclei, from one nuclear cycle to the next such that by cellularization a ~3-fold increase is realized compared to previous nuclear cycles. In addition, Dorsal levels oscillate with each and every nuclear cycle, dropping rapidly as nuclei divide and Dorsal escapes into the cytoplasm. Following nuclear division, import of Dorsal back into the nucleus is relatively slow leading to a gradual increase. This relatively slow import of Dorsal into the nucleus compared with other transcription factors acting at this time such as Bicoid, for example, likely relates to the requirement of Toll-mediated signaling to mediate entry of Dorsal to the nuclei and explains why levels of Dorsal increase as the length of nuclear cycles increases (Figure 1.1) (BELVIN AND ANDERSON 1996). In contrast, the nuclear distribution of the Bicoid transcription factor stabilizes relatively quickly within every nuclear cycle and, moreover, stays relatively constant from one nuclear cycle to the next (GREGOR *et al.* 2007). The observation that the Dorsal morphogen gradient changes in time, within as well as between nuclear cycles, suggests time impacts gene regulatory network activation.

The levels of Dorsal transcription factor almost double from one nuclear cycle to the next, approximately every 10 minutes (REEVES *et al.* 2012). How might a factor act as morphogen, to control spatial patterning, if its levels constantly change? One possibility is

that these transcription factor dynamics also induce unappreciated gene expression dynamics. By close analysis of the expression associated with four Dorsal target genes within precisely-staged, fixed embryos, two distinct temporal trends were found associated with targets (REEVES *et al.* 2012). Expression of the gene *short-gastrulation* (*sog*) (FRANCOIS *et al.* 1994) was found to be 'plastic' (dynamic), with levels changing constantly both upwards and downwards in time. For *sog*, it appears possible to turn gene expression on/off in time, presumably, in response to changing levels of Dorsal above/below an activation threshold when nuclear concentration oscillates between syncytial divisions. In contrast, other genes expressed along the DV axis, also Dorsal targets, such as *snail* (*sna*) (KOSMAN *et al.* 1991; IP *et al.* 1992) exhibit more of a 'ratchet' (monotonic) effect in that levels continue to steadily increase and expression domains never refine to narrower patterns once established despite changes in Dorsal. This "ratchet effect" is similar to the target response of another morphogen, Activin, important for patterning in *Xenopus* (GURDON *et al.* 1998). Thus, this preliminary analysis of four genes expressed along the DV axis in the *Drosophila* embryo identified two different temporal responses: dynamic (e.g. *sog*) versus monotonic (e.g. *sna*) (REEVES *et al.* 2012). However, as only a small number of targets were examined, it was not possible to distinguish whether these temporal changes were gene-specific responses or general network-wide trends. Furthermore, these dynamics may relate to differences in mRNA stability of transcripts or other post-transcriptional effects that have been little studied in the early embryo in relation to zygotic transcripts.

Case Studies in Transcription Factor Localization and Concentration

Two prominent transcription factors active early in *Drosophila* development are Dorsal and Bicoid. The nuclear concentration, gradients, and embryonic localization of both transcription factors have been characterized, and present a contrast in nuclear import strategies (GREGOR *et al.* 2007; REEVES *et al.* 2012). Both are imported into nuclei during syncytial nuclear cycles, but the dynamics and import rate are different between the two (Figure 1.1).

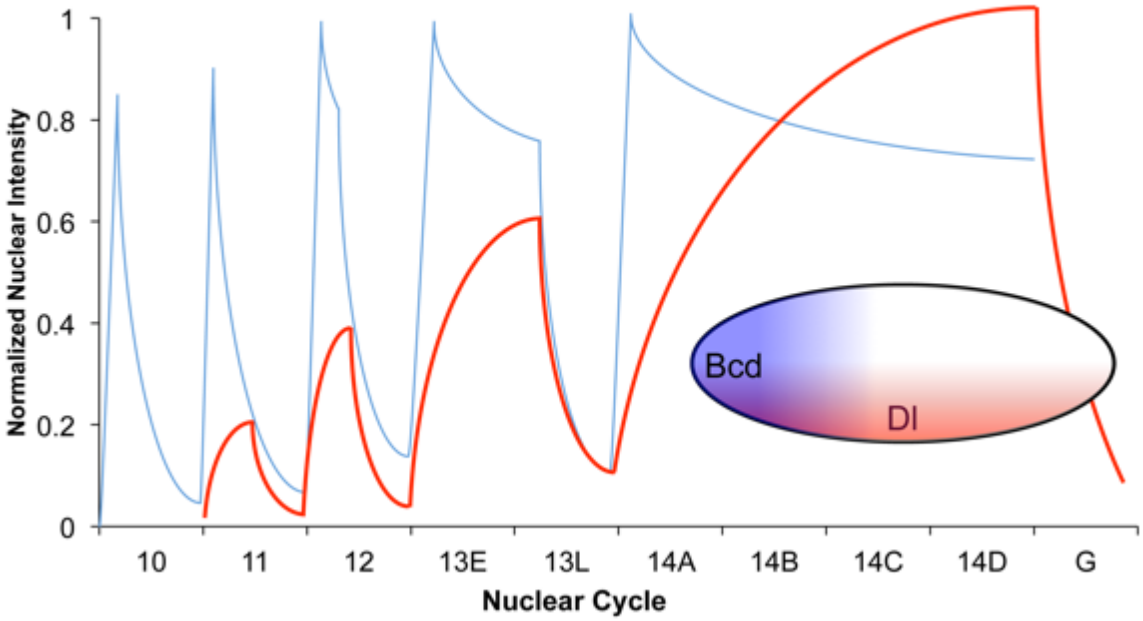


Figure 1.1. Bicoid and Dorsal Dynamics - Comparison of Nuclear Levels. A conceptual representation of the concentration of transcription factors Bicoid (blue) and Dorsal (Dl; red) in nuclei during late nuclear cycles based on data from previous studies (GREGOR *et al.* 2007; REEVES *et al.* 2012). Measurements were obtained by monitoring live fluorescent intensity of a bcd-GFP or dl-Venus fusion molecules from a single nucleus at 10% along the AP axis for Bicoid or ventral most position for Dorsal. Nuclear intensity is normalized to the maximum for each transcription factor and overlaid. Inset is an illustration of a *Drosophila* embryo with transcription factor concentration gradients for Bicoid and Dorsal.

While Bicoid undergoes a rapid uptake, it also undergoes a decrease in concentration before nuclear division, indicating an overshoot and reduction in concentration to a lower steady state. Nuclear cycles 10-12 are too short to reach this overshoot and reduction, but nuclear cycles 13 and 14 show this characteristic, with the concentration of Bicoid stabilizing before mitosis, when it then drops to low levels before being imported again. Dorsal, on the other hand, undergoes a slower increase to maximum levels at each nuclear cycle, with no overshoot. While Dorsal never reaches a steady state during early nuclear cycles, the concentration of Dorsal begins to level off during nuclear cycle 13 and finally achieves a steady state during nuclear cycle 14, demonstrating a different import mechanism than that of Bicoid. Both Bicoid and Dorsal leave nuclei at very similar rates and times between nuclear cycles, indicating that export is likely due to rapid diffusion of the transcription factors when the nuclear envelope breaks down during mitosis.

A temporally fine-scale, quantitative assay of gene expression provides insights into step-wise activation of *Drosophila* embryogenesis

An assay of gene expression dynamics was performed recently using NanoString nCounter technology (Figure 1.2) to measure the levels of expression for ~70 genes in the early *Drosophila* embryo, focusing on those expressed along the DV axis and providing further insight into the dynamics of genes expressed in the early embryo (GEISS *et al.* 2008; SANDLER AND STATHOPOULOS 2016). Ten time points spanning nuclear cycles (NC) 10 through 14 and also including gastrulation were investigated through assay of gene expression within individual, carefully-staged *Drosophila* embryos (Figure 1.2 A-C). Nuclear cycle 14 was divided into four time points, 14A-14D, providing data from before

(14A), during (14B and 14C), and after (14D) cellularization. In this analysis, the data suggested that tight temporal regulation of gene expression is key in the activation of the zygotic gene regulatory network and important for a properly developing embryo.

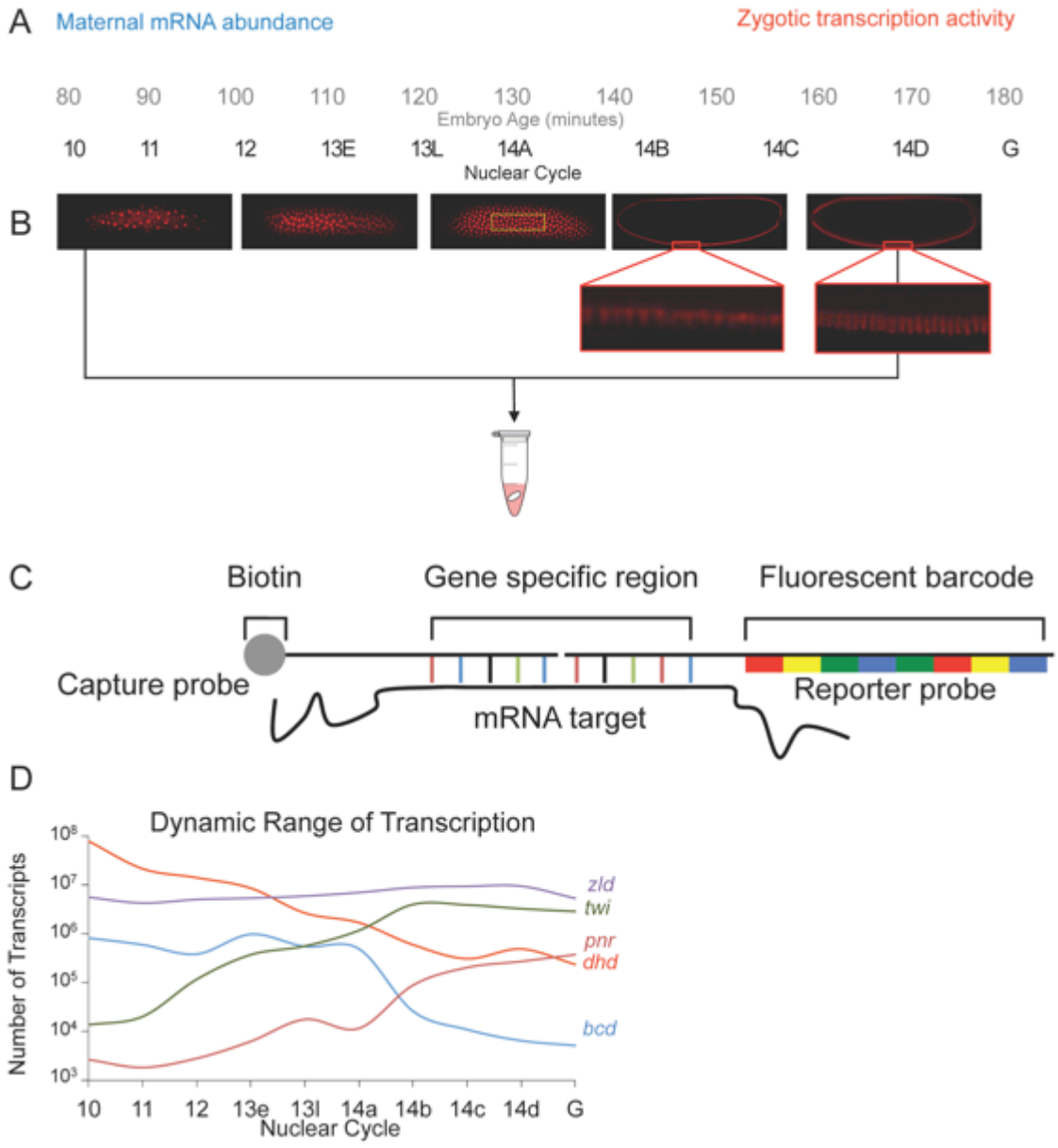


Figure 1.2. Timeline of Embryonic Development and Dynamic Range of Gene Expression (A) A timeline of early embryonic development in *Drosophila*. Maternal transcripts deposited during oogenesis are degraded while zygotic transcripts increase in abundance as the genome is first activated. The embryo age in minutes after egg laying (grey text) and corresponding nuclear cycle (black text) are aligned. (B) Nuclear density increases as nuclear cycles progress until nuclear cycle

14, then nuclei elongate until gastrulation, as shown in expanded images. Microfuge tube: individual embryos were selected at specific time points for analysis and to create a developmental time course. (C) A representation of NanoString probes hybridized to a target mRNA molecule. Both probes anneal to contiguous 50bp regions of the mRNA molecule. The reporter probe contains a target-specific fluorescent barcode, and the capture probe is conjugated to biotin for binding to a streptavidin-coated imaging cartridge. (D) mRNA abundance is highly variable and dynamic between different genes at the same time point and for the same gene at different time points. Different gene counts can vary by over four orders of magnitude simultaneously (*dhd* vs. *pnr* NC 10) or the same gene can vary by over 200-fold in around an hour (*dhd, twi,* and *pnr*).

In particular, it was found that not all time points during early embryonic development are equal in terms of changing gene expression. While maternal genes are constantly being degraded and zygotic genes are constantly being expressed during the blastoderm stage (Figure 1.2 D), the average fold-change in expression between various time points can differ greatly. Both the greatest increase in transcription and decrease in abundance occur during the first part of NC 14 (i.e. the transition from NC 14A to 14B). In fact, the rapid increase in transcription seen at this stage is over four times higher than the increase later in NC 14 (i.e. between NC 14C and 14D) less than 30 minutes later. This drastic difference may relate to Dorsal transcription factor dynamics. Prior to NC 14, nuclei divide too rapidly to allow Dorsal to build to high levels. Also, some active transcription may be aborted at every division due to the limited time available (SHERMOEN AND O'FARRELL 1991; O'FARRELL 1992; LEE *et al.* 2014).

This transition at the beginning of NC 14 is the first time in development that both Dorsal nuclear import and transcription can proceed uninterrupted for over 15 minutes. There are also more zygotic transcription factors present at the start of NC 14 as the result of their transcription and translation into functioning proteins during the previous nuclear cycles. These factors combine to make the short time period of around 15 minutes the most transcriptionally active during the blastoderm stage. By mid NC 14 (i.e. NC 14C), many

genes have reached a steady state of abundance, and while there are more transcripts present than 30 minutes before during the period of rapid transcription, the overall change is the lowest of any time point studied. This steady state and period of relatively little change occurs just after Dorsal reaches its own maximum concentration in nuclei and ceases increasing. It is not coincidental, therefore, that the expression rate of genes that rely so closely on Dorsal match the nuclear concentration dynamics of Dorsal itself.

Another benefit of the fine time scale quantitative profile provided by NanoString experiments is the ability to observe and dissect sub-circuits within the overall developmental Gene Regulatory Network (GRN). One of the most common sub-circuits found in GRN topologies is the feed forward loop, where an initial activator works cooperatively with one of its own targets to further activate more genes (DAVIDSON *et al.* 2002; MANGAN AND ALON 2003). A key property of feed forward loops is that the activating effects of individual components are additive or synergistic, and that each input alone is unable to activate target genes at full strength (DE-LEON AND DAVIDSON 2009). An example of a feed forward loop in the *Drosophila* developmental GRN is found in the mesoderm, where Dorsal first activates Twist, and then Dorsal and Twist together activate many other mesoderm genes (Figure 1.3 A) (IP *et al.* 1992; STATHOPOULOS AND LEVINE 2005). Since Twist has been shown to also activate mesoderm genes in the *Drosophila* embryo, it is a prime candidate for investigation and use in dissecting such network circuitry (KOSMAN *et al.* 1991; SANDMANN *et al.* 2007; SEHER *et al.* 2007; OZDEMIR *et al.* 2011).

The additive nature of feed forward loops can be observed by comparing the dynamics of Dorsal-Twist cooperative activation in wild type embryos to the activating

ability of Dorsal alone in *twi-* flies. Using NanoString, it can be observed that during the blastoderm stage between NCs 10 and 13, the expression of mesoderm genes slowly and steadily increases at every nuclear cycle, but then undergoes a very rapid increase starting at NC 14 until a steady state in transcript levels is reached. This bimodal profile may relate to temporal increases in Dorsal levels and/or to the additive effect of a second factor joining a feed forward loop (Figure 1.3 B).

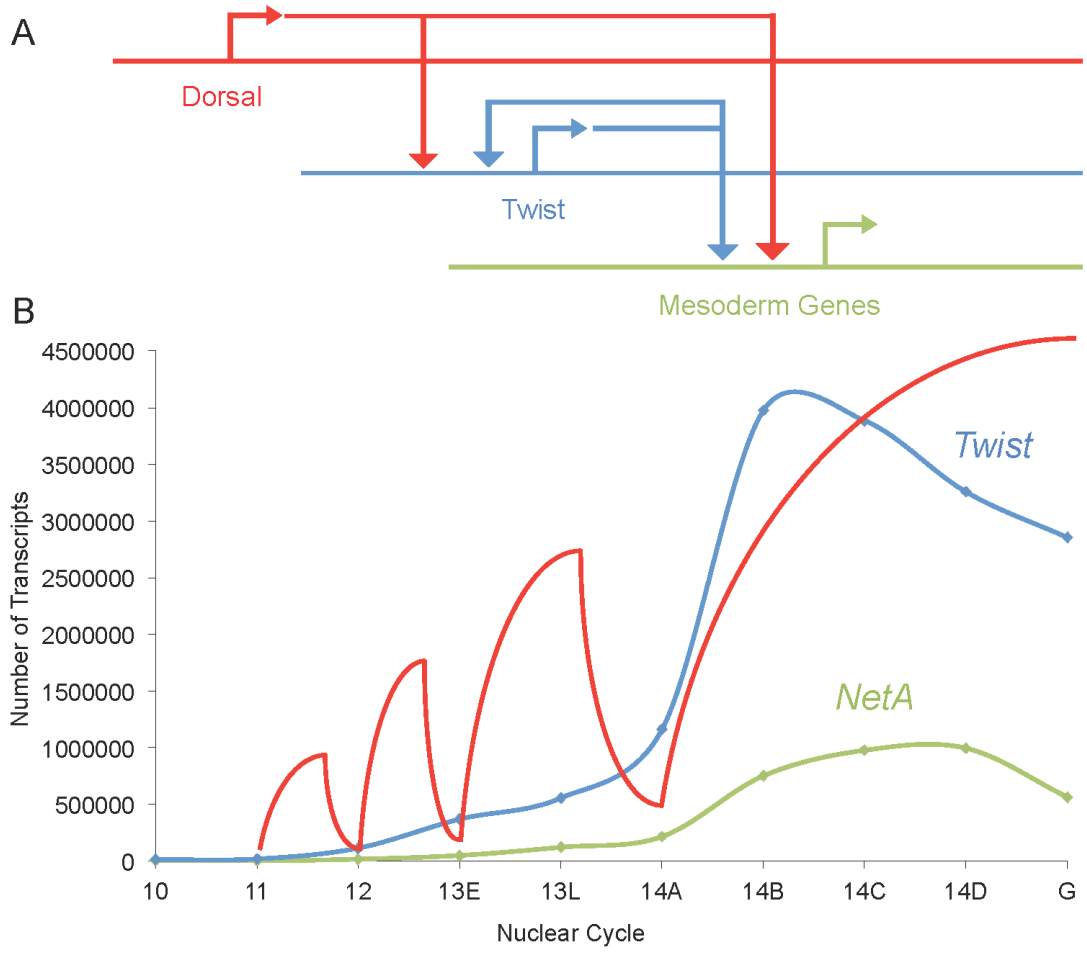


Figure 1.3. Dynamics of the Maternal to Zygotic Transition: Dorsal/Twist Feed Forward Loop. (A) A schematic of the early gene regulatory network architecture of the mesoderm showing the feed forward loop between Dorsal, Twist, and the rest of the mesoderm genes. Length of line for each network component corresponds to the nuclear cycle of activation and detectable presence on the plot below. (B) Transcriptional activity of *twi* (blue) and downstream mesoderm gene *NetA* (green) overlaid with a representation of the Dorsal concentration (red) in ventral nuclei.

When *twi* is mutated so it can no longer bind to DNA and mutant embryos are assayed by NanoString, the rapid increase in transcription usually observed in NC 14 does not occur, and the slower rate of transcription observed in NCs 10-13 is maintained (SANDLER AND STATHOPOULOS 2016). This difference in transcription rate demonstrates the additive nature of feed forward loops; at NC 14, Dorsal alone is able to activate its targets at a moderate level, but the input of Twist is able to provide an additional boost transcription that is added to the input of Dorsal to support high level expression. While a role for Twist in supporting expression of genes in the early embryo has been appreciated, using the NanoString to quantify levels of expression in individual, staged embryos illuminated the temporal role for Twist in supporting expression of genes, specifically, at NC 14 (SANDMANN *et al.* 2007).

Possible roles for step-wise progression of embryonic gene expression programs

Moving forward, an important goal in the field is to understand the role of dynamics of gene expression in supporting proper embryonic development (MANU *et al.* 2009; LOTT *et al.* 2011; RUSHLOW AND SHVARTSMAN 2012; ALI-MURTHY *et al.* 2013; WU *et al.* 2015). The recent quantitative analysis of gene expression in *Drosophila* embryos has highlighted activation of genes expressed along the DV axis occurs in a step-wise manner (REEVES *et al.* 2012; SANDLER AND STATHOPOULOS 2016). We contend this step-wise activation program is instrumental for DV patterning and suggest three ideas regarding its roles, below.

Activation of signaling pathways

Cell-cell signaling is not thought to broadly impact DV patterning until cellularization at the 14th nuclear cycle, when cells form, as before this point the embryo develops as a syncytium in which nuclei are not separated from each other by cell membranes (STATHOPOULOS AND LEVINE 2005; PERRIMON *et al.* 2012). It is, presumably, for this reason that genes requiring input from Notch or EGFR signaling such as *single-minded* (*sim*) and *intermediate neuroblasts defective* (*ind*), respectively, exhibit delayed expression that coincides with cellularization (MOREL AND SCHWEISGUTH 2000; LIM *et al.* 2013). However, recent studies have found that nuclei become compartmentalized before cellularization is complete (MAVRAKIS *et al.* 2009), suggesting that cell-cell signaling may be possible earlier.

The progressive activation of the DV patterning GRN in the early *Drosophila* embryo may promote activation of signaling pathways in a step-wise manner. It is appreciated that subdivision of the embryo into distinct domains of expression, through patterning, is necessary to set-up activation of signaling pathways through differential expression of receptors and ligands. However, findings that signaling pathway components are expressed before NC14, some as early as NC10, suggest that activation of signaling may occur as a step-wise progression influenced by the gene network program to impact activation and/or levels of signaling. Studies in other systems have provided evidence that "fold-change" may trigger signaling activation rather than a particular threshold level of ligand; arguing that step-wise activation of signaling may be important (GOENTORO *et al.* 2009; SHOVAL *et al.* 2010). Furthermore, in such a system, the temporal presentation of ligands may be more influential than absolute levels in

supporting signaling pathway activation, supporting the recent view that concentration-dependence is not pivotal to the action of morphogens (BRISCOE AND SMALL 2015; WOLPERT 2016).

Control of gene expression cohorts

Another finding from the NanoString study is that, while early embryogenesis is a dynamic time in general, there are stages of rapid coordinated changes in gene expression. For example, a gene cohort of Dorsal targets expressed in the mesoderm exhibit a gradual increase in abundance between NC 10 and 13, but then all exhibit a rapid and coordinated increase in transcription rate as NC14 begins (Figure 1.4 A). This coordinated increase occurs at the same time for all six mesoderm genes included in the NanoString study (*twi, mes3, sna, hbr, NetA,* and *htl)*, and coincides with the time of high dynamic change between NCs 14A and 14B. In contrast, target genes of Bicoid expressed along the anterior-posterior (AP) axis, such as *hb* and *otd*, show no signs of a coordinated increase in expression between NCs 14A and 14B, or any other time point (SANDLER AND STATHOPOULOS 2016). The AP targets of Bicoid increase gradually during the time course without a rapid change in expression strength. This is likely due to the relatively stable levels of Bicoid found along the AP axis during early embryogenesis.

A second group of six genes expressed in the dorsal ectoderm as targets of the TGF-β pathway (ASHE *et al.* 2000) behaved in a somewhat different way compared to the mesoderm genes (Figure 1.4 B). Like the group of mesoderm genes, the transcription of all six TGF-β target genes is also coordinated temporally. Unlike the mesoderm genes that all behave similarly, two classes of TGF-β target genes were uncovered based on different

kinetics of expression at the onset of NC14. One set exhibited slow and steady transcription whereas the other exhibited rapid expression. Despite these differences between mesoderm genes and TGF-β targets, the temporal coregulation of different groups of genes reinforces the idea that coordinated and precise timing of transcription is a key feature of the early GRN and has been observed in other systems (DUBRULLE *et al.* 2015).

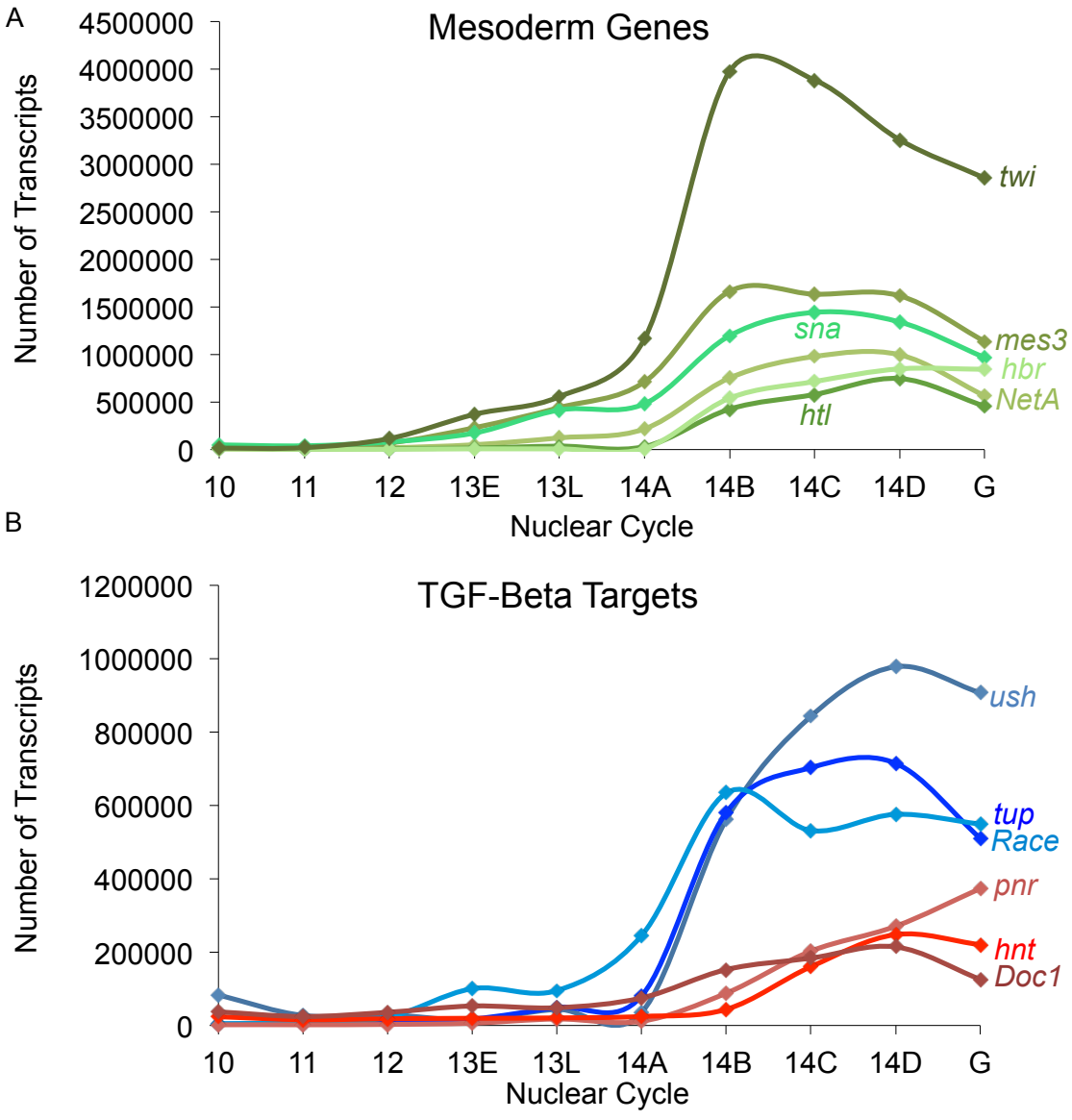


Figure 1.4. Timing of Gene Cohort Expression. (A) Genes expressed in the mesoderm are regulated as a cohort, with two coordinated phases of activation. Early coordinated activation begins

at NC 12, while robust coordinated activation begins at NC 14A for all genes examined (*mes3*, *sna*, *NetA, hbr, htl*) except for *twi* that is upregulated much faster and likely serves as input to other mesodermally-expressed genes together with Dorsal. (B) A cohort of genes expressed at NC 14A as targets of the TGF-β pathway in the dorsal ectoderm. The genes are temporally co-regulated, but diverge in their transcription rates. Genes *ush, tup,* and *Race* are transcribed quickly and reach a steady state, while genes *pnr, hnt,* and *Doc1* are transcribed moderately.

Furthermore, identification of additional gene expression cohorts such as these will facilitate approaches aimed to identify shared regulatory motifs in enhancers and promoters that support shared dynamics.

An additional difference is uncovered between the mesoderm and TGF-β target genes when the number of transcripts per cell is calculated instead of overall number of transcripts per embryo. When the overall number of transcripts for each group is divided by the number of cells expressing each gene, the mesoderm genes are maintained in a rank-order of abundance through the entire time course, while TGF-β target genes are expressed in a very similar number of transcripts per cell. A possible explanation for the persistent differences in expression per cell for mesoderm genes compared to the similar levels of expression for TGF-β targets is their position in the GRN. The mesoderm genes are some of the first zygotic genes to be activated in the network, while the TGF-β target genes are at the output level of a signaling pathway at the end of the pre-gastrulation network. It may be important to maintain different levels of gene expression early in developmental pathways in order to activate or repress targets in varying ways, while genes at the output level of signaling pathways are programmed to be expressed in similar levels to each other as the signaling pathway integrates changing inputs into a stable output.

Irreversibility of the embryonic patterning process

Another factor contributing to the irreversible nature of the step-wise activation of the GRN is syncytial nuclear division leading to increasingly stronger pulses of nuclear Dorsal. Beginning at NC 10, when nuclei migrate to the periphery of the embryo, nuclei in the ventral portions of the embryo are exposed to the highest concentrations of Dorsal and begin transcribing early mesoderm-determining transcription factors such as *twist* (*twi*) and *sna* (ALBERGA *et al.* 1991; KOSMAN *et al.* 1991; LEPTIN 1991). Although the first few nuclear cycles during the syncytial blastoderm stage are brief, around 10 minutes each, the short length of many early transcription factors allows them to be fully transcribed before nuclear division aborts active transcription. This brief pulse of transcription supplies mature transcripts to allow for the translation of full-length and functional proteins, able to either activate additional mesoderm genes in the case of *twi* or repress the expression of neurogenic ectoderm genes in the case of *sna*. The first active transcription factors set into motion cascades of activation and repression, with each subsequent nuclear cycle being accompanied by higher concentrations of Dorsal leading to the presence of even more early transcription factor gene products.

Each nuclear cycle can be thought of as a developmental step, leading nuclei or cells down a one-way trajectory towards their ultimate fate. GRN activation is a natural consequence of early Dorsal-mediated expression of the first transcription factors. The rapid nature of syncytial nuclear divisions combined with ever-increasing concentrations of Dorsal ensures that regulatory states established early in development are robustly transmitted and engrained in nuclei during subsequent nuclear divisions. By the time cellularization occurs in the middle of NC 14, the previous rounds of nuclear division

have set up a situation where the cells have no choice but to follow the path laid out for them, and the rapid onset of intracellular signaling pathways only serves to further cement these fates. On the other hand, Bicoid levels do not change as dramatically. It is possible that Bicoid is required only early in the AP patterning GRN, to set a chain of events in motion that relies more heavily on duration of Bicoid signal than absolute concentration in nuclei. Dorsal may remain continuously necessary as its concentration increases, supporting early patterning as well as late patterning events, up to when its levels plummet at gastrulation.

Key challenges in studying spatiotemporal regulation of gene expression programs

As discussed above, dynamic gene expression likely relates to proper timing of signaling pathway activation and also the step-wise progression of gene expression programs helps support irreversibility of the process. However, we have only scratched the surface in understanding how dynamic gene expression within this gene network is controlled at a mechanistic level. Below we comment on three areas of future research that will provide insight and better understanding.

Roles of additional transcription factors, both ubiquitous and spatially localized, in controlling temporal gene expression programs

It is clear that the ubiquitous zinc-finger transcription factor Zelda is very important for supporting early zygotic expression in the early *Drosophila* embryo (LIANG *et al.* 2008; NIEN *et al.* 2011). However, other factors also contribute to timing of gene expression. STAT92E, another ubiquitous factor, broadly influences early zygotic

transcription, as does the Grainy head transcriptional activator, exemplified by its support of *ind* expression (HARRISON *et al.* 2010; GARCIA AND STATHOPOULOS 2011; TSURUMI *et al.* 2011). It is likely that a number of ubiquitous activators including Zelda, STAT92E, and Grainy head impact patterning and that these factors may exhibit different timing of action. Also, as discussed above for Dorsal, transcription factors known primarily for their roles in supporting spatial patterning may also regulate timing of gene expression. Lastly, globally-acting repressors likely function to counterbalance this activation, to regulate spatial (OZDEMIR *et al.* 2014) as well as temporal expression. Understanding how these factors, ubiquitous or spatially-localized, collectively influence timing of gene expression programs is an important area of future research. Additionally, synthetic reporter constructs combining transcription factor binding sites [e.g. Dorsal, Zelda, and the early transcriptional repressor Suppressor of Hairless, Su(H)] have begun to examine the relationship between number and organization of binding sites, 'cis-regulatory logic' or 'grammar', to spatial regulation of expression (JIANG AND LEVINE 1993; LIBERMAN AND STATHOPOULOS 2009; OZDEMIR *et al.* 2014). Another promising future direction is to study how combinatorial control and organization of sites relates to timing and levels of gene expression (ERIVES AND LEVINE 2004; CROCKER *et al.* 2008; FARLEY *et al.* 2015).

Coordinate action of cis-regulatory modules and role in supporting gene expression dynamics

Transcription factors, for the most part, act on cis-regulatory modules (CRMs), and to understand how timing of gene expression is regulated a better understanding of

how CRMs cooperate to support gene expression must be acquired. Recent studies have found that multiple cis-regulatory modules are often associated with genes and are co-acting (HONG *et al.* 2008; PERRY *et al.* 2009; BAROLO 2012; STALLER *et al.* 2015). Some CRMs work concurrently to control spatial domains and levels of expression (PERRY *et al.* 2010; DUNIPACE *et al.* 2011), whereas others work sequentially to control the changing expression of genes in time (DUNIPACE *et al.* 2013). These insights lead directly to the question of how multiple CRMs coordinate in space and time. Recent studies have identified autoregulatory feedback as the mechanism regulating the switch from an early-acting CRM to a later-acting CRM at the *brinker* gene locus, which regulates the spatiotemporal expression of this gene (DUNIPACE *et al.* 2013). Therefore, understanding of both (i) spatiotemporal inputs (i.e. transcription factor dynamics) as well as (ii) CRMs acting and their coordinate action is required to understand how temporal gene expression is controlled.

Role of post-transcriptional regulation in temporal gene expression and gene functions

In an analysis of spatiotemporal profiles for the genes *sog* and *sna*, evidence was obtained that *sog* transcripts are degraded at the transition from NC 13 to NC 14 whereas *sna* transcripts are retained (REEVES *et al.* 2012). A simple explanation is that the short timeframe of NC 13, under 15 minutes, is not long enough to support transcription of long genes and therefore nascent transcripts that do not reach maturity are degraded (SHERMOEN AND O'FARRELL 1991; LEE *et al.* 2014). However, an alternate possibility (not mutually exclusive) is that post-transcriptional mechanisms influence the abundance and stability of zygotic transcripts present in the embryo (KUERSTEN AND GOODWIN

2003; SEMOTOK AND LIPSHITZ 2007; LASKO 2011). An exciting future direction would be to uncover how post-transcriptional regulation factors into the timing of developmental progression and, specifically, to uncover how it influences DV patterning and signaling pathway activation.

Concluding remarks

In summary, recent studies highlight the need to consider the dorsal-ventral gene regulatory network as a step-wise process in which the status of the system (i.e. gene expression) is assayed with fine temporal resolution. Use of the NanoString technology has supported generation of a time-series from carefully staged, individual *Drosophila* embryo fixed samples (SANDLER AND STATHOPOULOS 2016). From these data, dynamic trends within gene regulatory networks can be inferred such as identification of gene expression cohorts and specific, temporal roles for transcription factors (DUBRULLE *et al.* 2015). Furthermore, imaging transcripts directly and dynamically in living embryos over time is a complementary approach that is also able to assay dynamics of nascent transcripts associated with a single gene (GARCIA *et al.* 2013; LUCAS *et al.* 2013; BOTHMA *et al.* 2014). Identifying technologies that make it possible to assay expression levels for tens of genes *in vivo* live would be an exciting future frontier (DEAN AND PALMER 2014). The ultimate goal is to attain a complete understanding of the control of spatiotemporal gene expression, how it results from the action of transcription factors on one or more cis-regulatory modules. Model organisms are an excellent choice for such system level analyses aimed at deciphering regulatory logic that can help us better understand GRNs acting in humans. As more GRN studies emerge, it is becoming clear

that a common set of subcircuit designs is used (ALON 2007; DAVIDSON 2010). Additional trends, or even differences, may emerge from more comparative studies.

References

Alberga, A., J. L. Boulay, E. Kempe, C. Dennefeld and M. Haenlin, 1991 The snail gene required for mesoderm formation in Drosophila is expressed dynamically in derivatives of all three germ layers. Development 111**:** 983-992.

Ali-Murthy, Z., S. E. Lott, M. B. Eisen and T. B. Kornberg, 2013 An Essential Role for Zygotic Expression in the Pre-Cellular Drosophila Embryo. Plos Genetics 9.

Alon, U., 2007 Network motifs: theory and experimental approaches. Nat Rev Genet 8**:** 450-461.

Anderson, K. V., L. Bokla and C. Nusslein-Volhard, 1985 Establishment of dorsal-ventral polarity in the Drosophila embryo: the induction of polarity by the Toll gene product. Cell 42**:** 791-798.

Anderson, K. V., and C. Nusslein-Volhard, 1984 Information for the dorsal--ventral pattern of the Drosophila embryo is stored as maternal mRNA. Nature 311**:** 223-227.

Ashe, H. L., M. Mannervik and M. Levine, 2000 Dpp signaling thresholds in the dorsal ectoderm of the Drosophila embryo. Development (Cambridge) 127**:** 3305-3312.

Barolo, S., 2012 Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. Bioessays 34**:** 135-141.

Belvin, M. P., and K. V. Anderson, 1996 A conserved signaling pathway: the Drosophila toll-dorsal pathway. Annu Rev Cell Dev Biol 12**:** 393-416.

Biemar, F., D. A. Nix, J. Piel, B. Peterson, M. Ronshaugen *et al.*, 2006 Comprehensive identification of Drosophila dorsal-ventral patterning genes using a whole-genome tiling array. Proc Natl Acad Sci U S A 103**:** 12763-12768.

Bothma, J. P., H. G. Garcia, E. Esposito, G. Schlissel, T. Gregor *et al.*, 2014 Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos. Proc Natl Acad Sci U S A 111**:** 10598-10603.

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

Carmena, A., S. Gisselbrecht, J. Harrison, F. Jimenez and A. M. Michelson, 1998 Combinatorial signaling codes for the progressive determination of cell fates in the Drosophila embryonic mesoderm. Genes & Development 12**:** 3910-3922.

Chen, H., Z. Xu, C. Mei, D. Yu and S. Small, 2012 A system of repressor gradients spatially organizes the boundaries of Bicoid-dependent target genes. Cell 149**:** 618-629.

Chopra, V. S., and M. Levine, 2009 Combinatorial patterning mechanisms in the Drosophila embryo. Brief Funct Genomic Proteomic 8**:** 243-249.

Crocker, J., Y. Tamori and A. Erives, 2008 Evolution acts on enhancer organization to fine-tune gradient threshold readouts. PLoS Biol 6**:** e263.

Davidson, E. H., 2010 Emerging properties of animal gene regulatory networks. Nature 468**:** 911-920.

Davidson, E. H., and M. S. Levine, 2008 Properties of developmental gene regulatory networks. Proc Natl Acad Sci U S A 105**:** 20063-20066.

Davidson, E. H., J. P. Rast, P. Oliveri, A. Ransick, C. Calestani *et al.*, 2002 A genomic regulatory network for development. Science 295**:** 1669-1678.

de-Leon, S. B. T., and E. H. Davidson, 2009 Modeling the dynamics of transcriptional gene regulatory networks for animal development. Developmental Biology 325**:** 317-328.

Dean, K. M., and A. E. Palmer, 2014 Advances in fluorescence labeling strategies for dynamic cellular imaging. Nat Chem Biol 10**:** 512-523.

Dubrulle, J., B. M. Jordan, L. Akhmetova, J. A. Farrell, S. H. Kim *et al.*, 2015 Response to Nodal morphogen gradient is determined by the kinetics of target gene induction. Elife 4.

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

Dunipace, L., A. Saunders, H. L. Ashe and A. Stathopoulos, 2013 Autoregulatory Feedback Controls Sequential Action of cis-Regulatory Modules at the brinker Locus. Dev Cell 26**:** 536-543.

Erives, A., and M. Levine, 2004 Coordinate enhancers share common organizational features in the Drosophila genome. Proc Natl Acad Sci U S A 101**:** 3851-3856.

Farley, E. K., K. M. Olson, W. Zhang, A. J. Brandt, D. S. Rokhsar *et al.*, 2015 Suboptimization of developmental enhancers. Science 350**:** 325-328.

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

Garcia, H. G., M. Tikhonov, A. Lin and T. Gregor, 2013 Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. Curr Biol 23**:** 2140-2145.

Garcia, M., and A. Stathopoulos, 2011 Lateral gene expression in Drosophila early embryos is supported by Grainyhead-mediated activation and tiers of dorsally-localized repression. PLoS One 6**:** e29172.

Geiss, G. K., R. E. Bumgarner, B. Birditt, T. Dahl, N. Dowidar *et al.*, 2008 Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol 26**:** 317-325.

Goentoro, L., O. Shoval, M. W. Kirschner and U. Alon, 2009 The Incoherent Feedforward Loop Can Provide Fold-Change Detection in Gene Regulation. Molecular Cell 36**:** 894-899.

Graveley, B. R., A. N. Brooks, J. W. Carlson, M. O. Duff, J. M. Landolin *et al.*, 2011 The developmental transcriptome of Drosophila melanogaster. Nature 471**:** 473-479.

Gregor, T., E. F. Wieschaus, A. P. McGregor, W. Bialek and D. W. Tank, 2007 Stability and nuclear dynamics of the bicoid morphogen gradient. Cell 130**:** 141-152.

Gurdon, J. B., S. Dyson and D. S. Johnston, 1998 Cells' perception of position in a concentration gradient. Cell 95**:** 159-162.

Harrison, M. M., M. R. Botchan and T. W. Cline, 2010 Grainyhead and Zelda compete for binding to the promoters of the earliest-expressed Drosophila genes. Dev Biol 345**:** 248-255.

Hong, J.-W., D. A. Hendrix and M. S. Levine, 2008 Shadow Enhancers as a Source of Evolutionary Novelty. Science 321**:** 1314-.

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

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

Jiang, J., C. A. Rushlow, Q. Zhou, S. Small and M. Levine, 1992 Individual dorsal morphogen binding sites mediate activation and repression in the Drosophila embryo. EMBO J 11**:** 3147-3154.

Kanodia, J. S., R. Rikhy, Y. Kim, V. K. Lund, R. DeLotto *et al.*, 2009 Dynamics of the Dorsal morphogen gradient. Proc Natl Acad Sci U S A 106**:** 21707-21712.

Kosman, D., Y. T. Ip, M. Levine and K. Arora, 1991 Establishment of the mesoderm-neuroectoderm boundary in the Drosophila embryo. Science 254**:** 118-122.

Kuersten, S., and E. B. Goodwin, 2003 The power of the 3' UTR: translational control and development. Nat Rev Genet 4**:** 626-637.

Lasko, P., 2011 Posttranscriptional regulation in Drosophila oocytes and early embryos. Wiley Interdiscip Rev RNA 2**:** 408-416.

Lee, J. Y., D. J. Marston, T. Walston, J. Hardin, A. Halberstadt *et al.*, 2006 Wnt/Frizzled signaling controls C-elegans gastrulation by activating actomyosin contractility. Current Biology 16**:** 1986-1997.

Lee, M. T., A. R. Bonneau and A. J. Giraldez, 2014 Zygotic genome activation during the maternal-to-zygotic transition. Annu Rev Cell Dev Biol 30**:** 581-613.

Leptin, M., 1991 twist and snail as positive and negative regulators during Drosophila mesoderm development. Genes Dev 5**:** 1568-1576.

Levine, J. H., Y. Lin and M. B. Elowitz, 2013 Functional roles of pulsing in genetic circuits. Science 342**:** 1193-1200.

Liang, H. L., C. Y. Nien, H. Y. Liu, M. M. Metzstein, N. Kirov *et al.*, 2008 The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature 456**:** 400-403.

Liberman, L. M., G. T. Reeves and A. Stathopoulos, 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, L. M., and A. Stathopoulos, 2009 Design flexibility in cis-regulatory control of gene expression: synthetic and comparative evidence. Dev Biol 327**:** 578-589.

Lim, B., N. Samper, H. Lu, C. Rushlow, G. Jimenez *et al.*, 2013 Kinetics of gene derepression by ERK signaling. Proc Natl Acad Sci U S A 110**:** 10330-10335.

Lott, S. E., J. E. Villalta, G. P. Schroth, S. Luo, L. A. Tonkin *et al.*, 2011 Noncanonical compensation of zygotic X transcription in early Drosophila melanogaster development revealed through single-embryo RNA-seq. PLoS Biol 9**:** e1000590.

Lucas, T., T. Ferraro, B. Roelens, J. De Las Heras Chanes, A. M. Walczak *et al.*, 2013 Live imaging of bicoid-dependent transcription in Drosophila embryos. Curr Biol 23**:** 2135-2139.

MacArthur, S., X. Y. Li, J. Li, J. B. Brown, H. C. Chu *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 Biol 10**:** R80.

Mangan, S., and U. Alon, 2003 Structure and function of the feed-forward loop network motif. Proceedings of the National Academy of Sciences of the United States of America 100**:** 11980-11985.

Manu, S. Surkova, A. V. Spirov, V. V. Gursky, H. Janssens *et al.*, 2009 Canalization of gene expression and domain shifts in the Drosophila blastoderm by dynamical attractors. PLoS Comput Biol 5**:** e1000303.

Mavrakis, M., R. Rikhy and J. Lippincott-Schwartz, 2009 Cells within a cell: Insights into cellular architecture and polarization from the organization of the early fly embryo. Commun Integr Biol 2**:** 313-314.

Morel, V., and F. Schweisguth, 2000 Repression by suppressor of hairless and activation by Notch are required to define a single row of single-minded expressing cells in the Drosophila embryo. Genes Dev 14**:** 377-388.

Moussian, B., and S. Roth, 2005 Dorsoventral axis formation in the Drosophila embryo--shaping and transducing a morphogen gradient. Curr Biol 15**:** R887-899.

Nien, C. Y., H. L. Liang, S. Butcher, Y. Sun, S. Fu *et al.*, 2011 Temporal coordination of gene networks by Zelda in the early Drosophila embryo. PLoS Genet 7**:** e1002339.

O'Farrell, P. H., 1992 Developmental biology. Big genes and little genes and deadlines for transcription. Nature 359**:** 366-367.

Ozdemir, A., K. I. Fisher-Aylor, S. Pepke, M. Samanta, L. Dunipace *et al.*, 2011 High resolution mapping of Twist to DNA in Drosophila embryos: Efficient functional analysis and evolutionary conservation. Genome Res 21**:** 566-577.

Ozdemir, A., L. Ma, K. P. White and A. Stathopoulos, 2014 Su(H)-Mediated Repression Positions Gene Boundaries along the Dorsal-Ventral Axis of Drosophila Embryos. Dev Cell 31**:** 100-113.

Perrimon, N., C. Pitsouli and B. Z. Shilo, 2012 Signaling mechanisms controlling cell fate and embryonic patterning. Cold Spring Harb Perspect Biol 4**:** a005975.

Perry, M. W., A. N. Boettiger, J. P. Bothma and M. Levine, 2010 Shadow enhancers foster robustness of Drosophila gastrulation. Curr Biol 20**:** 1562-1567.

Perry, M. W., J. D. Cande, A. N. Boettiger and M. Levine, 2009 Evolution of insect dorsoventral patterning mechanisms. Cold Spring Harb Symp Quant Biol 74**:** 275-279.

Purvis, J. E., and G. Lahav, 2013 Encoding and decoding cellular information through signaling dynamics. Cell 152**:** 945-956.

Ray, R. P., K. Arora, C. Nusslein-Volhard and W. M. Gelbart, 1991 The control of cell fate along the dorsal-ventral axis of the Drosophila embryo. Development 113**:** 35-54.

Reeves, G. T., and A. Stathopoulos, 2009 Graded Dorsal and Differential Gene Regulation in the *Drosophila* Embryo in *Perspectives on Generation and Interpretation of Morphogen Gradients*, edited by J. Briscoe, P. Lawrence and J.-P. Vincent. Cold Spring Harbor Laboratory Press.

Reeves, G. T., N. Trisnadi, T. V. Truong, M. Nahmad, S. Katz *et al.*, 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 A. F. Schier, 2011 Morphogen gradients: from generation to interpretation. Annu Rev Cell Dev Biol 27**:** 377-407.

Roth, S., D. Stein and C. Nussleinvolhard, 1989 A Gradient of Nuclear-Localization of the Dorsal Protein Determines Dorsoventral Pattern in the Drosophila Embryo. Cell 59**:** 1189-1202.

Rushlow, C. A., K. Han, J. L. Manley and M. Levine, 1989 The graded distribution of the dorsal morphogen is initiated by selective nuclear transport in Drosophila. Cell 59**:** 1165-1177.

Rushlow, C. A., and S. Y. Shvartsman, 2012 Temporal dynamics, spatial range, and transcriptional interpretation of the Dorsal morphogen gradient. Curr Opin Genet Dev 22**:** 542-546.

Sandler, J. E., and A. Stathopoulos, 2016 Quantitative Single-Embryo Profile of Drosophila Genome Activation and the Dorsal-Ventral Patterning Network. Genetics.

Sandmann, T., C. Girardot, M. Brehme, W. Tongprasit, V. Stolc *et al.*, 2007 A core transcriptional network for early mesoderm development in Drosophila melanogaster. Genes Dev 21**:** 436-449.

Seher, T. C., M. Narasimha, E. Vogelsang and M. Leptin, 2007 Analysis and reconstitution of the genetic cascade controlling early mesoderm morphogenesis in the Drosophila embryo. Mech Dev 124**:** 167-179.

Semotok, J. L., and H. D. Lipshitz, 2007 Regulation and function of maternal mRNA destabilization during early Drosophila development. Differentiation 75**:** 482-506.

Shermoen, A. W., and P. H. O'Farrell, 1991 Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. Cell 67**:** 303-310.

Shoval, O., L. Goentoro, Y. Hart, A. Mayo, E. Sontag *et al.*, 2010 Fold-change detection and scalar symmetry of sensory input fields. Proceedings of the National Academy of Sciences of the United States of America 107**:** 15995-16000.

Shvartsman, S. Y., M. Coppey and A. M. Berezhkovskii, 2008 Dynamics of maternal morphogen gradients in Drosophila. Curr Opin Genet Dev 18**:** 342-347.

Simpson, P., 1983 Maternal-zygotic gene interactions during formation of the dorsoventral pattern in Drosophila embryos. Genetics 105**:** 615-632.

Staller, M. V., B. J. Vincent, M. D. Bragdon, T. Lydiard-Martin, Z. Wunderlich *et al.*, 2015 Shadow enhancers enable Hunchback bifunctionality in the Drosophila embryo. Proc Natl Acad Sci U S A 112**:** 785-790.

Stathopoulos, A., and M. Levine, 2002 Dorsal gradient networks in the Drosophila embryo. Dev Biol 246**:** 57-67.

Stathopoulos, A., and M. Levine, 2004 Whole-genome analysis of Drosophila gastrulation. Curr Opin Genet Dev 14**:** 477-484.

Stathopoulos, A., and M. Levine, 2005 Genomic regulatory networks and animal development. Developmental Cell 9**:** 449-462.

Stein, D. S., and L. M. Stevens, 2014 Maternal control of the Drosophila dorsal-ventral body axis. Wiley Interdiscip Rev Dev Biol 3**:** 301-330.

Steward, R., 1989 Relocalization of the Dorsal Protein from the Cytoplasm to the Nucleus Correlates with Its Function. Cell 59**:** 1179-1188.

Tsurumi, A., F. Xia, J. Li, K. Larson, R. LaFrance *et al.*, 2011 STAT is an essential activator of the zygotic genome in the early Drosophila embryo. PLoS Genet 7**:** e1002086.

Wolpert, L., 2016 Positional Information and Pattern Formation. Curr Top Dev Biol 117**:** 597-608.

Wu, H., Manu, R. Jiao and J. Ma, 2015 Temporal and spatial dynamics of scaling-specific features of a gene regulatory network in Drosophila. Nat Commun 6**:** 10031.

Zeitlinger, J., R. P. Zinzen, A. Stark, M. Kellis, H. Zhang *et al.*, 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.

Zinzen, R. P., C. Girardot, J. Gagneur, M. Braun and E. E. Furlong, 2009 Combinatorial binding predicts spatio-temporal cis-regulatory activity. Nature 462**:** 65-70.

Chapter 2

QUANTITATIVE SINGLE-EMBRYO PROFILE OF *DROSOPHILA* GENOME ACTIVATION AND THE DORSAL–VENTRAL PATTERNING NETWORK

ABSTRACT

During embryonic development of *Drosophila melanogaster,* the Maternal to Zygotic Transition (MZT) marks a significant and rapid turning point when zygotic transcription begins and control of development is transferred from maternally deposited transcripts. Characterizing the sequential activation of the genome during the MZT requires precise timing and a sensitive assay to measure changes in expression. We utilized the NanoString nCounter instrument, which directly counts mRNA transcripts without reverse transcription or amplification, to study over 70 genes expressed along the dorsal-ventral (DV) axis of early *Drosophila* embryos, dividing the MZT into 10 time points. Transcripts were quantified for every gene studied at all time points, providing the first data set of absolute numbers of transcripts during *Drosophila* development. We found that gene expression changes quickly during the MZT, with early Nuclear Cycle (NC) 14 the most dynamic time for the embryo. *twist* is one of the most abundant genes in the entire embryo and we use mutants to quantitatively demonstrate how it cooperates with Dorsal to activate transcription and is responsible for some of the rapid changes in transcription observed during early NC14. We also uncovered elements within the gene regulatory network that maintain precise transcript levels for sets of genes that are spatiotemporally co-transcribed within the presumptive mesoderm or dorsal ectoderm. Using this new data, we show that a fine-scale, quantitative analysis of temporal gene expression can provide new insights into

developmental biology by uncovering trends in gene networks, including coregulation of target genes and specific temporal input by transcription factors.

INTRODUCTION

The Maternal to Zygotic Transition (MZT) is a key step in animal embryonic development, when maternally deposited transcripts are degraded in the embryo, and the embryonic genome is first activated. In *Drosophila melanogaster,* the MZT takes place within the first three hours of development, during the late syncytial nuclear divisions and ending at the cellular blastoderm stage with gastrulation (FOE AND ALBERTS 1983; PRITCHARD AND SCHUBIGER 1996; TADROS AND LIPSHITZ 2009). Gene expression during the MZT is highly dynamic, with patterns of zygotic genes first being established and changing between and within nuclear cycles (STATHOPOULOS AND LEVINE 2005; REEVES *et al.* 2012). It is clear therefore that each syncytial nuclear cycle can be treated as a single, or even multiple developmental time points. A few recent RNA-seq based studies have in fact divided embryonic development into time points based on syncytial nuclear divisions for this very reason (LOTT *et al.* 2011; ALI-MURTHY *et al.* 2013). In previous studies, however, the syncytial nuclear stage, especially nuclear cycles 10-14, has been grouped together in a small number of developmental stages or time points (BOWNES 1975; BATE AND MARTINEZ ARIAS 1993; GRAVELEY *et al.* 2011). These pioneering studies provided the basis for studying embryonic development of *Drosophila,* and the modENCODE transcriptome provided a depth of sequencing data never before achieved for *Drosophila.* We choose, however, to focus on a fine time scale approach and fewer genes to provide a detailed analysis of a specific period in development (REEVES *et al.* 2012).

The top-level network inputs appear to be more dynamic on the DV axis than on the Anterior-Posterior (AP) axis. An activator of AP transcription is maternally deposited *bicoid*, which is transported to the anterior pole and forms a concentration gradient. The nuclear concentration of Bicoid during the final five nuclear cycles remains mostly constant during each nuclear cycle, indicating that Bicoid itself activates transcription of AP genes at a constant rate through these nuclear cycles (GREGOR *et al.* 2007). In contrast the protein product of the maternal gene *dorsal*, found in a DV gradient, increases in concentration within nuclei during each of the final five nuclear cycles (REEVES *et al.* 2012). This increase in nuclear Dorsal concentration suggests that the DV network is activated differently at each nuclear cycle, both by Dorsal itself, and by a network of transcription factors that respond to different levels of Dorsal. The combination of the rapidly changing transcriptional landscape during the MZT, the increasing nuclear concentration of Dorsal on the DV axis, and the small number of studies that have examined embryogenesis at the single nuclear cycle level present an opportunity to use emerging technologies to provide additional insight into this gene patterning network.

In this study, we examine the MZT and gene expression dynamics of the DV network at 10 time points during *Drosophila* embryonic development between NC10 and gastrulation at a 10-15 minute resolution. We utilize the NanoString nCounter instrument to directly detect and quantify 68 early embryonic genes from single embryos, and we calculate the absolute number of transcripts per embryo for every gene at every time point in the study (GEISS *et al.* 2008b). The NanoString system is able to precisely quantify transcripts across five orders of magnitude from a single embryo without the need to fragment, amplify, or reverse transcribe the RNA (GEISS *et al.* 2008a). The direct detection

of mRNA molecules minimizes steps between sample collection and data acquisition, reducing error, sample loss, or contamination. RNA-seq has been used in past studies of the *Drosophila* MZT to quantify the number of transcripts for a gene in the early embryo, and while these studies provide an abundance of data for all genes transcribed, the methods used have been shown to introduce bias in transcript count and read coverage that can hamper absolute quantification of transcripts (HANSEN *et al.* 2010; LOTT *et al.* 2011; ROBERTS *et al.* 2011; ALI-MURTHY *et al.* 2013; PETKOVA *et al.* 2014).

MATERIALS & METHODS

Fly stocks. Embryo collection and live imaging was done on flies with a His2Av-RFP fusion [Bloomington Drosophila Stock Center (BDSC) 23650]. *twist- (twi)* embryos were obtained using a twi^{1}/CyO stock (BDSC 2381). PCR for LacZ was done on all mutant embryos to confirm the absence of the balancer chromosome and the presence of homozygous *twi-* mutant chromosomes.

Live imaging and embryo collection. Flies with the His2Av-RFP fusion were allowed to lay eggs for four hours at 25°C. Individual embryos were hand de-chorionated and mounted on a microscope slide using a modified version of the hanging-drop method (REED *et al.* 2009). Nuclear divisions were monitored using epifluorescence, and confocal images of individual embryos were captured when embryos reached a desired developmental stage (Figs. 1A and B). NC13 was broken into two stages based on number of minutes into interphase, with early NC13 at five minutes into interphase, and NC late 13 at 12 minutes into interphase. NC14 was divided into four stages, 14A, 14B, 14C, and

14D, with embryo stage determined by three criteria: time elapsed in interphase, nuclear elongation, and progression of cellularization. NC14A was staged at 10-15 minutes into interphase, with a 1:1 ratio of nuclear length to width, and before the start of cellularization. NC14B was staged at 25-30 minutes with a nuclear elongation ratio of 2:1 and cellularization progressed less than 33%. NC14C was staged at 40-45 minutes with a nuclear elongation ratio of 3:1 and cellularization progressed less than 66%. NC14D was staged at 55-60 minutes with a nuclear elongation ratio greater than 3:1 and cellularization progressed greater than 66%. Selected embryos were placed in 100uL Trizol Reagent and snap-frozen in liquid nitrogen within one minute of imaging and stored at -80°C. Confocal images of collected embryos were analyzed and the precise nuclear cycle determined by calculating nuclear density.

RNA extraction and NanoString analysis. Embryos of desired developmental stage were selected based on confocal image analysis, thawed and crushed, and 900uL Trizol Reagent was added. Additionally, 1ul of Affymetrix GeneChip Poly-A RNA Control was added at a dilution of 1:10000. RNA was extracted from Trizol Reagent according to the standard protocol, except an additional chloroform extraction and an additional 70% Ethanol wash were preformed to increase the purity of RNA for hybridization. Purified RNA was resuspended in 10uL RNAse free dH2O and 1uL was analyzed on a NanoDrop 2000 UV-Vis Spectrophotometer to determine RNA purity and concentration. 5uL of RNA from a single embryo was hybridized with NanoString probes at 65°C for 18 hours and transcripts were quantified on the NanoString Digital Analyzer using the high sensitivity protocol and 1155 fields of view. Three single embryos were analyzed for each time point and the

average transcript count was used after normalization with GeneChip Poly-A RNA Controls and NanoString positive controls. Any NanoString experiments with abnormally high or low RNA spike-in counts were excluded from final data analysis and additional embryos were used to generate data.

A NanoString bioinformatics team carried out probe design so that all probes had similar binding properties and bound to one single exon that covered as many isoforms as possible for each gene. NanoString specifications indicate that hybridization efficiencies may vary by up to two-fold. After data was collected from the NanoString nCounter, background was removed by averaging three RNA negative runs on the nCounter, averaging the count for each probe, and subtracting probe specific background from each gene. For 75 probes, the background count was in the single digits, with the background count of a single probe giving 250 counts. This probe was deemed defective by the manufacturer and excluded from the study. Figure S1 shows the raw background counts for all probes. Table S2.1 lists probe sequences used for NanoString code set. Table S2.2 provides quantified counts for all *Drosophila* genes in the code set.

RESULTS

Creation of a Developmental Time Series.

We selected nuclear cycle 10 through gastrulation as the extent of the time series in order to focus on the beginning of the syncytial blastoderm stage when maternal transcripts are abundant and zygotic transcription is beginning, until gastrulation, when zygotic transcription is robust and many signaling pathways are functioning (Figure 1.2 A). We staged individual embryos at each time point using a transgenic line of flies carrying a

Histone-RFP fusion, using fluorescence to visually inspect and capture an image of each embryo immediately before collection. Nuclear cycle was confirmed by calculating nuclear density using confocal images. Immediately after imaging, embryos were immersed in Trizol and snap-frozen in liquid nitrogen (Figure 1.2 B).

Control mRNA spike-ins were added during extraction to determine NanoString efficiency and calculate absolute number of transcripts per embryo in a manner not biased by number of cells or other measures that rely on embryonic transcription (LOVEN *et al.* 2012). RNA was hybridized with NanoString probes according to standard protocols, and the RNA-Probe hybrid molecules were bound to slides using the nCounter Prep Station and counted using the nCounter Digital Analyzer. Raw counts were normalized using both NanoString positive controls added to the probe mix during synthesis and mRNA spike in controls added during extraction.

Quantification of Transcripts and Dynamic Range of Transcription.

To compute the absolute number of transcripts for genes included in the data set, we calculated a linear regression (R^2=0.966) for the mRNA spike-ins comparing input to NanoString counts, and fit counts for all other genes to this regression line. Using this fit, we calculated a scaling factor of 232.84 ± 11.52 (confidence interval $p \leq 0.001$) between NanoString counts and number of RNA molecules in the sample. Linear regressions for control mRNA input and NanoString positive controls are displayed in Figure 2.1 C.

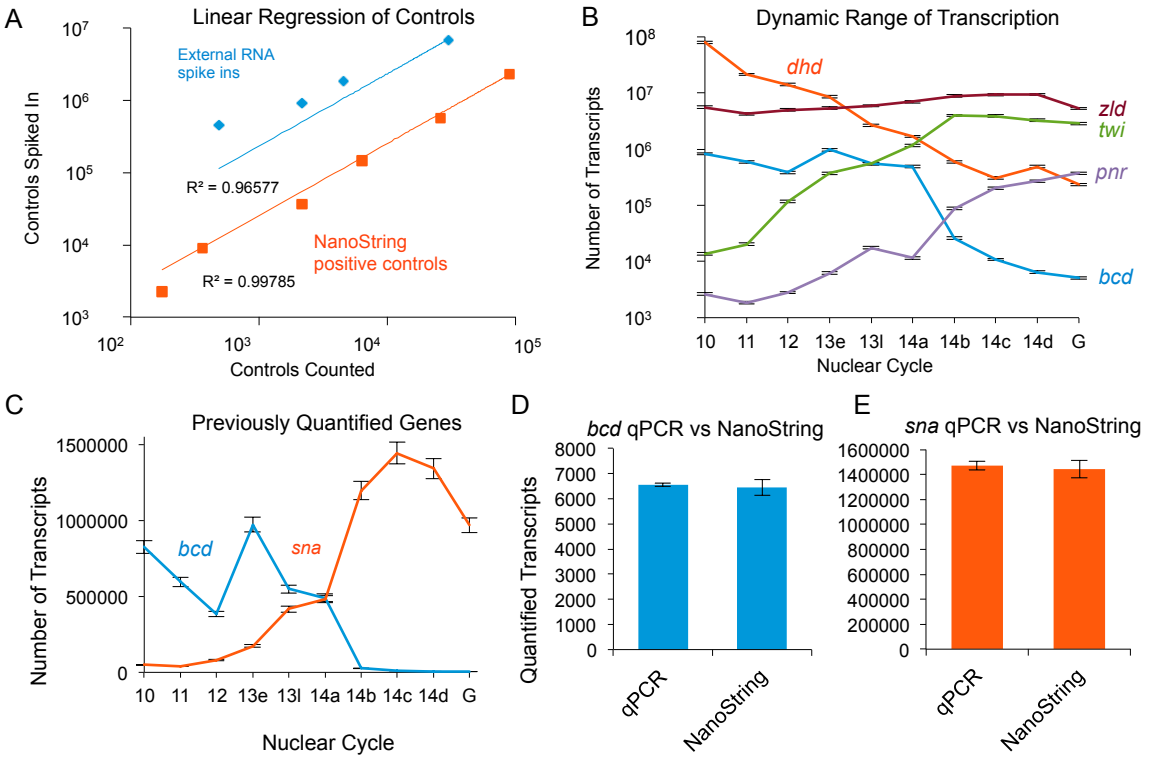


Figure 2.1. Quantification of NanoString counts and comparison to previously quantified genes. (A) Linear regression of RNA spike in controls (blue) and NanoString positive controls (orange). The graph displays both absolute number of control molecules added and number counted per sample for four foreign RNA spike-ins added to embryonic RNA during extraction and positive controls added to the NanoString probe mix during manufacture. (B) The dynamic range of transcription varies over four orders of magnitude between the least abundant (*pnr*) and most abundant (dhd) gene in the code set, but still completely within the six log dynamic range detection limit of the NanoString instrument. Error bars represent confidence interval $p \leq 0.001$. In this and other figures, number of transcripts refers to counts measured from single embryos, done in triplicate and averaged. (C) The genes *bcd* and *sna* have previously been quantified in the embryo during a single time point or subset of time points within the time course covered by this study. Their expression profiles calculated using NanoString, as measured in number of transcripts per embryo, are plotted. Error bars represent confidence interval $p \leq 0.001$. (F,G) qPCR comparing the abundance of *bcd* (D) and *sna* (E) to spike-in RNA controls shows that the ratio between *bcd* and *sna* transcripts and the controls is highly similar to the ratio calculated using NanoString. Error bars represent SEM.

We found that the temporal variation in transcript abundance for individual genes was large, with some genes changing by over three orders of magnitude in under an hour (Fig. 1D). In addition, the difference between the most and least abundant transcript within

a single time point was four orders of magnitude. In NC10, there were 7.97x10^7 copies of *dhd* and 2.63x10^3 copies of *pnr*, a fold-difference of over 30,000 (Figure 2.1D). The change in expression for single genes and the differences in expression between various genes further reinforce our division of the MZT into 10 time points to capture rapid changes and highlight the dynamic nature of embryonic development during this time period.

In order to validate the accuracy of the NanoString instrument, we performed qPCR on two embryonic genes included in the study. We selected *snail* at peak expression during NC14C as a representative of expression level of many genes during this time, and *bicoid* during NC14D when the majority of transcripts have been degraded, to validate the ability of NanoString to detect rare transcripts. We extracted total RNA from single embryos using the same method as NanoString experiments and the same exogenous mRNA spikes to quantify the number of transcripts. Using qPCR, we calculated 6,566±72 *bcd* transcripts present at NC14D, and 6458±320 using NanoString, a difference of 1.68% (Fig. 2.1 D). For *sna*, we calculated 1,472,568±3,681 transcripts in the embryo during NC14C using qPCR, and 1,442,597±71,409 transcripts using NanoString, a difference of 2.04% between qPCR and NanoString (Figure 2.1 E). Because of the essentially identical values calculated with qPCR and NanoString, we concluded that our use of external mRNAs with NanoString to quantify all genes in the dataset is accurate.

Dynamic Change Between Nuclear Cycles is Highly Variable.

When measuring the overall positive and negative change in transcript abundance from one nuclear cycle to the next, we noticed that the transition from NC14A to 14B is the most dynamic in the time course. Between NC14A and 14B, the greatest increase in

transcription and greatest amount of degradation both occur, measured as positive or negative relative change for all genes from the previous nuclear cycle (Figure 2.2 A,B). The average fold-increase for genes between NC14A and 14B was 5.6±1.2, while the average fold-increase between all other NCs was 1.9±0.2. The decrease from NCs 14A to 14B was slightly less pronounced, at 3.0±0.8 fold, compared to 1.6±0.1 for all other NCs.

Of the genes with the greatest increase from NC14A to 14B, the majority are Dorsal targets expressed in the mesoderm or neurogenic ectoderm, as well as genes also expressed in the dorsal ectoderm as part of the TGF-β pathway (Figure 2.2 C). Genes that rapidly decrease between NCs 14A and 14B are maternally deposited transcripts or are zygotic genes refined from broad to narrow patterns (Figure 2.2 D). Purely maternal genes *dhd* and *yl* were among the most reduced transcripts, as well as zygotically refined genes *zen, scw,* and *hb.*

Interestingly, the genes *bcd* and *spz*, both commonly thought of as purely maternal, showed evidence both of degradation of maternal products and zygotic transcription. Transcript counts for both *bcd* and *spz* first increased, then declined sharply between NCs 14A and 14B, indicating a quick burst of zygotic transcription as maternal products were being degraded (Figure 2.3 A). The number of transcripts remains at a higher level than the minimum counted at the maternal to zygotic switch point for three or four additional time points, adding more weight to the finding that there is new embryonic transcription of these genes.

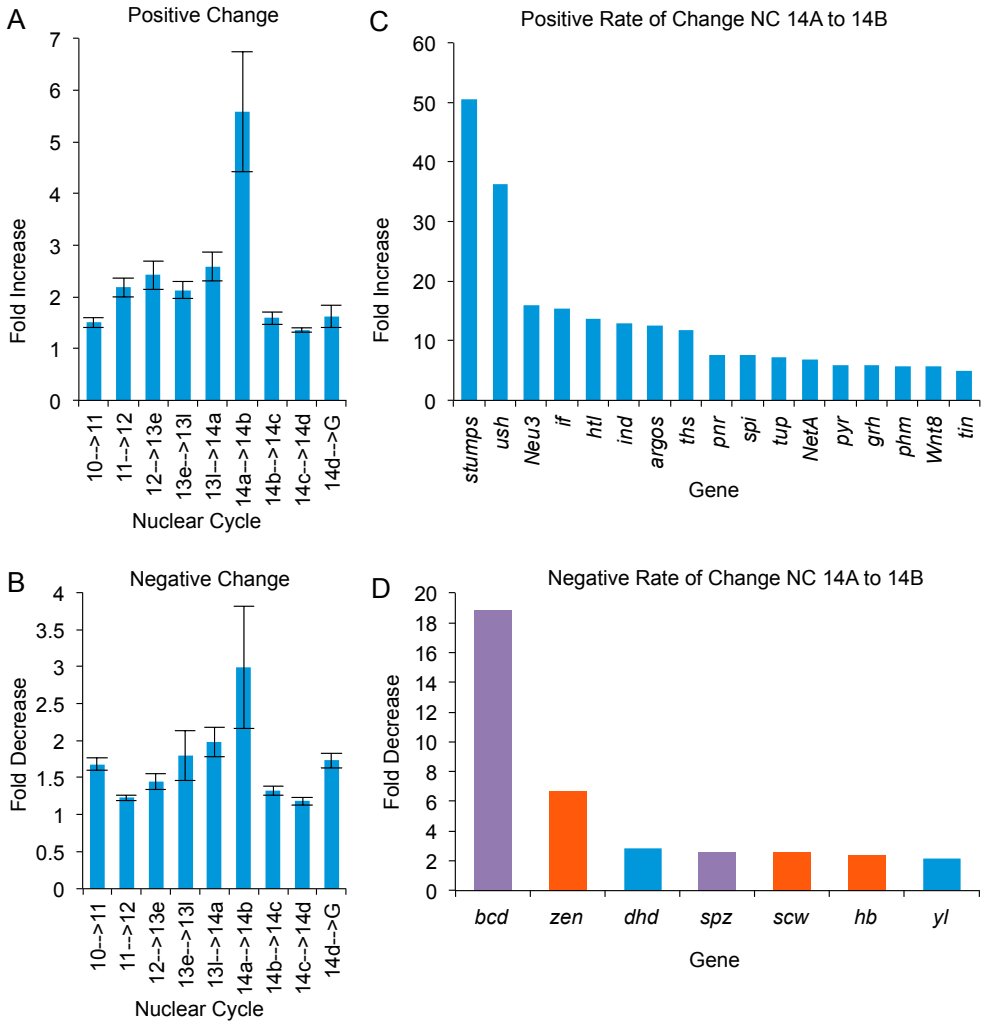


Figure 2.2. Figure 2. Dynamic change between nuclear cycles. (A, B) Average fold-increase or decrease for genes changing between each nuclear cycle. The transition from nuclear cycle 14A to nuclear cycle 14B is the most dynamic in the entire time course, both in terms of the overall increase and decrease in number of transcripts detected for genes. Between these two time points, the amount of transcripts for some genes increases more than 50-fold in around 15 minutes. Error bars represent SEM. (C) There are 17 genes with a 5-fold or greater increase between nuclear cycle 14A and 14B, most of which are direct Dorsal targets in the mesoderm and ventral ectoderm, or targets of the TGF-β pathway. (D) There are seven genes with a 2-fold or greater decrease in this period, with genes maternally deposited and being degraded (blue), broadly expressed and being spatially refined (orange), or both maternally deposited and zygotically transcribed before being degraded (purple).

In situ hybridizations using intronic probes show that there is in fact zygotic transcription of *bcd* detected as early NC11 (Figure S2.2), with dots of nascent nuclear

signal visible in many nuclei throughout the embryo. Since maternal *bcd* is spliced and mature before the egg is laid, signal from intronic probes must indicate new zygotic transcription.

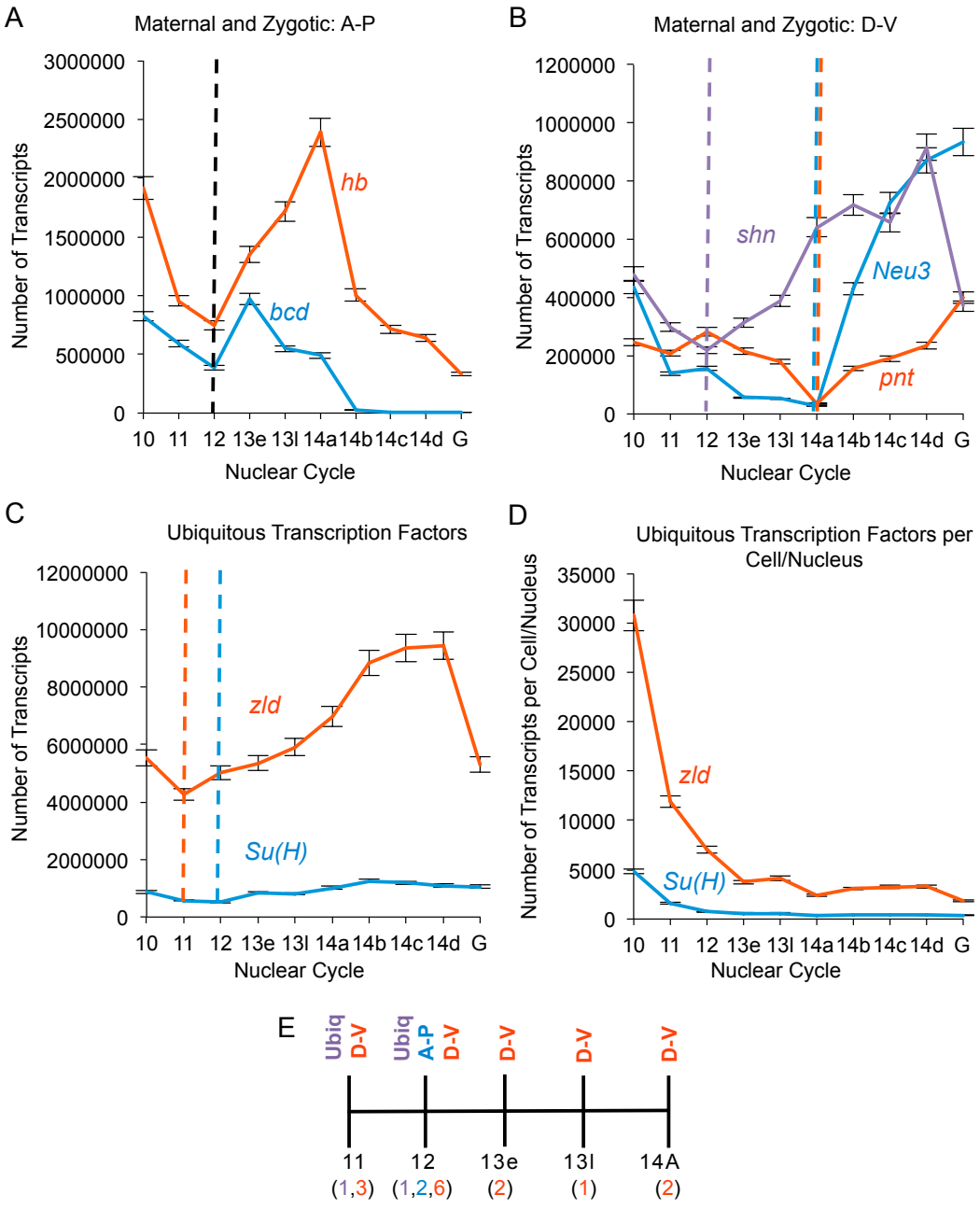


Figure 2.3. Diversity in Maternal to Zygotic Switch Points. (A) AP axis genes *hb* and *bcd* and (B) DV axis genes *shn, neu3,* and *pnt* are both maternally deposited and zygotically transcribed. (C)

The broadly acting transcription factors *zld* and *Su(H)* are both maternally deposited and zygotically expressed, with the zygotic activation occurring early at NC11 and 12 for *zld* and *Su(H),* respectively. (D) Despite the overall increase in number of transcripts for both *zld* and *Su(H),* the highest number of copies per nucleus occurs at NC10, before the maternal transcripts are completely degraded and zygotic transcription takes place. Transcription of both genes is strong enough, however, to cause a slight increase in number of transcripts per cell during NC14. (E) A timeline of maternal to zygotic switch points, with the number of each class of gene that switches at every time point. All error bars represent confidence interval $p \leq 0.001$.

It is possible that embryonic transcription is needed to maintain the correct level of protein if mRNA degradation occurs too quickly. This finding provides a new insight into the transcription and regulation of two genes and shows the strength of the NanoString system to acquire highly sensitive data that can be validated by other traditional experimental methods.

In addition to the change between nuclear cycles being highly variable, the switch from maternal to zygotic control is variable for genes that are both maternally deposited and zygotically transcribed. We define the maternal to zygotic switch point as the time when degradation of maternal input is overwhelmed by zygotic transcription, and counts increase. We included 19 dual maternal and zygotic genes in the study, and found that the maternal switch points occur as early as NC11 and as late as NC14A (Figure 2.3 A-C). Both dual switching AP genes included, *bcd* and *hb*, switch at NC12 (Figure 2.3 A) along with seven other DV genes; however DV genes *med, E(spl)m8,* and *sax* switch at NC11, *spi* and *cic* switch at early NC13*,* and *Neu3* and *pnt* switch at NC14A. The ubiquitous transcription factors *zld* and *Su(H)* have switch points at NCs 11 and 12 respectively. Because they are ubiquitous, we calculated the number of transcripts per nucleus or cell (for pre-cellularized or post-cellularized embryos, depending on nuclear cycle) in addition to the number of transcripts per embryo. Overall, maximum expression for *zld* and *Su(H)*

occurs at NCs 14D and 14B respectively, but when number of transcripts are divided by number of nuclei or cells present, transcripts are most abundant at NC10. This is consistent with studies showing that *zld* acts as an early activator of expression, with effects from lack of *zld* transcripts observed much earlier than NC14 (NIEN *et al.* 2011). Robust transcription late in the time course is able to compensate for nuclear division and dilution of transcripts, and the number of transcripts per cell for both *zld* and *Su(H)* increase during NC14.

The relative rate of transcript degradation between each nuclear cycle follows the pattern of diversity observed in maternal to zygotic switch points, in that there is a wide range of rates at which maternal transcripts are degraded. We computed relative degradation between maternal genes by calculating the percentage of transcript decrease for each gene at nuclear cycle transitions, and then comparing rates between genes. Degradation rates differ by up to 31.9% between genes, and degradation occurs until NC14A for some genes.

Zygotic Genome Activation and Mesoderm Gene Network Properties.

The mesoderm presents an opportunity to study a set of genes that are spatiotemporally co-activated. We selected the genes *twi, sna, htl, hbr, NetA,* and *mes3,* which are all dependent on the binding of the transcription factor Dorsal for their expression. When the transcripts per embryo for the mesoderm genes are compared, it is clear that there is a specific rank-order of abundance maintained throughout the time series (Figure 2.4A). *twi* is more than twice as abundant as the next gene, *mes3*, and more than seven times as abundant as the weakest gene, *htl.* All six mesoderm genes have similar boundaries on the DV axis (Figure 2.4 C'-F'), but have different boundaries on the AP axis

(Figure 2.4 C-F). The *twi* domain extends to the anterior and posterior poles of the embryo, while the *htl* domain is found in the middle ~75% of the AP axis. We counted the number of nuclei expressing all six mesoderm genes and determined the number of transcripts per nucleus.

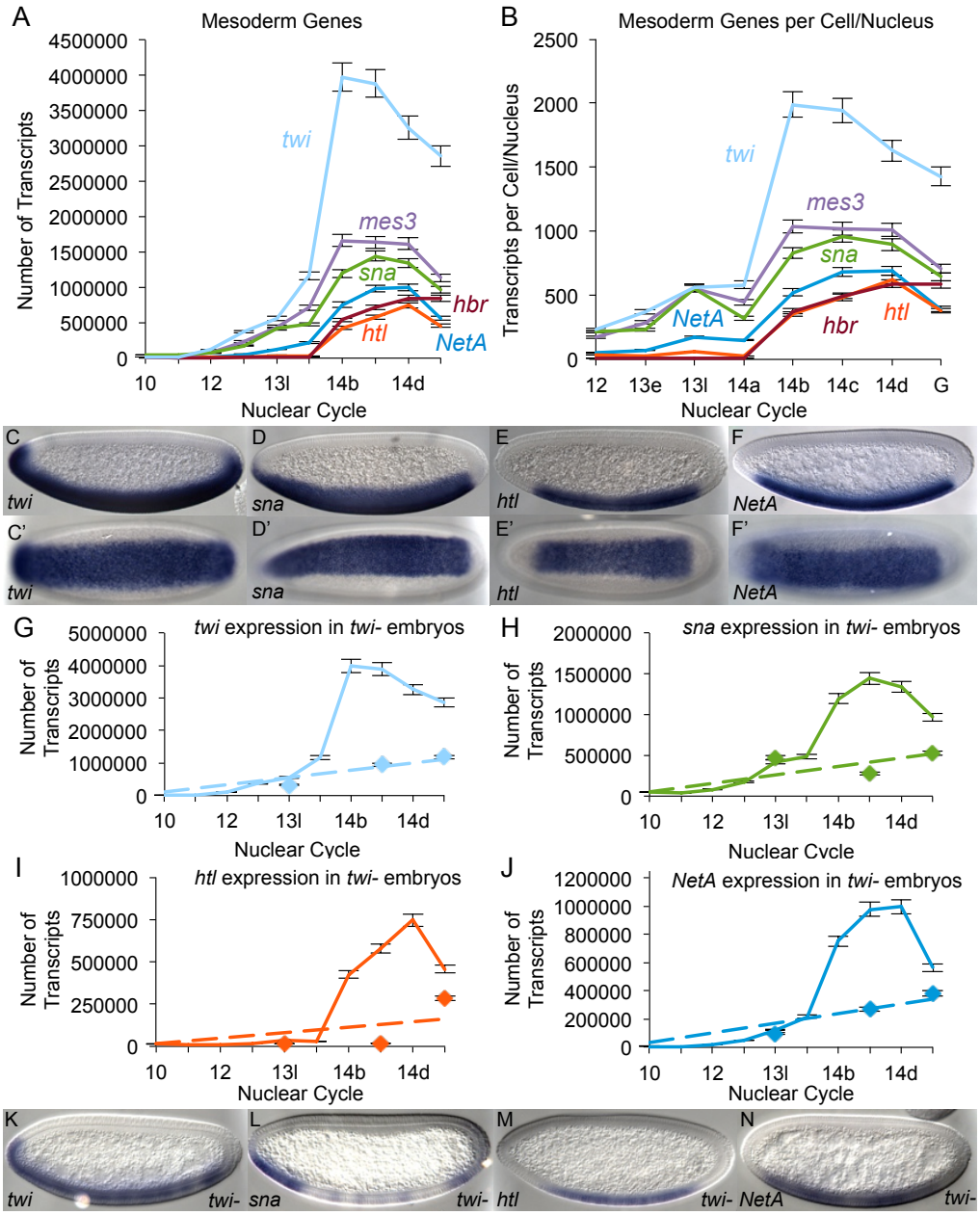


Figure 2.4. Mesoderm gene expression and transcription rates. (A) Expression profiles of the mesoderm genes *twi, htl, mes3, sna, NetA,* and *hbr*. (B) Number of transcripts per cell was

calculated by dividing the absolute number of transcripts by number of cells expressing each gene. (C-F) In situ hybridization using riboprobes against mesoderm genes *twi, sna, htl,* and *NetA,* showing their respective expression domains laterally and dorsally. (G-J) Expression of *twi, sna, htl* and *NetA* in *twi-* embryos, with mutant expression data collected at late NC13, NC14C, and gastrulation. Dashed linear regression trend lines show trajectory of expression without *twi*. (K-N) In situ hybridizations of *twi, sna, htl,* and *NetA* in *twi-* embryos. All error bars represent confidence interval $p \leq 0.001$. Embryos are staged to NC14C.

Even after normalizing for number of nuclei, the rank-order of abundance remains the same for all six genes throughout the time series, although several genes that were differentially expressed in whole-embryo counts are more similarly expressed in individual nucleus counts. In late NC13, there are ~25% more *twi* than *sna* transcripts in the whole-embryo count, and the difference between the two genes drops to less than 1% in per-nucleus counts for a short time, however the order is established again in NC14 (Figure 2.4 A, B). A similar change of 20% more *hbr* than *htl* in NC14 for the entire embryo drops to less than 3% per cell. Still, the rank order remains the same even when transcripts per cell are calculated. NCs 10 and 11 were difficult to estimate, since robust patterns do not appear until NC12, and therefore we did not include the earliest two time points in the per nucleus calculations.

It is also clear that transcription of mesoderm genes is biphasic. In NCs 10-13, there is a moderate and steady increase for each of the six genes. In NC14, the increase in number of transcripts becomes much more rapid. Since all six mesoderm genes depend on Dorsal and Twist for activation, and Dorsal is maternally deposited, we analyzed embryos from *twi*$^{-}$ flies in an attempt to explain the rapid increase in transcription observed in NC14. We selected late NC13, 14C, and gastrulation for the *twi*$^{-}$ analysis, which cover early, peak, and declining Twist activation. We found that in late NC13 *twi*$^{-}$ embryos, the average

expression of mesoderm genes was 76.4% ±11.6% of wild type, indicating that Dorsal activation accounts for around 76% of transcription at that time point, with some variability between genes, while Twist supports the rest of the activating input. At NC14C, the average expression level of mesoderm genes in *twi*¯ embryos was 22.5% ± 8.5% that of wild type. This drastic drop suggests that Twist is responsible for over 77% of the expression of mesoderm genes at this time. During gastrulation, the average expression of mesoderm genes slightly recovered to 55.9% ±3.7% of wild type levels, implying that Twist is responsible for less than half of the activation. When the data from *twi*¯ embryos is plotted with wild type data, it is evident that without Twist activation, the transcription rate of the mesoderm genes matches the early transcription rate, when Dorsal is the predominant activating transcription factor (Figure 2.4 G-J). When in situ staining is performed for all four mesoderm genes in *twi*¯ embryos, it is clear that both the expression domain and level of expression are both reduced without activation from Twist (Figure 2.4 K-N). Therefore, the input of Twist is responsible for the rapid increase in transcription observed for the mesoderm genes during nuclear cycle 14.

Sequential Activation of the TGF-β Signaling Pathway and Compensatory Transcription.

The TGF-β signaling pathway is one of the best-studied signaling pathways in *Drosophila,* and model organisms in general, and because the components are well known, presents an opportunity to observe how the MZT activates a complete signaling pathway (WU AND HILL 2009; AKHURST AND PADGETT 2015). We included 18 members of the TGF-β pathway, as well as others peripherally related. The two primary ligands are Dpp

and Scw, both purely zygotically transcribed. While peak TGF-β signaling takes place in the dorsal ectoderm, both *scw* and *dpp* are initially expressed in broader regions of the embryo. The expression of *dpp* extends to the ventral midline during NC13 and the expression of *scw* is ubiquitous starting as early as NC11, and both genes refine to the dorsal ectoderm during NC14. Our NanoString data confirms this initial broad expression and subsequent refinement of both *dpp* and *scw* (Figure 2.5 A). Furthermore, both *scw* and *dpp* decrease at very similar rates from NC14B onwards, including a pause in decreasing from NC14C to 14D, when they are both in the last stage of refining to their final expression domain. We included six TGF-β targets in the study, and found that they are all strongly activated beginning in NC14. We separated TGF-β targets into two classes, based on how they are activated in NC14. Genes *pnr, hnt,* and *Doc1* are expressed in a gradually increasing manner throughout NC14, until gastrulation when the rate of transcription levels off (Figure 2.5B). In contrast, *Race, tup,* and *ush* increase very quickly at the beginning of NC14 and reach a plateau as early as NC14B or 14C (Figure 2.5 B).

The TGF-β targets are expressed in the same general domain of the embryo, but the exact patterns differ between the genes. We counted the number of cells expressing the genes *ush, Race*, and *hnt* for NCs 14C, 14D, and at the onset of gastrulation. We focused on these three genes because they are expressed purely along the DV axis, unlike the other three that are expressed in AP-modulated patterns as well, and these time points because they fall during the peak of TGF-β signaling, when the genes are expressed in their final domains. TGF-β target expression during earlier time points is still developing and final patterns are not yet established. When the whole embryo transcript levels for the three

genes are compared, *ush* is always the most abundant, with *Race* at around 60% of the *ush* levels and *hnt* at around 22% of *ush* levels (Figure 2.5 F).

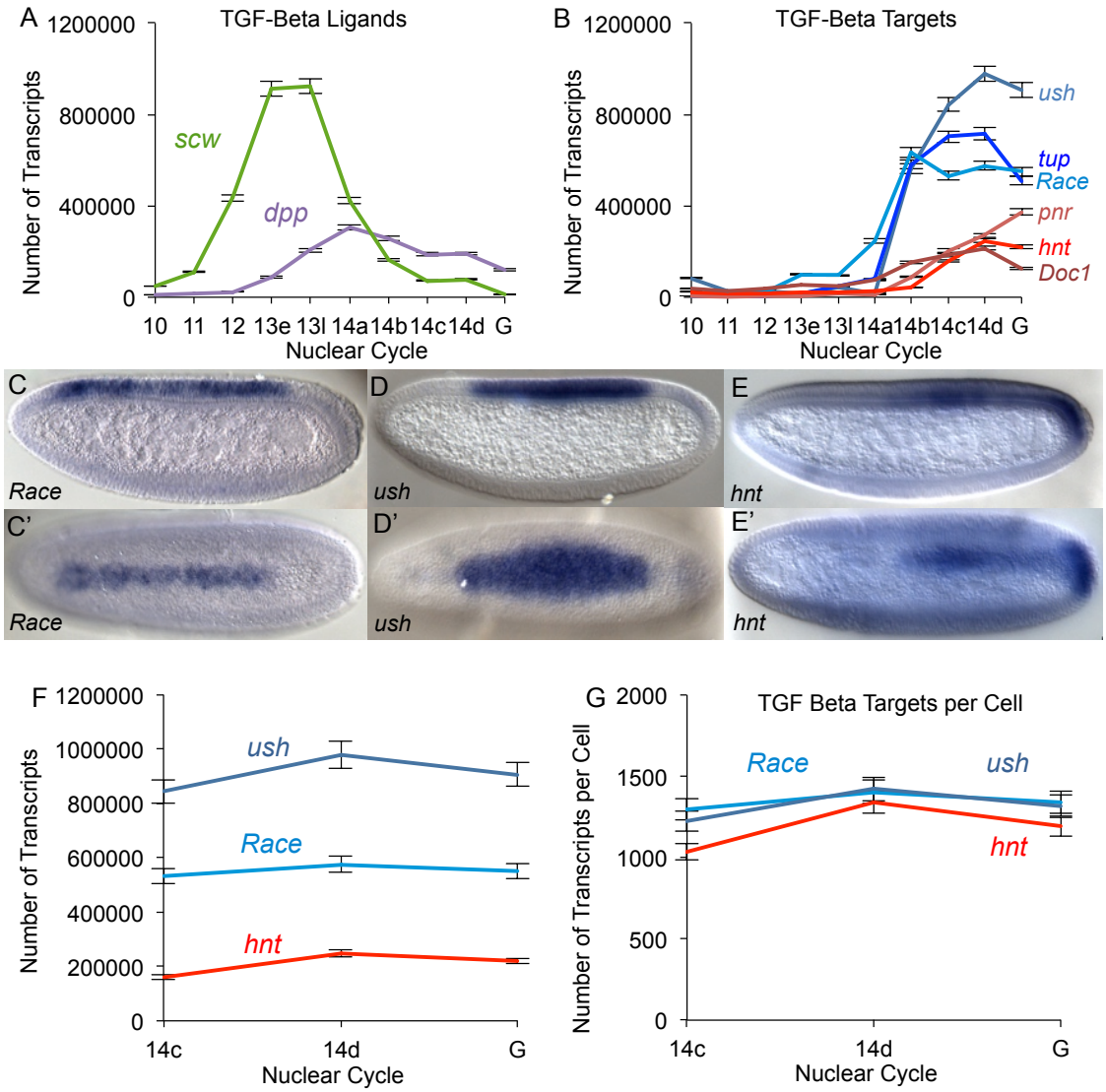


Figure 2.5. Activation and properties of the TGF-β signaling pathway and targets. (A, B) Expression profiles TGF-β ligands *scw* and *dpp* (A) and TGF-β target genes *ush, tup, Race, pnr, hnt,* and *Doc1* (B) showing number of transcripts per embryo. (C-E) In situ hybridizations showing expression patterns for *ush, Race,* and *hnt*, both laterally and dorsally. (F) Total number of transcripts per embryo during peak expression for *ush, Race,* and *hnt*. (G) Number of transcripts per cell for *ush, Race,* and *hnt*. All error bars represent confidence interval $p \leq 0.001$.

However when the number of transcripts per each cell is calculated based on expression domain (Figure 2.5 C-E), the results change drastically. The difference between

Race and *ush* drops to 1%-4% depending on the time point, and the difference between *hnt* and *Race* and *ush* drops to 6%-15% depending on time point (Figure 2.5 G). The similarity in number of transcripts expressed in each cell for *Race*, *ush,* and *hnt* suggests that the genes respond in a comparable way to common transcriptional activators. There may be repressors that define the extent of each gene's patterns, but in the cells where each of the genes is active, the genes are transcribed at similar levels.

DISCUSSION

Our use of NanoString technology combined with our fine time scale developmental window has provided a novel way to examine transcription during the MZT in *Drosophila*. The dynamic change between NCs reveals new insights into the development of *Drosophila* embryos. The transition from NC14A to 14B is the most dynamic in the study, and is unique for three reasons. First, the concentration of Dorsal in cells is at its highest level at this time point, allowing activation of genes on the dorsal edges of Dorsal gradient that were not activated by lower levels in previous NCs. Second, this is the first time that transcription proceeds uninterrupted for longer than 15 minutes, allowing a greater ramp-up time for highly expressed genes to accumulate to levels not reached before. Lastly, the combination of increased Dorsal concentrations and more time available for transcription allows novel gene interactions and cell signaling to take place within the DV gene network that were not possible before, further increasing the number of genes expressed and the levels at which they are expressed.

One exception to the biphasic transcription modes for mesoderm genes is *twi*, which begins its increase in expression rate at the end of NC13, slightly earlier than the

other genes. The combination of this earlier increase in transcription rate for *twi*, the overall highest abundance of *twi*, and the role of Twist as a master co-activator with Dorsal in the mesoderm lead us to hypothesize that the biphasic transcription for mesoderm genes is due to the input of *twi* in the gene network. In *twi*- embryos, when there would usually be the lowest endogenous abundance of *twi,* other genes change the least, and where there would usually be the highest endogenous level of *twi,* other genes are affected to the greatest degree observed. We therefore conclude that the moderate expression observed from NCs 10 to 13 is due to the input of Dorsal, while the exponential increase in expression in NC14 is due to the input of Twist and a combinatorial effect of the Dorsal-Twist feed forward loop. With *twi* as a top-level activator in the mesoderm and early target of Dorsal, high levels of *twi* are needed so Twist can robustly bind its targets in every cell where it is needed (SANDMANN *et al.* 2007). It is inline with this prevailing view, therefore, that *twi* is consistently the most abundant mesoderm gene quantified.

Two studies have quantified the number of transcripts for two genes included in this study, using FISH to estimate the number of transcripts (BOETTIGER AND LEVINE 2013; PETKOVA *et al.* 2014). One study of *bcd* transcripts prior to the syncytial blastoderm stage and NC10 found 890,000 transcripts. Our study found 824,064 transcripts during NC10, at the closest stage to the embryos used in the previous study; however significant transcript degradation occurs between the time point in the previous study and NC10. A second previous study quantified *sna* transcripts and found a maximum of around 250 transcripts per nucleus during NC13 and 200 transcripts per cell during NC14, while our data shows a maximum of around 550 transcripts per nucleus in NC13 and around 1000 transcripts per cell during peak expression at NC14C, a 2-fold to 5-fold difference.

Using FISH to count single points of fluorescence can be challenging, with probe design and microscopy techniques affecting the counts (FEMINO *et al.* 1998; RAJ *et al.* 2008; LUBECK AND CAI 2012). In addition, the combination of dense points of fluorescence signal making it difficult to distinguish individual spots, and the use of a threshold to exclude fluorescent signal, may reduce the number of transcripts counted and account for the differences between our quantification and the numbers calculated for *bcd* and *sna*. The authors of the *sna* study counted only cytoplasmic signal, excluding nuclear transcripts, which might have reduced the count and, by design, did not account for active transcription. One factor that could slightly inflate the number of *sna* transcripts per cell we calculated is a low level of background transcription in non-mesoderm cells. If *sna* is expressed at a very low level in cells outside the mesoderm, our calculations would attribute these transcripts to mesoderm cells and slightly increase our quantification. This would lead to a negligible increase, since the transcriptional activity of cells expressing *sna* is so much stronger than non-mesoderm cells that *sna* transcripts are undetectable using standard in situ hybridization.

Furthermore, our qPCR data reinforce the accuracy of our quantification method and post-collection data analysis and processing. Previous foundational studies have compared changes in gene expression using NanoString and qPCR for different time points in the development of sea urchin embryos, and found that the relative fold-changes calculated between time points were highly correlated between NanoString and qPCR data (GEISS *et al.* 2008a; MATERNA *et al.* 2010).

The diversity observed for both the maternal to zygotic switch point and the degradation rate for can be explained by the increasing concentration of Dorsal in nuclei

during successive NCs. As the concentration of Dorsal increases, the activation of target genes occurs at different rates and times, depending on whether genes depend directly on Dorsal, the concentration of Dorsal required, or the necessity of an intermediate gene. It is possible that degradation rates alone for genes are much more similar than we have observed, but since genes are activated at different rates and times, the varying influx of embryonic transcripts may cause the observed degradation rate to differ from the basal level.

Although NanoString technology does not provide spatial information on gene expression *a priori*, interesting trends or new insights from this data can be validated using other methods. In the case of mesoderm genes, using NanoString we determined that a rank-order of abundance is established early in development and is maintained robustly through the time series. The rank-order of genes was first observed for the entire embryo, meaning that spatial variations were not originally taken into account, but remained the same after transcripts were normalized for number of cells, indicating that the order is established and maintained at the level of gene regulation (e.g. enhancer and gene network properties). This combination of NanoString data and spatial information strengthens the finding and provides an example of how NanoString can be used to investigate multiple genes simultaneously and integrate with other methods.

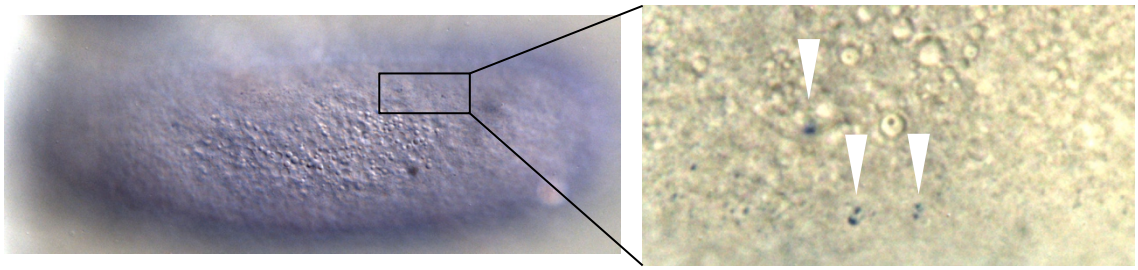
Of the six TGF-β targets studied, *Race, ush,* and *hnt* are expressed only in the dorsal ectoderm, while *pnr*, *tup,* and *Doc1* are also regulated along the AP axis, expressed in stripes or laterally towards the midline of the embryo. The TGF-β targets respond to activation in two distinct ways, with half of the genes rapidly transcribed between NCs 14A and 14B and quickly reaching a steady state, and half of the genes being continuously

transcribed at moderate rates until gastrulation. The different modes of transcriptional activation do not appear to correlate with the genes based on expression patterns, indicating that there could be an unknown factor involved in rapidly activating one set of genes, just as we have shown that *twi* rapidly activates mesoderm genes. Once the TGF-β targets are activated and reach their peak expression, the maintenance of final levels might no longer depend on this initial activating signal, just as the mesoderm genes depend on *twi* the least at gastrulation, after peak expression. While these TGF-β target genes are diverse in terms of function, the convergence of transcript abundance in each cell, we propose, may demonstrate a unique property of the signalling pathway to integrate changing levels of input to maintain stable and reliable transcription of target genes. This property can be contrasted with the six mesoderm genes, where even after normalizing for number of nuclei expressing each gene, many differences in expression remained present throughout the time course. This difference may exist because the six mesoderm genes are at the top level of signaling pathways (e.g. *htl* FGF receptor) while the TGF-β targets are at the output level. Varying levels of top-level input signal may be integrated (i.e. coordinated) in order to provide a similar output level of many downstream target genes within a tissue.

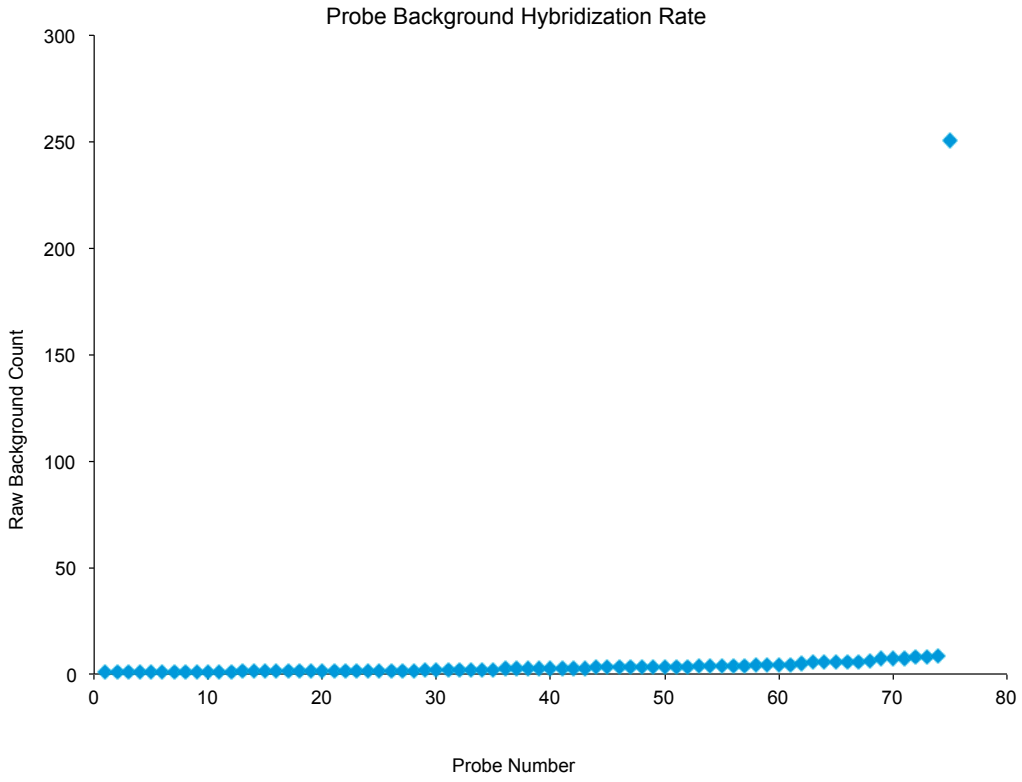
We have demonstrated the use of NanoString as a new technology to precisely quantify transcripts and create a fine scale time course of *Drosophila* embryonic development. In addition to being the first large-scale quantification during *Drosophila* development, this study has provided new insights into the sequential activation of gene regulatory networks and suggested that network properties regulate levels of transcription for clades of genes. We believe the most promising future use of NanoString is in the

characterization of mutant phenotypes and accurately measuring changes in expression of large numbers of genes in mutant backgrounds, as we show with *twi* mutant data.

SUPPLEMENTAL MATERIAL



Supplemental Figure 2.1. In situ hybridization with proves for *bcd* introns, showing an embryo in NC 11.



Supplemental Figure 2.2. Background hybridization rate for all NanoString probes. Raw background counts are an average of three RNA-negative runs and sorted from lowest to highest counts. Probe for *LacZ* as a control had a high background rate and was removed from analysis.

Symbol	Probe sequence
ASPP	GTCGGCATCGACTGCTAATAAGACGCAAACGGCCACCTTTGGACTGCAATCCCTTAACATCAACGACAATTTACCCATCAAAGCGAAGCCACTGACCATC
CIAPIN1	CTGCATTTGGTGTCCTACATCGGTCCGGCGGCTAGTTTGCTGCAGGAGATCAAGCTATCCGGATTCATAAACTGTCGCGAGGACTCTCCAGATGCTTTGA
Dl	CTGGCAGCAACAGCGGTCTCACCTTCGATGGCGGCAACCCGAATATCATCAAAAACACCTGGGACAAGTCGGTCAACAACATTTGTGCCTCAGCAGCAGC
Doc1	CAGCGCCAGGATTTTGAGCAGGACTCAACGGGTAGTGTTTCGCCAGCTTACTATGGAGCAACTCATCCAGTTGCCAATCCTATGTCGCCGTATCTCCGCC
Egfr	AAACTACAAATCATTCGCGGACGCACGCTGTTCAGCTTATCCGTGGAGGAGGAGAAGTATGCCTTGTTCGTCACTTATTCCAAAATGTACACGCTGGAGA
LacZ	CATCAAAAAATGGCTTTCGCTACCTGGAGAGACGCGCCCGCTGATCCTTTGCGAATACGCCCACGCGATGGGTAACAGTCTTGGCGGTTTCGCTAAATAC
Mhs/Dr	CCCTTCGGACCGCCTGGCATGTTTCCGGGAGCAGGATTCGGTGGAGATGCCAACGAACCGCCGCGGATCAAGTGCAACCTGCGGAAGCACAAGCCCAACC
N	GATGGAACTTGCCACGACAAGATCAATGGCTTTAAATGCAGTTGTGCCCTGGGCTTTACGGGCGCACGATGTCAAATCAATATAGACGACTGTCAGTCGC
NetA (probe 1)	TCCTGTCAGCCCAGTTTAAGCATTTGTGAACAATTTACGTACTAGGGCCGTCTCGAAATCAAACAGAGCGCCAGTAAGCAATAATTTATGTGTAAGAAGC
NetA (probe 2)	CGCGCCTGCATTCCGGACTTTGTGAACGCCGCCTACGATGCTCCAGTGGTGGCCAGTTCCACGTGCGGATCGTCGGGTGCGCAGCGATATTGCGAGTACC
Neu3/Meltrin	ACGAGATGGGGCACAACTTTGGCATGGAGCACGATACATCGGATTGTCATTGTCGGGATGAGAAGTGTGTGATGGCTGCCTCGAGTACCTCGTTTATTCC
Race/Acne	TCTACGACAAGGCCGGAACCGCCGTGCGATCTCAATTCGAACGCTACGTGGAGCTCAACCACCAAGGCAGCCAAACTGAATAACTTTACTTCTGGTGCCGA
RpL32	TGCCCACCGGATTCAAGAAGTTCCTGGTGCACAACGTGCGCGAGCTGGAGGTCCTGCTCATGCAGAACCGCGTTTACTGCGGCGAGATCGCCCACGGCGT
RpL7-like	GTGTCATCTGTCTGGAGGATATAATCCACGAAATTTGCACAGTGGGTCCAAACTTCGCCGCTGTAAATGAGTTTCTGTGCGCCTTCACGTTGTCCAGTCC
SoxN	TGGGCCAGTCGCACCTGACCCACTCGCAGCTGAGCCATCACAACCACCATCATCACCACATGAGCGCCCATATCGCGGCCAGCCAGAGTCCCAATCCGCT
Su(H)	GAAGCCGACGATCCCGTTTCTCAGCTGCACAAATGCGCCTTCTACATGAAGGATACGGATCGGATGTATCTCTGTTTGTCGCAGGAGAAAATCATACAGT
Tl	GGGCAACCTTGTGACCCTCGTCATGAGCAGGAATCGACTGCGTACCATCGATTCCAGAGCCTTTGTAAGCACAAATGGACTGCGACATTTGCATTTGGAT
wntD	GCGGTACTAAAGCCATTCGAAGCCATTGCCCAGGATCTCCTCCAAATGTACGACGATGCCATCCAACTAGAAGGAGCCAGTAGCAATCTAAAGATTATGT
YFP	GAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGTGTATATCACC
aos	ATGTCGTTGTCCTGAATCGAATCGCATGCCCAACAATGTGATCATCCATCATCACAGTCATTCCTCGGGATCGGTGGATTCCCTGAAGTACAGGAACTAC
bcd	TGAAGGGTCTGGACAAGAGCTGCGACGATGGCAGTAGCGACGACATGAGCACCGGAATAAGAGCCTTAGCAGGAACCGGAAATCGTGGAGCGGCATTTGC
brk	CCGCCGTCAAATGCAGTTGCACGAAAAGGATCCATCTGGTGTGGATCTCACTTTCCGCAAACGTAAGGTCATCACTAGTCCGATGCAGCCGGATAAGATT
cic	GGCAGCTGCGGTGGCCAGGGCATATCATCCGATTTGCAGGGCGATATTATACCGCTGACAATTGACAACTACAATAGCACCTGTGATGAGGCGCCAACTA
dhd	ATACTCCAGCAAGGCGGTGGTGCTCAAGATCGATGTGGACAAATTCGAGGAGCTGACGGAGCGCTACAAGGTGCGCAGCATGCCAACGTTTGTCTTTTTG
dpp	CAGCGAGCCCGCCTCGTTCAGTGATAGTGATAAAAGCCATCGGAGTAAAACAAACAAAAAACCTAGCAAAAGTGACGCGAACCGACAGTTCAACGAAGTG
eve	AGACGCCGCCGCTGGATCTGCAGTCGTCGTCATCGCCGCACTCCTCCACGCTGTCGCTCTCGCCAGTGGGATCCGATCACGCCAAGGTGTTCGACCGCAG
fdy	CAACAACAACCAACGCTACGGAAACAAGGAGTCGATCGGAGAATTCGGAAACGAGTTGCCGCAGCGTCAATTCAACAATCGTGACAACAGGGGACCACCA
grh	AGATCAACATTGCGTTCAGTGCTTGAGCACGGATTTCAGCAGTCAAAAGGGAGTTAAGGGCCTGCCGCTGCACGTACAAATCGACACATTTGAGGACCC
hb	CACCGCAAGTCGCACAGTTCTGTGTATCAGTACCGTTGTGCGGATTGTGATTACGCCACCAAGTATTGCCACAGCTTCAAGCTGCATCTGCGCAAGTATG
hnt/peb	CCATTACGGCGCCCAGTACTAATCCCAGCAGTTTGAAAACGATGATCGCTCAAGCCGAATATGTGGGAAAATCCCTCAAAGAAGTGGCCAGTTCACCGTT
htl	TAATCAATCGAGCACATCCAGTGCGGATTTGGATGATGGCGCCGCAGATGATGATGATAACAAGGCCGATTTGCCGGTGAATGTCAGTTCGAAACCCTAC
if	GTGCCAGCGTTCACATCGAGAAGGCTCGCGGTGAAGGTTTCGTTCGAGGAGTCTTGGTCAGCAATAGCACCGATGCGGGTGACAAATTAAGTCCCAAACA
ind	CCATCCATACGCCCAGCTGTTCGCCAGCAAGAGGAAGTCCAGTGGATTCAGCAACTACGAAGGATGTTATCCATCGCCACCGCTCTCGGCCAATCCCAAC
m8/E(spl)m8-HLH	CTCCACACCTGGCGATGAGCGTCGACCTGGGCAAATCGGTGATGACTCACCTGGGACGCGTCTACAAGAACTTGCAGCAATTCCACGAAGCACAGTCCGCC
mad	GAACAGTGGATTCAACTCGCACAGCTTGAGCACCAGCAACACATCGGTGGGCAGTCCGAGTTCCGTCAACTCCAATCCCAATTCGCCGTACGACAGCTTG
med	CCATCGCATACTTCGAGTTGGACACGCAAGTGGGCGAGACCTTCAAGGTGCCGTCAGCGAAACCGAACGTAATCATCGACGGCTATGTGGATCCCTCTGG
mes3/llp4	CGCCGAAAGATGTGCGGCGAGGCTCTGATCCAGGCACTGGATGTGATTTGTGTTAATGGATTTACACGCCGTGTCAGGCGGAGCAGTGCGTCTAAGGATG
otd/oc	TTTAGCAATCAGAATCAGGTCAACTACAACATGGGCCATTCGGGCTACACGGCCTCCAATTTTGGTCTGTCGCCATCGCCATCCTTCACGGGCACCGTGT
phm	CCTACCAGGCGCCGCCACAGTTCATCCCATTCTCGTCCGGCTATCGAATGTGTCCCGGCGAAGAGATGGCTCGCATGATACTCACGCTCTTTACGGGTCG
pnr	AACAATGCTTTCCCCTTTACGGACAGACGACGACGCAACAGCAGCACCAGCAACATGGCCACAGCATGACATCATCCTCTGGACAGGCACATCTGAGTGC
pnt	GACAAGACGTGTCAGAGTTTCATTTCGTGGACCGGCGATGGCTGGGAGTTCAAGCTAACAGATCCTGATGAGGTGGCTCGTCGTTGGGGAATTCGCAAAA
put	CGCCATCCGAACATCCTCGAATTCCTGGGCGTTGAGAAGCACATGGACAAGCCGGAATATTGGCTGATATCCACCTACCAGCATAACGGATCACTATGCG
pyr	AAAGCCAAGTTGAGACAACAGCGAACCCCGAAGGTGAGACAGAGACAGTTACTACCGAAACTGTAATGCAGAATCGCAATCATATTGATGCTAATAATAT
rho	ATCCTAGTGATCTCCATCATTGAGATTGCCATCTTCGCCTACGACCGCTACACAATGCCCGCCCAGAATTTCGGGCTACCCGTTCCGATTCCGTCGGATT
sax	GTTCTGGGTCCTTTTCTGGTCATCGCTCTGCTGGGCGCCGTGACCATCTTCTTCATTCGTCGTAGCCATCGCAAGCGTCTGGCTGCCTCGCGAACCAAAC
scw	CAATAATCACATCGAGATGCCCATATACCAGAAGCGTCCTTTGAGCGAGCAAATGGAGATGATTGATATCCTGGATTTGGGTGACCGTCCGAGACGACAG
shn	CAGTCCTGCAAACAGCATTGGCTCTTCGGGTTCGCAACCTAAAAGATTAGTGTGCTCCTTTACATCGCCGAAGCCGCCTTTCGACTACCAAAAGCAGGAA
sim	CGTGCATTTGGGCCTCAGTCAGGTTGAGCTGACGGGCAACTCGATATTCGAGTACATACACAACTACGATCAGGACGAGATGAATGCCATTTTGTCGCTG
sli	ATAGTCTTCAGTAGCGCGGAGCAGAATGGAATTCTCATGTACGACGGCCAGGATGCTCATCTTGCAGTGGAACTGTTTAATGGGCGTATTCGGGTTAGCT
sna	TTGCCGCCAATCATGCCAAAAACTACCGCTTCAAGTGCGACGAGTGCCAGAAGATGTACTCCACCTCGATGGGCCTGTCCAAGCACCGTCAGTTCCACTG
sog (probe 1)	TGCTCATCATCGCCGGACTGCTGATCGTCTGCTTGGCGGGCGTGACGGAGGGCCGCCGGCATGCGCCGCTCATGTTCGAGGAGTCCGACACGGGCAGGCG
sog (probe 2)	GCAGTGCCCAGGATTCCTGTCAGATGTGCGCCTGTTTGCGTGGCCAATCCAGTTGCGAGGTGATCAAGTGTCCGGCTCTCAAGTGCAAGTCCACGGAGCA
spi	AAACCTTCGATGCCTGGTACTGTTTGAACGATGCCCATTGCTTTGCGGTGAAGATAGCCGATCTACCGGTTTACAGCTGCGAGTGCGCGATTGGCTTTAT
spz	TCAGACGAGCGATTCCTTTGCAGGAGCATCAGGAAGCTGGTGTACCCAAAAAAGGGCTTGAGGGCGGACGACACCTGGCAGTTAATTGTCAATAACGATG
star/S	GCTGCCAGAAGTCCACGGGCATGGAGTCCCGGGCGAGAACATCTCCGCAACAGATCCAGCATCCGCACAGACAGCACCACCAGCAGCAGCACCATCACCA
stumps	TACTGGCCATTTTGTTGGATGTCAGCGAGGAGAAGGTGAAGGAGATGCACAAGTCGGCACTGCCCAGTTACTACAAGTGGCGTCGTTGTGTGGTGCGCGA
ths	CAACCTGAACAACACACTAACGGAGGCTTCTCCCACGGCAACATCAACATCCAGCATCACGTCCGCAACAACCATGAAGAGACCCATACGGAAGTTCACC
tin	GGAGCATCTACGGTTCAGATGAGGGCGCCAAAAAAAAGGATAACAGCCAGGTGACCTCCTCGCGTTCCGAGCTGCGAAAAAACAGCATCAGTGGGAACAG
tkv	CAACCTGAACAACACACTAACGGAGGCTTCTCCCACGGCAACATCAACATCCAGCATCACGTCCGCAACAACCATGAAGAGACCCATACGGAAGTTCACC
tld	GAACTGGATGTCATTCACATTGCACAGAATCCACGTGTCGGCACTCGACGCTACATGGCTCCAGAAGTATTGAGTCAGCAGCTGGATCCCAAGCAGTTTG
tsg	CCATGAAAATGAAAGCGAAATTAGTTGTTCTCTCTGTGGGAGCTCTCTGGATGATGATGTTTTTTCTGGTGGATTACGCCGAGGGCAGGCGACTGAGTCA
tup	AGCCGTCGTATCGCGAACATCGACCGTCTTAGCTCGATAGCCGTCGGATTTGCCTTTTCGGATGCGAGAGACTCTTAAATCCCTCACCATAGATCCAAAT
twi	GATGCACTTCCAAAACGCCTACAGACAAAGTTTCGAGGGCTACGAGCCGGCCAACTCGCTGAACGGCAGTGCTTACTCCAGCTCGGATCGGGATGACATG
ush	TAAAACCAGCCAGATTTATGTGCCTGCCCTGTGGGATTGCATTCAGTTCGCCCTCAACTCTGGAGGCCCATCAAGCCTACTATTGCTCTCATAGAATTAA
vn (probe 1)	GGCTTATTCGATGCAAAGGTGCCGTCGTCCCAACGCTCTGGCTTCCATATATCGGACATCTTGAATTTGGAGGGCTCTGAGCTGAAGAATGCAGCAGCTG
vn (probe 2)	AATATCTTAAGTAGGCTGCTCAGCCTCAACAGCAATAGTCTTAGTAGCCGTAGCAATGTTAAGTTAAAGCCAGCAACAGTATTCGACGCCGGCAGCAGTA
vnd	GGCTTATTCGATGCAAAGGTGCCGTCGTCCCAACGCTCTGGCTTCCATATATCGGACATCTTGAATTTGGAGGGCTCTGAGCTGAAGAATGCAGCAGCTG
yl	AGTTTCAGCCAGAGTCGCACGAGAGCAACGTCCTGGTCCGCTCCAATTTGGGCAATGTGAGTGCCCTGGCCTTCGATCATTTGAGCAGGAATCTATACTG
zen	TGAGATTGCTCAGCGCTTGTCCCTGTGCGAACGCCAGGTGAAGATCTGGTTCCAGAACCGACGAATGAAGTTCAAGAAGGACATACAAGGTCACCGCGAG
zld/vfl	GGCACTGATTACGCCTTTGGCACCGTGGGCGTTGACTATGGCAAGAGCGCCGGTGTTCGCTATCATCCCTACCTGCAGACGCCAACCTCCGGACTGGGAG
Rm62	TTTGGGCAACTACATCCAGATCAACATTGGATCGCTGGAGCTGTCCGCCAATCACAACATCCGCCAGGTGGTGGACGTTTGTGACGAGTTCAGCAAGGAG
dap	AATTAGTCATTGCGGGACCGCGTGGAAGAATGGGGCAGGAAGCTGTTAAATTGGCAGAACGAACACCACATTTTGACCTTGTAGGGGCCATAGACCATAC
lys	TATGCGAGCAAAGCATTCTCATCAGTCGCAATGATTCAGCTCGCTGAGGAAGAGGGACTTTCTTTAGATGTCGTATCCGGAGAGAGCTATATACGGCTGT
phe	AAAGCGTTTGATGATGTATTGATTCCAGGGGCCATGCAGGAGCTTGAAGCACTCGGCTGCAAAGTGAGGCTTCTGGGTGCATACCAGTCTTACCAATTAT
thr	CAGCAAGGTCGTAGCTGTGTTAACAGGAAACGGACTGAAAGATCCGAACACAGCGGTCGACATTTCAGAAATCAAGCCTGTCACATTGCCGACTGATGAA

Supplemental Table 2.1. NanoString probe sequences.

Gene	NC10	NC11	NC12	NC13 early	NC13 late	NC14A	NC14B	NC14C	NC14D	Gastrulation	twi-NC13 late	twi- NC14C	twi- G
ASPP	8288	6688	7916	8155	10422	112335	450505	551935	876282	563944			
CIAPIN1	761129	542467	830513	734745	661469	467461	823911	695686	956214	430563			
Dl	935709	663601	518982	585214	888970	875290	1660717	1731491	2372444	1801880			
Doc1	37712	24320	35864	53056	48315	75013	152344	183487	214242	125614			
Egfr	16083	17321	61660	592139	669584	1868709	2109015	1957941	3003563	1421675			
LacZ	182186	293310	583454	232979	129289	547485	316991	409355	525149	287243			
Msh	7162	8986	11232	11402	15824	22560	50759	109471	165301	258317			
N	1558777	1189694	1111618	1211288	1546959	1167607	995446	1132822	1686271	866623			
NetA	6724	3591	18350	50983	123138	217278	752457	978334	996502	565661	96717	269647	382261
NetA	7842	2870	4869	4856	21973	60994	415134	430662	900211	324375	5090	124160	226021
Neu3	435331	139034	157481	56727	52197	27079	431172	724790	870291	934613			
Race	5129	7538	22196	100307	95235	245341	634014	531354	575584	549055			
RpL32	16905105	13720549	23781430	20722641	13362584	15668506	20545028	19085664	16268679	15827045			
RpL7-like	1326432	787879	753764	1092939	999343	938951	1615813	1393553	1402098	1120626			
SoxN	5309	5502	5098	15390	60970	220887	455703	664135	1123229	816014			
Su(H)	862571	572156	518462	847601	819821	1018784	1253770	1195990	1094373	1053576			
Tl	4389119	2970065	2936472	3842825	3701868	2677429	3675885	2532735	2410680	601151			
Wnt8	9774	6597	20135	21129	29529	9309	52747	191934	180746	248468			
YFP	11384	16091	54591	92940	9106	35942	37433	97376	52526	170385			
argos	31026	14868	15948	18819	25707	13444	168035	184339	180897	107737			
bcd	824064	593817	383500	972066	547987	490086	26033	10922	6458	5122			
brk	8398	5245	8674	32586	97685	344800	354206	422148	645661	283417			
cic	957247	673918	497843	436273	781548	612413	403992	360468	435226	263707			
dhd	79674243	21449661	14059531	8572801	2660381	1685924	595044	307924	486971	235065			
dpp	8632	14373	23964	86638	204534	304993	257196	187936	191752	119736			
eve	45060	103559	335322	458662	358758	810429	2109262	1718719	1451811	1093716			
fdy	8190983	2562347	2854680	1258724	814592	273225	763763	881542	809285	950438			
grh	5160	7656	6322	12892	46858	166102	982027	1135535	1440205	731692			
hb	1914781	955723	745552	1349177	1716938	2389012	1003582	713784	640024	332298			
hnt	23844	15592	18270	19535	20618	24522	43901	160272	247744	220214			
htl	13707	8416	9961	16187	36252	31042	424855	578598	746791	456443	15257	18281	284297
if	7483	5567	9669	9148	13287	8248	126433	250528	374144	242609			
ind	9782	6833	12585	13634	29097	9643	123908	242090	184619	180276			
m8	49895	30608	36453	119173	85588	61094	77560	148526	226797	147635			
mad	2320229	1606187	1499279	2120802	1992640	1775663	1566217	955707	1015417	364573			
med	769012	515610	744110	710495	585845	483242	514247	441612	410031	351410			
mes3	14145	13770	69304	229645	442062	712089	1663073	1633260	1616914	1138062			
otd	7397	4204	9937	65295	177593	479657	813160	746499	1064261	537093			
phm	2799	2375	7769	12665	47528	114560	651684	956336	792661	647817			
pnr	2627	1837	2797	6243	17618	11523	87999	202723	270344	373276			
pnt	247152	208096	282070	215594	179415	34791	156893	189810	235059	399086			
put	1225353	744442	525653	773866	956389	954800	706486	580462	542102	253371			
pyr	3800	5535	11320	9898	16944	36753	220649	230789	360600	188092			
rho	13199	6920	8970	23837	31396	167800	143865	130785	174815	64487			
sax	327269	236816	405846	361637	213050	156841	134686	124414	154724	104473			
scw	48284	110085	434630	913014	925065	422228	162527	73058	76205	12681			
shn	480283	298802	218485	313696	389072	641353	717375	658500	915969	370918			
sim	19353	11501	17189	11473	10200	13827	23298	60815	124627	145286			
sli	41021	29031	29068	109052	208056	557530	1890349	2128538	3005397	1938660			
sna	48743	39533	80456	174096	416652	481341	1196169	1442597	1341323	967958	473406	274218	527574
sog	8108	7109	7152	81861	160970	699008	2203870	1668058	2108290	1548549			
sog	23511	26939	163448	193414	147083	888861	1483394	897275	1235709	797801			
spi	301687	171465	141912	113561	174344	174985	1330121	1741162	2377326	1828567			
spz	1774940	1582678	3340007	2240101	829977	190455	73095	52781	70276	53562			
star	1494056	686253	536176	631415	756432	748682	775022	844008	1107001	775543			
stumps	8053	3612	3017	7669	7195	10694	540227	714826	848024	843642			
ths	5129	5136	5727	9809	18330	22865	270168	359123	663613	517465			
tin	5012	4259	7211	4327	15468	4852	24035	81743	117617	234121			
tkv	798740	487457	401224	561750	606528	539503	641227	481711	728126	512325			
tld	71046	126383	287898	661702	636138	1197603	1004434	827969	818131	319511			
tsg	78961	184946	956782	874300	657930	740006	723239	503869	412145	197618			
tup	3542	3689	7538	17393	49082	80574	580765	702873	714659	510458			
twi	13573	20122	116034	370665	557544	1164521	3974135	3884520	3257041	2854478	325489	944528	1187181
ush	82242	27256	29118	15126	43747	15505	561491	842846	977968	907121			
vn	26053	22694	14342	31297	25129	29354	96505	81081	87258	74668			
vn	6779	8620	10006	15198	23184	99396	145492	108087	201069	137041			
vnd	10048	4296	8142	18357	68824	191621	267890	306959	406459	209536			
yl	11487866	5404336	5176009	5078951	1382270	410441	190539	134058	146212	94954			
zen	117039	190098	444769	1375704	1091425	2617444	389283	341477	311618	194090			
zld	5522624	4260426	5020075	5357893	5900820	6979035	8844677	9350096	9438153	5305075			
Rm62	7297964	5060832	9263082	7311270	5693838	3719954	6563911	6556429	5546910	4898260			

Supplemental Table 2.2. Transcript counts for all *Drosophila* genes included in the NanoString code set.

REFRENCES

Akhurst, R. J., and R. W. Padgett, 2015 Matters of context guide future research in TGF beta superfamily signaling. Science Signaling 8.

Ali-Murthy, Z., S. E. Lott, M. B. Eisen and T. B. Kornberg, 2013 An essential role for zygotic expression in the pre-cellular Drosophila embryo. PLoS Genet 9**:** e1003428.

Bate, M., and A. Martinez Arias, 1993 *The Development of Drosophila melanogaster*. Cold Spring Harbor Laboratory Press, Plainview, N.Y.

Boettiger, A. N., and M. Levine, 2013 Rapid Transcription Fosters Coordinate snail Expression in the Drosophila Embryo. Cell Reports 3**:** 8-15.

Bownes, M., 1975 A photographic study of development in the living embryo of Drosophila melanogaster. J Embryol Exp Morphol 33**:** 789-801.

Femino, A. M., F. S. Fay, K. Fogarty and R. H. Singer, 1998 Visualization of single RNA transcripts in situ. Science 280**:** 585-590.

Foe, V. E., and B. M. Alberts, 1983 Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis. J Cell Sci 61**:** 31-70.

Geiss, G. K., R. E. Bumgarner, B. Birditt, T. Dahl, N. Dowidar *et al.*, 2008a Direct multiplexed measurement of gene expression with color-coded probe pairs. Nature Biotechnology 26**:** 317-325.

Geiss, G. K., R. E. Bumgarner, B. Birditt, T. Dahl, N. Dowidar *et al.*, 2008b Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol 26**:** 317-325.

Graveley, B. R., A. N. Brooks, J. Carlson, M. O. Duff, J. M. Landolin *et al.*, 2011 The developmental transcriptome of Drosophila melanogaster. Nature 471**:** 473-479.

Gregor, T., E. F. Wieschaus, A. P. McGregor, W. Bialek and D. W. Tank, 2007 Stability and nuclear dynamics of the bicoid morphogen gradient. Cell 130**:** 141-152.

Hansen, K. D., S. E. Brenner and S. Dudoit, 2010 Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Res 38**:** e131.

Lott, S. E., J. E. Villalta, G. P. Schroth, S. Luo, L. A. Tonkin *et al.*, 2011 Noncanonical compensation of zygotic X transcription in early Drosophila melanogaster development revealed through single-embryo RNA-seq. PLoS Biol 9**:** e1000590.

Loven, J., D. A. Orlando, A. A. Sigova, C. Y. Lin, P. B. Rahl *et al.*, 2012 Revisiting Global Gene Expression Analysis. Cell 151**:** 476-482.

Lubeck, E., and L. Cai, 2012 Single-cell systems biology by super-resolution imaging and combinatorial labeling. Nat Methods 9**:** 743-748.

Materna, S. C., J. Nam and E. H. Davidson, 2010 High accuracy, high-resolution prevalence measurement for the majority of locally expressed regulatory genes in early sea urchin development. Gene Expr Patterns 10**:** 177-184.

Nien, C. Y., H. L. Liang, S. Butcher, Y. Sun, S. Fu *et al.*, 2011 Temporal coordination of gene networks by Zelda in the early Drosophila embryo. PLoS Genet 7**:** e1002339.

Petkova, M. D., S. C. Little, F. Liu and T. Gregor, 2014 Maternal Origins of Developmental Reproducibility. Current Biology 24**:** 1283-1288.

Pritchard, D. K., and G. Schubiger, 1996 Activation of transcription in Drosophila embryos is a gradual process mediated by the nucleocytoplasmic ratio. Genes Dev 10**:** 1131-1142.

Raj, A., P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden and S. Tyagi, 2008 Imaging individual mRNA molecules using multiple singly labeled probes. Nature Methods 5**:** 877-879.

Reed, B. H., S. C. McMillan and R. Chaudhary, 2009 The preparation of Drosophila embryos for live-imaging using the hanging drop protocol. J Vis Exp.

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

Roberts, A., C. Trapnell, J. Donaghey, J. L. Rinn and L. Pachter, 2011 Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol 12**:** R22.

Sandmann, T., C. Girardot, M. Brehme, W. Tongprasit, V. Stolc *et al.*, 2007 A core transcriptional network for early mesoderm development in Drosophila melanogaster. Genes Dev 21**:** 436-449.

Stathopoulos, A., and M. Levine, 2005 Genomic regulatory networks and animal development. Developmental Cell 9**:** 449-462.

Tadros, W., and H. D. Lipshitz, 2009 The maternal-to-zygotic transition: a play in two acts. Development 136**:** 3033-3042.

Wu, M. Y., and C. S. Hill, 2009 TGF-beta Superfamily Signaling in Embryonic Development and Homeostasis. Developmental Cell 16**:** 329-343.

Chapter 3

UNCOVERING GENETIC INTERACTIONS AND PLACING GENES IN THE REGULATORY NETWORK USING NANOSTRING

ABSTRACT

Gene Regulatory Networks (GRNs) describe the interactions between genes, transcription factors, and signaling pathways that establish the transcriptional landscape of cells or tissues. They are compiled using a variety of methods, including manipulation of individual genes, studies of the physical interactions of proteins, and obtaining evidence of transcription factors binding to DNA. Data used to build GRNs has traditionally been gathered in experiments for single genes or small numbers of genes, using techniques such as in situ hybridization, Real-Time Quantitative Polymerase Chain Reaction (qPCR), and various proteomic approaches. More recently, high-throughput RNA-seq has allowed the entire transcriptome to be observed simultaneously. The gap between single-gene studies with few data points at a time and ~15,000-gene studies with millions of individual sequencing reads is quite disparaging in terms of type of data produced and the skills needed to analyze the data. Recently, NanoString technology has bridged this gap and allowed quantitative studies of tens to hundreds of genes in parallel. We describe how NanoString technology can be used to study an entire GRN to uncover new genetic interactions, as is the case with the gene *twi*, or place unknown genes in the GRN, with the gene *neu2*.

INTRODUCTION

The process of patterning the early *Drosophila melanogaster* embryo has been the subject of much study, as many signaling pathways characterized in *Drosophila* development are conserved among vertebrate and human disease models (BELVIN AND ANDERSON 1996; ROCK *et al.* 1998; MEDZHITOV 2001; CRIPPS AND OLSON 2002; OLSON 2006; KLAUS AND BIRCHMEIER 2008; GRAVELEY *et al.* 2011b; PANDEY AND NICHOLS 2011). Understanding how many genes work together in patterning tissues, specifying cell fate, or contributing to disease states is aided by the creation of a Gene Regulatory Network (GRN) to map the complex interactions between tens to hundreds of individual genes (DAVIDSON *et al.* 2002). In *Drosophila*, the early embryo is arranged on two major axes: anterior-posterior (AP), specifying head to tail; and dorsal-ventral (DV), specifying back to belly. Maternally deposited signals determine the basic polarities of both axes, with the transcription factors Bicoid (Bcd) and Dorsal (Dl) on the AP and DV axes, respectively, activating the expression of many downstream genes in the early embryo that control development (DRIEVER AND NUSSLEINVOLHARD 1988; ROTH *et al.* 1989). The early embryonic GRN for the DV axis has been well characterized, with the interactions of ~75 genes described in a network model (LEVINE AND DAVIDSON 2005; LONGABAUGH *et al.* 2005; STATHOPOULOS AND LEVINE 2005).

The construction of the DV GRN took many years of research, with most contributions coming from mutating or ectopically expressing single genes and observing the behavior of small numbers of genes in response to the changes. Sequencing the *Drosophila* genome allowed the widespread computational exploration of genes and their regulatory sequences to predict and test more interactions taking place in the GRN and add

additional connections (ADAMS *et al.* 2000; MARKSTEIN AND LEVINE 2002). Additionally, microarray experiments allowed data to be gathered on the expression changes of thousands of genes in parallel in different mutant backgrounds affecting the activity of Dl in activating the DV GRN (STATHOPOULOS AND LEVINE 2002). Following the trend of microarray experiments, high throughput RNA-seq allowed the entire developmental transcriptome to be observed simultaneously, with hundreds of millions of individual sequence reads providing data depth never before achieved (GRAVELEY *et al.* 2011a). RNA-seq has also been used in experiments similar to microarray experiments, by changing the activating input of a signaling pathway and observing gene changes in response (DEIGNAN *et al.* 2016).

There exists a large disparity in the types of data collected from different methods used to build a GRN. While microarrays and RNA-seq can give an overview of how the entire organism or network reacts to a perturbation, they do not focus on the effects to single genes in terms of expression domain or other specific functions. In situ hybridizations, qPCR, and analyses of protein interactions give a very detailed view on a small number of genes, but scaling up to the whole GRN level, let alone the whole genome level is a daunting task. In this study, we expand upon our previous work using NanoString technology to address experimental gaps in probing GRNs and identify new GRN members and connections (SANDLER AND STATHOPOULOS 2016).

NanoString uses fluorescent mRNA barcodes that are designed to bind to mRNAs of specific genes and provide a count of transcripts in a sample (GEISS *et al.* 2008). The technology achieves this without using reverse transcription, library amplification, or library fragmentation, which are used in RNA-seq and qPCR and can introduce bias and

reduce reliability of quantitative data (HANSEN *et al.* 2010; ROBERTS *et al.* 2011). In addition, NanoString can provide this quantitative data for up to 800 genes in parallel from a single sample, drastically reducing the number of experiments needed to obtain the same results with qPCR, and without the computational analysis required of RNA-seq.

MATERIALS & METHODS

Fly stocks. Embryo collection and live imaging was done on flies with a His2Av-RFP fusion [Bloomington Drosophila Stock Center (BDSC) 23650]. *twi-* embryos were obtained using a twi^{1}/CyO stock (BDSC 2381). *Neu2-* embryos were obtained using a PBac{SAstopDsRed}LL06458 P{FRT(w^{hs})}2A P{neoFRT}82B P{Car20y}96E/TM3, Hb LacZ stock (DGRC 141806). For both mutant lines, PCR for LacZ was done on embryos to confirm the absence of the balancer chromosome and the presence of homozygous *twi*$^{-}$ or *Neu2*$^{-}$ mutant chromosomes.

NanoString. NanoString and RNA extraction methods are described in chapter 2.

RESULTS

twist modulates TGF-β signaling at different GRN levels

To test our approach of using NanoString to make new connections and place previously uncharacterized genes in the GRN, we analyzed *twi*$^{-}$ embryos to confirm changes in expression of known target genes. As expected, genes expressed in the mesoderm exhibited greatly reduced expression in the pre-gastrulation blastoderm (see Chapter 2, figure 2.4, (SANDLER AND STATHOPOULOS 2016)). In addition, we noticed that

there were a number of changes in the expression of members of the TGF-β signaling pathway, ranging from pathway inputs to output genes (Figure 3.1 A-E).

All five TGF-β output genes were significantly decreased in *twi-* mutant embryos, expressed at 55%, 54%, and 64% of wild type levels in nuclear cycles 13L, 14C, and gastrulation respectively (Figure 3.1 A, B).

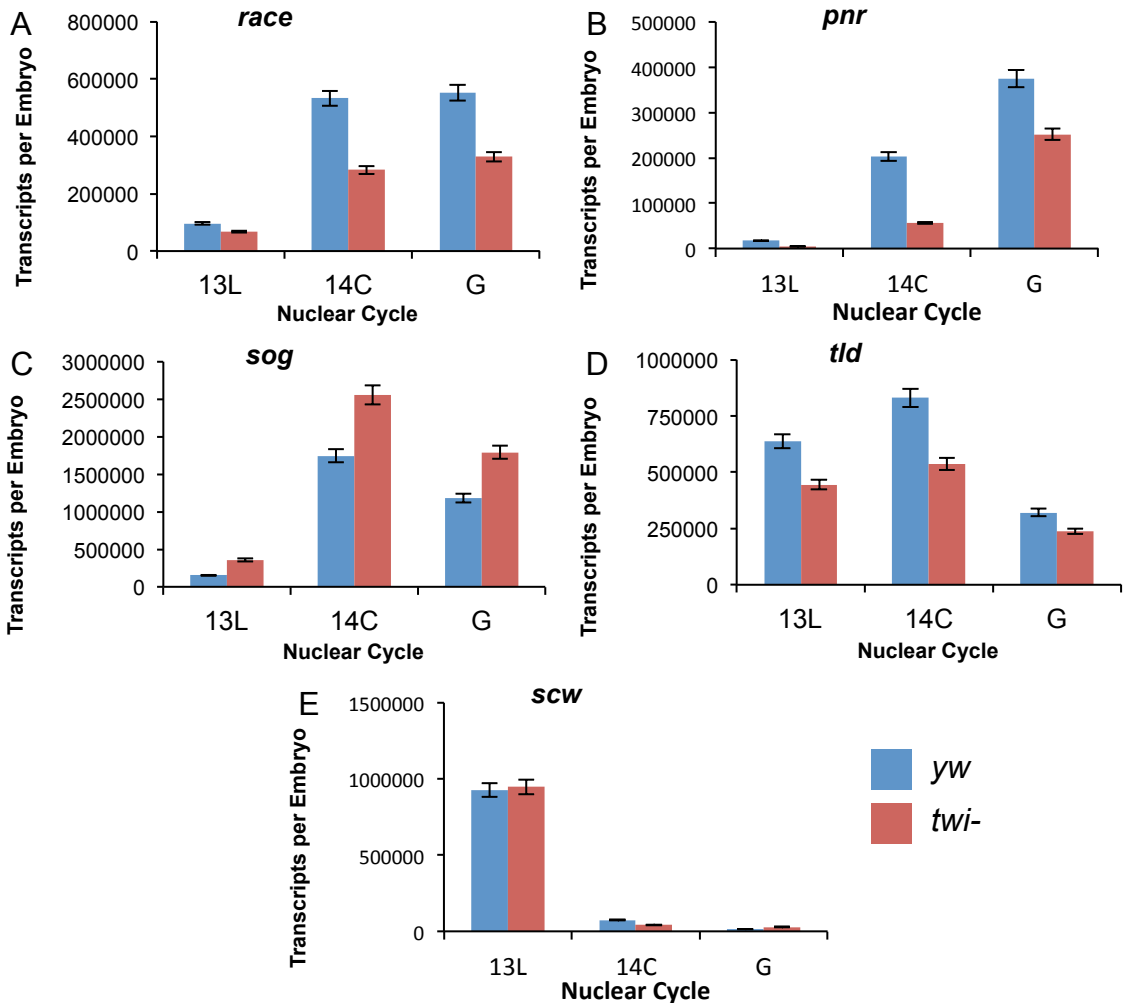


Figure 3.1. TGF-β pathway members in *twi*⁻ embryos. (A-E) Plots comparing number of transcripts counted in *yw* embryos to *twi*⁻ embryos, with *yw* in blue and *twi*⁻ in red. Expression of TGF-β output genes *race* (A) and *pnr* (B) is decreased. TGF-β ligand interacting genes show diverging expression: *sog* (C) is increased, while *tld* (D) is decreased. TGF-β ligand *scw* shows no change.

This drastic change for every output gene signifies that the activity of the entire signaling pathway is repressed. It is interesting though, that expression of the TGF-β ligands themselves, *dpp* and *scw,* is not changed (Figure 3.1 E), indicating that genes involved in processing of ligands or signal transduction are responsible for the differences observed. In fact, *sog*, which binds and sequesters Dpp and Scw, is expressed at 176%, 148%, and 115% of endogenous levels at nuclear cycles 13L, 14C, and gastrulation, respectively (Figure 3.1 C). Finally, *tld*, which cleaves Sog and releases Dpp and Scw to activate TGF-β signaling is expressed at 70%, 65%, and 74% of wild type levels for the same time points listed above (Figure 3.1 D).

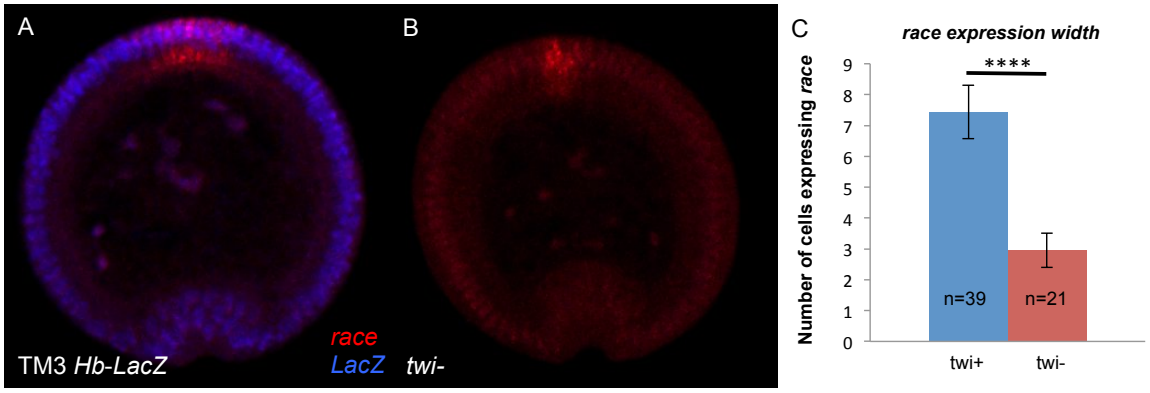


Fig 3.2. *race* expression domain in *twi*¯ embryos. Fluorescent In Situ Hybridization staining for *race* (red) and *LacZ* (blue) in *twi*¯/TM3 *Hb-LacZ* embryos. (A) Heterozygotes with functional *twi* and (B) mutants lacking *twi*. Embryos were chopped at the cephalic furrow to control for narrowing width of expression along the AP axis. (C) Quantification of *race* expression width showing a narrowing of expression in *twi*- embryos. P<0.0001, two-tailed Student's T-Test.

Together, these data paint a clear picture. Twist is an activating transcription factor that drives the expression of many genes in the mesoderm, including the repressor *snail (sna)* (LEPTIN 1991). In *twi*¯ embryos, *sna* expression is greatly decreased (SANDLER AND STATHOPOULOS 2016), which in turn de-represses *sog* in the mesoderm, and expression of *sog* increases (REMBOLD *et al.* 2014). The combination of more *sog* and less *tld* likely leads

to an overall repression of TGF-β signaling, as Dpp and Scw are sequestered by more Sog and released to signal at a lower rate due to lack of Tld to cleave sog. When an in situ hybridization is done with a riboprobe against *race* in *twi*$^{-}$ embryos at gastrulation, it is clear that the expression of *race* is greatly reduced from its normal width of 7-8 cells in the dorsal ectoderm to a restricted 3 cells wide (Fig. 3.2 A-C).

Placing a gene in the GRN

The gene *Neu2* was identified as one of several expressed along the DV axis in a screen comparing expression of genes in embryos with varying levels of Dl (STATHOPOULOS AND LEVINE 2002). *Neu2* expression is increased in *Toll*^{rm9}*/Toll*^{rm10} mutants, which have a low-level concentration of Dl in all nuclei, but no gradient. *Neu2* is normally expressed in narrow ventral-lateral stripes on either side of the embryo, and excluded from the mesoderm. When *Neu2*$^{-}$ embryos were analyzed using NanoString at five time points (NC 13L, NC 14A, NC 14B, NC 14D, and gastrulation), several changes were observed. The expression of many mesoderm genes increased, especially that of *sna*, which is especially sensitive to the width of the Dl nuclear gradient (GARCIA *et al.* 2013), while the expression of other genes, including *sog*, decreased (Figure 3.3 A, B). When FISH is performed in *Neu2*$^{-}$ embryos using a *sna* probe, it is clear that the width of *sna* expression expands signcantly, from a very consistent 20 cells wide in *yw* embryos to over 26 cells wide in *Neu2*- embryos (Figure 3.3 F-H). Interestingly, the expression of TGF-β output genes initially decreased, but recovered to wild type levels by gastrulation (Figure. 3.3 D, E).

These changes are consistent with an expansion of the Dl nuclear gradient. As *sna* is especially responsive to changes in the width of the Dl nuclear gradient, its increased expression is consistent with this expansion. Likewise, *htl,* another mesoderm gene is also transcribed in greater numbers (Figure 3.3 C). In contrast, *sog* expression is decreased from its wild type levels, consistent with an expanded *sna* domain repressing what are usually *sog*-expressing cells.

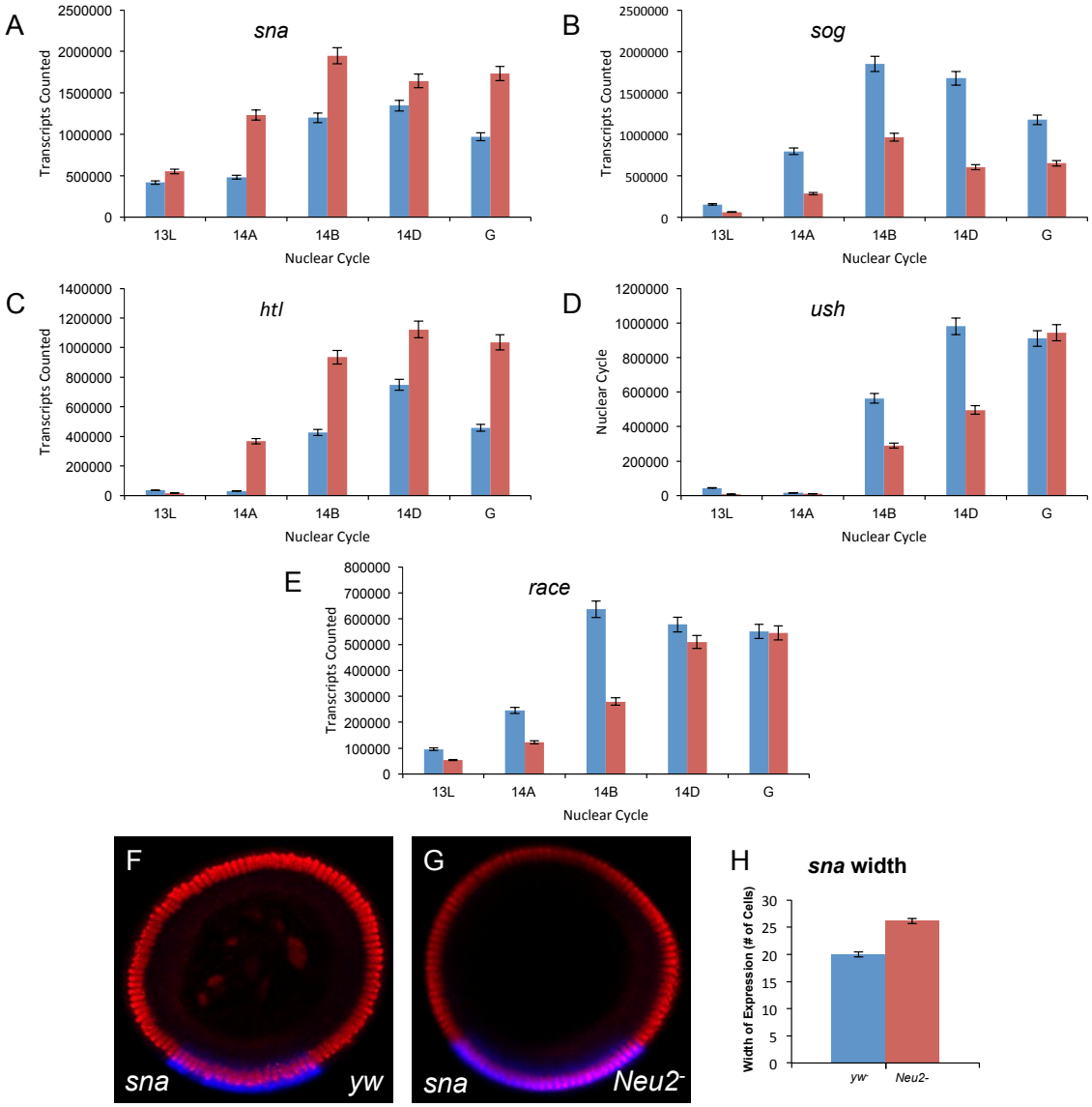


Figure 3.3. Gene expression changes in *Neu2*⁻ embryos. (A-E) Plots comparing number of

transcripts counted in *yw* embryos to *Neu2*$^{-}$ embryos, with *yw* in blue and *Neu2*$^{-}$ in red. Expression of mesoderm genes *sna* (A) and *htl* (C) is increased, while *sog* (B) expression is decreased. TGF-β output genes *ush* (D) and *race* (E) are originally expressed lower than in *yw* embryos, but recover at gastrulation. (F, G) FISH staining shows the expression width of *sna* in *yw* (F) and *Neu2*$^{-}$ (G) embryos. Width is quantified in (H), with error bars representing SEM.

The initial decrease in TGF-β output genes could be caused by a narrower dorsal ectoderm caused by the general dorsal expansion genes. The recovery of these TGF-β genes could be due to a switch in signaling input. At gastrulation, the TGF-β pathway could switch from relying purely on the DV genes to instead being activated by genes not affected by the loss of *Neu2*.

DISCUSSION

We have demonstrated the use of NanoString technology to provide quantitative data on the expression of dozens of genes in parallel and monitor an entire GRN for changes when perturbed. This combination of highly quantitative data and information on dozens of genes is made especially more powerful considering that it was obtained using single embryos. RNA-seq protocols require up to 10μg of RNA input (LIANOGLOU *et al.* 2013), while RNA input to NanoString can be 100ng or less, a 100-fold decrease in starting material and less total RNA than is extracted from one *Drosophila* embryo. The customizable NanoString probe sets also allow for flexible experiments, where the same initial input can yield a wide variety of results based on user directed specifications with no additional experimental effort.

In the case of *twi*, although the function of the gene had been well characterized for many years, NanoString analysis still yielded new insights into how *twi* acts when the

entire GRN is taken into account. Although the connection between *twi, sna,* and *sog* is understood, the involvement of *tld* and *race* is novel. NanoString does not provide positional information, as samples are homologized before RNA extraction, so there were two possibilities for how *race* expression changed. Either the expression domain could have remained unchanged and overall transcription level decreased, or the expression level in cells where *race* was transcribed could have remained unchanged while the expression domain was restricted. A single in situ experiment was able to confirm the later possibility and provide the spatial information not evident with NanoString alone. This combination of quantification of an entire GRN with experiments targeting genes with unexpected behavior is a powerful tool to provide new data on the procession of embryonic development.

The use of *Neu2* as a first attempt at placing an unknown gene into the GRN using NanoString has yielded a testable model that can be used to design specific experiments to fill in gaps in data and answer questions that arise from the analysis. When other data is incorporated with the NanoString results, a clearer picture emerges. As with *twi*¯ embryos, a target FISH experiment for *sna* confirms NanoString data and provides clarity on the level versus expression domain question.

It ha been suggested that Neu2 binds to the cell membrane bound protein Weckle (Wek) through a global two-hybrid screen (GIOT *et al.* 2003). A Neu2 immunopercipitation and western blot for Wek are necessary to conform this preliminary observation. Wek has been shown to facilitate the Toll/Tube complex, which facilitates transport of Dl into nuclei and drives activation of the DV GRN (CHEN *et al.* 2006). It is possible that Neu2 binding to Wek can disrupt the nuclear localization of Dl by disrupting the Wek/Toll/Tube complex.

The expression domain of *Neu2* in two stripes in the neurogenic ectoderm further points to this role, it is expressed in a location where the Dl nuclear gradient rapidly drops off in concentration and many genes are sensitive to very small changes in Dl concentration (REEVES *et al.* 2012). It is possible that Neu2 balances the action of Wek in this critical region to refine the Dl gradient. In *neu2*$^{-}$ embryos, this action of balancing Wek is absent, and Wek may be overactive in facilitating the import of Dl, leading to a widening of the gradient. The NanoString results support this model and help place *Neu2* in the GRN, demonstrating the usefulness of the technique and providing a new path of experiments to design and carry out.

REFERENCES

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne *et al.*, 2000 The genome sequence of Drosophila melanogaster. Science 287**:** 2185-2195.

Belvin, M. P., and K. V. Anderson, 1996 A conserved signaling pathway: the Drosophila toll-dorsal pathway. Annu Rev Cell Dev Biol 12**:** 393-416.

Chen, L. Y., J. C. Wang, Y. Hyvert, H. P. Lin, N. Perrimon *et al.*, 2006 Weckle is a zinc finger adaptor of the Toll pathway in dorsoventral patterning of the Drosophila embryo. Current Biology 16**:** 1183-1193.

Cripps, R. M., and E. N. Olson, 2002 Control of cardiac development by an evolutionarily conserved transcriptional network. Developmental Biology 246**:** 14-28.

Davidson, E. H., J. P. Rast, P. Oliveri, A. Ransick, C. Calestani *et al.*, 2002 A genomic regulatory network for development. Science 295**:** 1669-1678.

Deignan, L., M. T. Pinheiro, C. Sutcliffe, A. Saunders, S. G. Wilcockson *et al.*, 2016 Regulation of the BMP Signaling-Responsive Transcriptional Network in the Drosophila Embryo. Plos Genetics 12.

Driever, W., and C. Nussleinvolhard, 1988 The Bicoid Protein Determines Position in the Drosophila Embryo in a Concentration-Dependent Manner. Cell 54**:** 95-104.

Garcia, H. G., M. Tikhonov, A. Lin and T. Gregor, 2013 Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. Curr Biol 23**:** 2140-2145.

Geiss, G. K., R. E. Bumgarner, B. Birditt, T. Dahl, N. Dowidar *et al.*, 2008 Direct multiplexed measurement of gene expression with color-coded probe pairs. Nature Biotechnology 26**:** 317-325.

Giot, L., J. S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang *et al.*, 2003 A protein interaction map of Drosophila melanogaster. Science 302**:** 1727-1736.

Graveley, B. R., A. N. Brooks, J. Carlson, M. O. Duff, J. M. Landolin *et al.*, 2011a The developmental transcriptome of Drosophila melanogaster. Nature 471**:** 473-479.

Graveley, B. R., A. N. Brooks, J. W. Carlson, M. O. Duff, J. M. Landolin *et al.*, 2011b The developmental transcriptome of Drosophila melanogaster. Nature 471**:** 473-479.

Hansen, K. D., S. E. Brenner and S. Dudoit, 2010 Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Res 38**:** e131.

Klaus, A., and W. Birchmeier, 2008 Wnt signalling and its impact on development and cancer. Nature Reviews Cancer 8**:** 387-398.

Leptin, M., 1991 twist and snail as positive and negative regulators during Drosophila mesoderm development. Genes Dev 5**:** 1568-1576.

Levine, M., and E. H. Davidson, 2005 Gene regulatory networks for development. Proceedings of the National Academy of Sciences of the United States of America 102**:** 4936-4942.

Lianoglou, S., V. Garg, J. L. Yang, C. S. Leslie and C. Mayr, 2013 Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes Dev 27**:** 2380-2396.

Longabaugh, W. J. R., E. H. Davidson and H. Bolouri, 2005 Computational representation of developmental genetic regulatory networks. Developmental Biology 283**:** 1-16.

Markstein, M., and M. Levine, 2002 Decoding cis-regulatory DNAs in the Drosophila genome. Current Opinion in Genetics & Development 12**:** 601-606.

Medzhitov, R., 2001 Toll-like receptors and innate immunity. Nature Reviews Immunology 1**:** 135-145.

Olson, E. N., 2006 Gene regulatory networks in the evolution and development of the heart. Science 313**:** 1922-1927.

Pandey, U. B., and C. D. Nichols, 2011 Human Disease Models in Drosophila melanogaster and the Role of the Fly in Therapeutic Drug Discovery. Pharmacological Reviews 63**:** 411-436.

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

Rembold, M., L. Ciglar, J. O. Yanez-Cuna, R. P. Zinzen, C. Girardot *et al.*, 2014 A conserved role for Snail as a potentiator of active transcription. Genes & Development 28**:** 167-181.

Roberts, A., C. Trapnell, J. Donaghey, J. L. Rinn and L. Pachter, 2011 Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol 12**:** R22.

Rock, F. L., G. Hardiman, J. C. Timans, R. A. Kastelein and J. F. Bazan, 1998 A family of human receptors structurally related to Drosophila Toll. Proceedings of the National Academy of Sciences of the United States of America 95**:** 588-593.

Roth, S., D. Stein and C. Nussleinvolhard, 1989 A Gradient of Nuclear-Localization of the Dorsal Protein Determines Dorsoventral Pattern in the Drosophila Embryo. Cell 59**:** 1189-1202.

Sandler, J. E., and A. Stathopoulos, 2016 Quantitative Single-Embryo Profile of Drosophila Genome Activation and the Dorsal-Ventral Patterning Network. Genetics 202**:** 1575-+.

Stathopoulos, A., and M. Levine, 2002 Dorsal gradient networks in the Drosophila embryo. Dev Biol 246**:** 57-67.

Stathopoulos, A., and M. Levine, 2005 Genomic regulatory networks and animal development. Developmental Cell 9**:** 449-462.

Chapter 4

A DEVELOPMENTAL PROGRAM TRUNCATES LONG TRANSCRIPTS TO TEMPORALLY REGULATE CELL SIGNALING

ABSTRACT

During development, rapid mitotic divisions and a fixed transcription rate limit the maximal length of transcripts. In *Drosophila* embryos, previous studies suggested that transcription of long genes is initiated but aborted in early divisions with short interphases of 15 minutes or less. Here we identify that long genes are expressed during short nuclear cycles but as truncated transcripts. The RNA binding protein Sex-lethal is required to specifically support transcription termination of these short transcripts as it associates with truncated but not full-length forms. Furthermore, one short product of a truncated transcript for the gene *short-gastrulation* relates closely to a previously characterized dominant negative form that retains TGF-β signaling in the off-state. In summary, our results reveal a developmental program of short transcripts and concomitant protein products that helps prime the *Drosophila* embryo, keeping signaling at earlier stages to a minimum to support proper timing of cell signaling initiation at cellularization.

INTRODUCTION

Early embryonic development of the fruit fly *Drosophila melanogaster* has 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 2016). Transcription is aborted during mitosis between NCs, and nascent transcripts are degraded, with transcription restarting at interphase of the following NC (SHERMOEN AND OFARRELL 1991).

As the rate of transcription in *Drosophila* has been measured at ~1.1-1.5kb per minute of interphase, including in recent studies using MS2-MCP live imaging of nascent transcripts (ARDEHALI AND LIS 2009; GARCIA *et al.* 2013), transcription of zygotic genes during syncytial NCs is likely time constrained. In support of this view, 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.* 2015). It was thought that long genes, those over 20kb, are either not transcribed before NC14 or aborted mid-transcript, and no protein products were present (OFARRELL 1992). This observation is exemplified in the pair of duplicated adjacent genes *knirps (kni)* and *knirps-like (knrl)*. *kni* is 3kb long, while *knrl* is over 23kb long and has introns totaling over 19kb, yet the functional domains of the two proteins are almost identical. While *kni* is expressed as early as NC 11, with an interphase of around 10 minutes, *knrl* is expressed only in late NC 14, with an interphase of over 45 minutes that permits the presence of the full-length transcript (ROTHE *et al.* 1992).

Recently, studies have produced evidence that some long genes are transcribed during early NCs (LOTT *et al.* 2011; ALI-MURTHY *et al.* 2013; SANDLER AND STATHOPOULOS 2016). To explore such questions, 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 (Figure 1B).

RESULTS

Long Transcripts Are Truncated During Short Nuclear Cycles

Using an available RNA-seq dataset of *Drosophila* development, we selected four long genes, *short gastrulation (sog), Netrin-A (NetA), scabrous* (s*ca*)*,* and *Protein kinase cAMP-dependent catalytic subunit 3 (Pka-C3),* with evidence of transcription during NC13 (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, to search for alternate transcript isoforms. Only the previously defined 5' transcription start sites were recovered (GRAVELEY *et al.* 2011b), leading us to conclude that alternate start sites are not used for these genes, whereas 3' RACE products identified truncations in these four transcripts (Figure 4.1A). The short forms aligned to annotated transcripts at the beginning of the full-length genes, but ended with an alternate exon, including coding sequence and a 3' UTR, in what is usually an endogenous intron. The RACE products were all poly-adenylated, with no poly-A sequence in the genome at the locus of alignment.

To distinguish between full-length transcripts and short forms, we designed fluorescence in situ hybridization (FISH) riboprobes at the 5' and 3' ends of *sog, NetA, sca,*

and *Pka-C3*, with 3' probes covering full-length forms only (Figure 4.1A). In all cases, there was no observable signal from the 3' probes during NC13, while signal from the 5' probes was present, indicating that transcription did not reach the 3' ends of genes assayed (Figure 4.1 C, D, G, I, K-N). In contrast, full-length transcripts were present in NC14 when the duration of interphase was permissive (Figure 1 E, F, H, J, K-N).

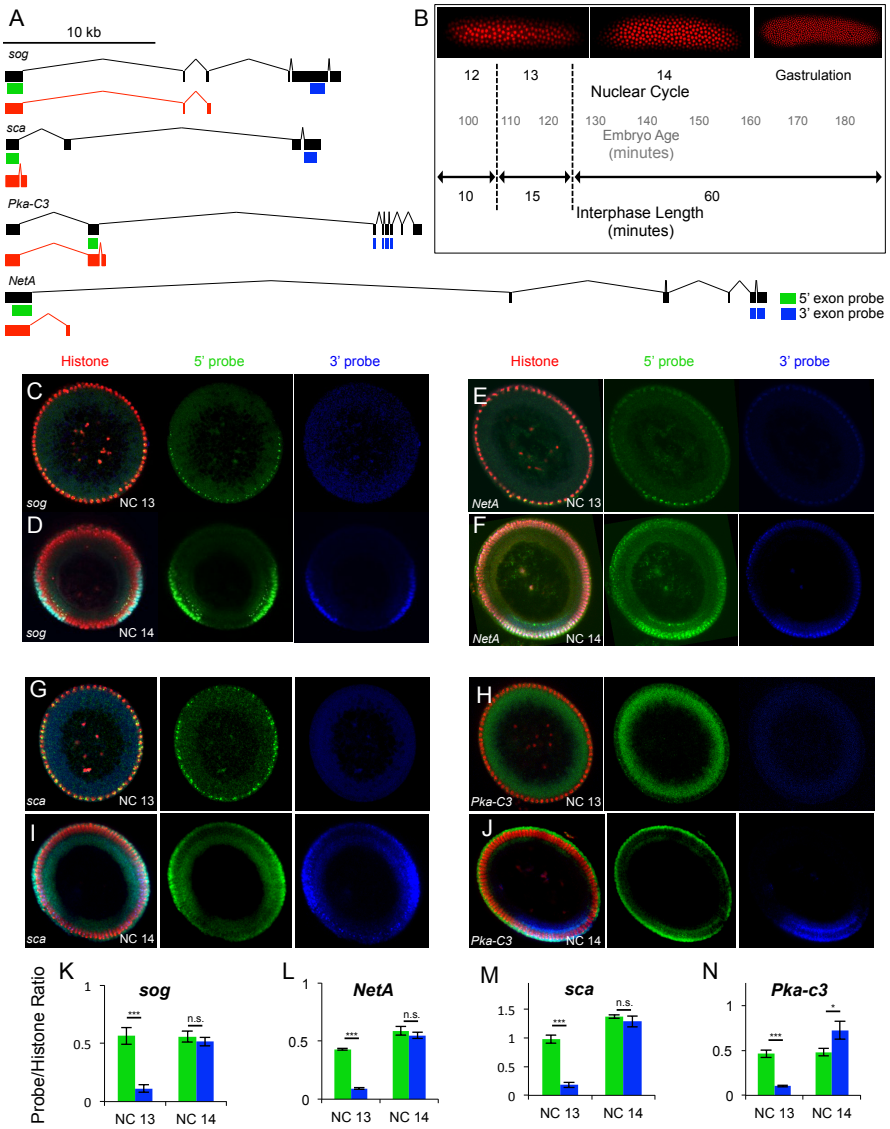


Figure 4.1. Long genes are transcribed as short forms in NC13. (A) Full-length transcripts (black) and mapped 3' RACE identified shorter transcripts (red) for each of the long genes investigated. Locations of 5' and 3' FISH riboprobes shown in green and blue, respectively. (B) A timeline of the syncytial 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* showing full-length transcript is present in NC14, but not NC13. Images depict manually chopped embryo cross-sections, as described in the methods. (K-N) Normalization of 5' and 3' FISH riboprobe stainings of genes *sog, NetA, sca,* and *Pka-C3* to immunostained Histone H3 to compare signal intensity. Differences are present for all genes in NC13 (Data are presented as means ± SEM. *** $p<0.0001$, * $p<.01$, two-tailed Student's T-Test). (See also Figure S4.3)

The RNA Binding Protein Sex-Lethal Controls Transcript Truncation

Since the short transcripts uncovered included intron-derived coding sequence, it is likely that transcriptional regulation by RNA binding proteins (RBPs) is a cause of truncation, as opposed to post-translational cleavage. We used two search algorithms to find RBP recognition sites in introns within 1kb downstream of truncations identified by 3' RACE (RAY *et al.* 2013; PAZ *et al.* 2014). A list of binding sites was compiled with the criteria that sites must be found within 1kb of the truncation point for all four genes and, the RBPs must be present in the early embryo or maternally deposited in the developing oocyte. Using Gal4-mediated RNAi to knockdown transcript levels specifically during late stages of oogenesis and in the early embryo (STALLER *et al.* 2013), we assayed a role for the RBP Sex-lethal (Sxl) based on presence of putative binding sites (Figure 4.2G; (PENALVA AND SANCHEZ 2003) in all four long genes, and PPS and U1 snRNP (70k subunit) based on evidence of physical association with Sxl (PENALVA AND SANCHEZ 2003; JOHNSON *et al.* 2010). To characterize RNAi phenotypes, riboprobes were used to assay transcription of intronic sequences 3' of the initially defined truncation sites. In embryos with RNAi against *sxl, PPS,* and *U1 snRNP,* intronic FISH signal past the truncation point was observed during NC13 for *sog, NetA, sca,* and *Pka-C3*, indicating that transcriptional read-through past the truncation point occurs

(Figures 4.2A and S4.1A-K). There was no intronic signal detected for the same probes in wildtype or UAS-RNAi negative controls (Figures 4.2B and S4.1L), indicating that transcription does not normally reach this point during NC 13.

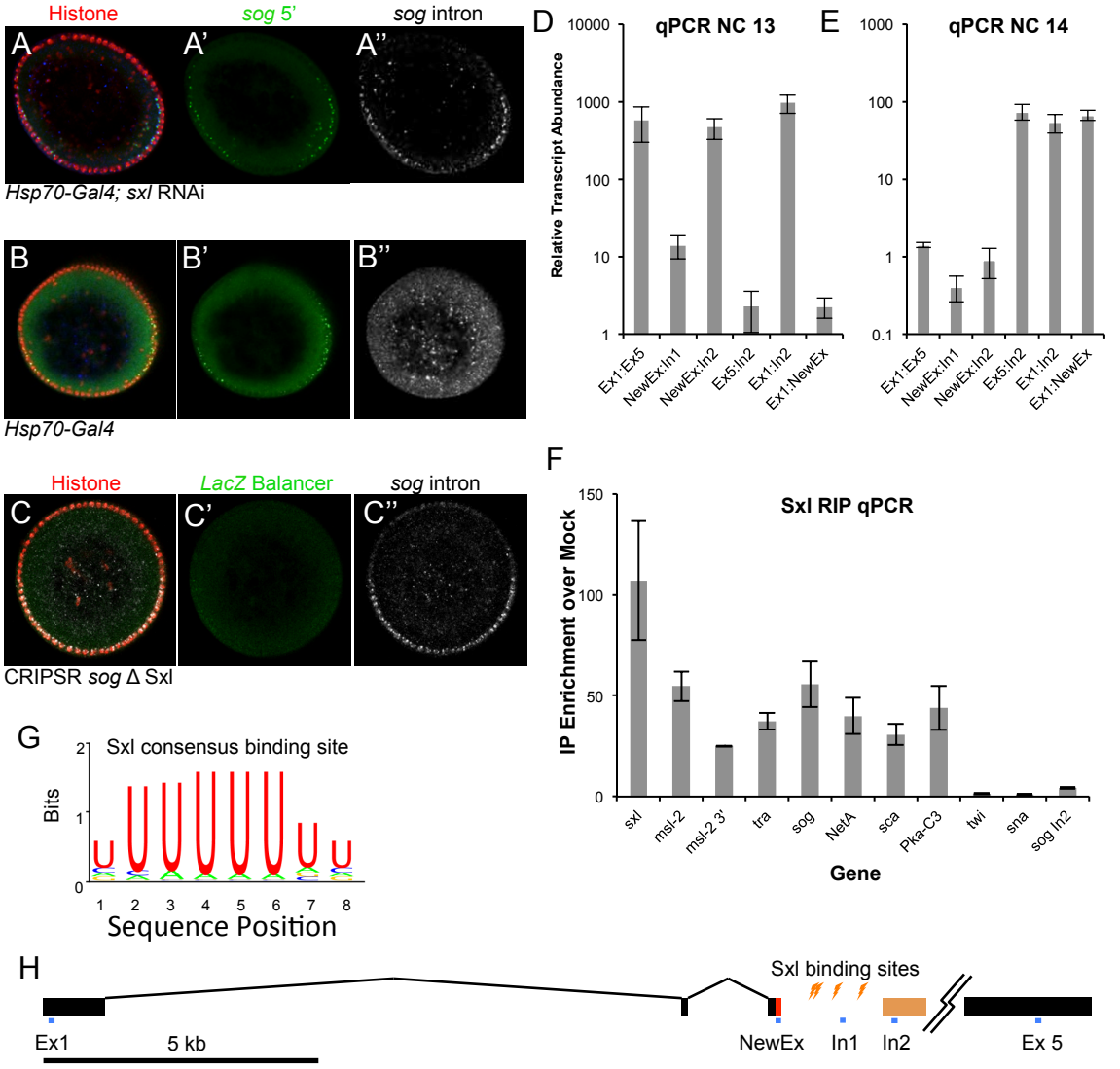


Figure 4.2. Sxl binds to truncated transcripts of long genes. (A) RNAi against *sxl*, showing signals for 5' FISH riboprobe in green, intronic FISH riboprobe in white, and anti-Histone H3 in red used for embryo staging. 5' Probe location shown in (Figure 1A), 3' probe location shown in (H) as orange box. (B) *sxl* RNAi negative control, probes as in (A). (C) FISH riboprobe for the *sog* intron downstream of Sxl binding sites shows transcriptional read-through when Sxl site cluster is deleted using CRISPR-Cas9 system. (D,E) qPCR using primer pairs along the *sog* locus comparing NC13 with NC14, respectively. Location of primers within *sog* locus are shown in (H). Data are presented as means ± SEM. (F) Quantification of enrichment of transcripts

immunoprecipitated with Sxl compared to mock IP, with previously known Sxl-bound genes *sxl, msl-2,* and *tra,* long genes *sog* (including intronic primer set In2 past truncation point)*, NetA, sca*, and *Pka-C3*, and dedicated short genes *twi* and *sna*. Data are presented as means ± SEM. (G) Consensus Sxl binding site (PENALVA AND SANCHEZ 2003) found in the introns of known Sxl-bound genes and long genes. (H) A schematic of the *sog* locus, showing the locations of qPCR primer sets, intronic FISH probe, and Sxl binding site cluster. All images are chopped embryo sections as described in supplemental methods. (See also Figure S4.1)

Furthermore, within the *sog* intron, a Sxl binding site cluster composed of four sites is located ~100-250 bp downstream of the truncation point identified by 3' RACE (Figure 4.2H). When this sequence was deleted using CRISPR-Cas9, transcriptional read-through past the truncation point was observed (Figure 4.2C), providing further evidence that Sxl plays a key role in truncation. In addition, Sxl, PPS, and U1 snRNP have been shown to form a protein complex that regulates splicing (JOHNSON *et al.* 2010), and a subunit of the U1 snRNP complex physically binds to RNA Polymerase II (MORRIS AND GREENLEAF 2000). Together these observations suggest a transcriptional and/or splicing-coupled mechanism involving the modulation of RNA polymerase dynamics with protein complexes assembled at the Sxl sites in nascent mRNA.

Sex-Lethal Directly Binds to Truncated Transcripts

If Sxl binds to RNA, the clusters of Sxl consensus binding sites (Figure 4.2H) must be transcribed for Sxl to act. Using qPCR primer sets spaced along the *sog* locus (Figure 4.2H, blue markers), we found that during NC14, the 5' and 3' exons of *sog* (probes Ex1 and Ex5) were expressed at approximately equivalent levels, while intronic probes were expressed an average of ~65-fold lower than the coding exons (Figure 4.2E), likely representing rapid excision of introns. During NC13, Ex1 is expressed ~600-fold

higher than Ex5, indicating that Ex5 is not transcribed, consistent with FISH observations (Figure 4.2D). Also during NC13, the novel coding region of truncated *sog* (probe NewEx) and the Sxl binding site cluster (probe In1) were expressed ~2-fold and ~10-fold lower than the Ex1 respectively, compared to a ~65-fold decrease for the same sequences during NC14 (Figure 4.2E), providing evidence that this section of the intron is retained during NC13 instead of being spliced out (Figure 4.2D). There was a marked difference between the three intronic probes during NC13; transcript abundance for probe In2 compared to NewEx decreased ~500-fold, even though they were equivalently transcribed during NC 14 (Figures 4.2D, E). This decrease is similar to the difference between Ex1 and Ex5 during NC13, suggesting a true truncation or absence of transcript after the Sxl binding sites (Figure 4.2D).

To assess a physical Sxl association with truncated transcripts, we immunoprecipitated Sxl protein and performed qPCR on eluted RNA. We found that mRNAs of the positive control genes *sxl, msl-2,* and *tra*, which are known to be bound and spliced by Sxl (PENALVA AND SANCHEZ 2003), were enriched an average of ~56-fold compared to a mock IP (Figure 4.2F). Transcripts of *sog, NetA, sca,* and *Pka-C3* were enriched an average of ~42-fold over mock IP (Figure 4.2F). This result, in combination with the presence of Sxl binding sites in the transcripts for these genes, strongly indicates that Sxl binds to all four mRNAs found to be truncated. The negative control genes *twi* and *sna* (short genes without long forms) and *sog* In2 (qPCR probe 3' of the cluster of Sxl binding sites; Figure 4.2H) were not significantly enriched in Sxl IP compared to Ubx mock IP (Figure 2F), indicating no Sxl binding to mRNA.

Protein Products of Short Transcripts Are Functional in Signaling Pathways

Next we investigated whether short products code for peptides that are functional in their normal signaling pathways. Of particular interest is that the short form of Sog contains the entire first cysteine-rich domain, which binds and sequesters TGF-β ligands Decapentaplegic (Dpp) and Screw (Scw) (Figure 4.3A,B) (MARQUES *et al.* 1997). This short form closely resembles a Sog fragment known as Supersog, both in structure and function (see below). However, Supersog was hypothesized to arise from proteolytic cleavage of full-length Sog (YU *et al.* 2000), which is not likely the case during NC13 because the majority of *sog* transcripts are expressed as truncated forms. While full-length Sog is cleaved by the protease Tolloid (Tld) to release the ligands for signaling, short Sog does not contain Tld cleavage sites (PELUSO *et al.* 2011), and may bind Dpp-Scw irreversibly (Figure 4.3B). To test this idea, we assayed the effect of ectopic expression of short Sog on the TGF-β signaling target gene *race* [Figure 4.3C,C'; (RUSCH AND LEVINE 1997)]. We placed the short *sog* cDNA under control of the *even-skipped* (*eve*) stripe 2 enhancer as has previously been done for full-length *sog* (ASHE AND LEVINE 1999), producing a stripe of expression along the anterior-posterior axis in addition to endogenous full-length expression present in a broad lateral domain (Figure 3F). In these embryos, expression of the TGF-β target gene *race* is lost within the trunk and retained only in a small patch at the anterior end of the dorsal ectoderm (Figure 4.3D,D' compare with C,C'), similar to embryos lacking functional Dpp, since only the trunk expression, but not anterior domain, is TGF-β signaling-dependent (XU *et al.* 2005). This indicates that the short Sog peptide acts as a dominant negative repressor unable to release Dpp and Scw to signal once bound.

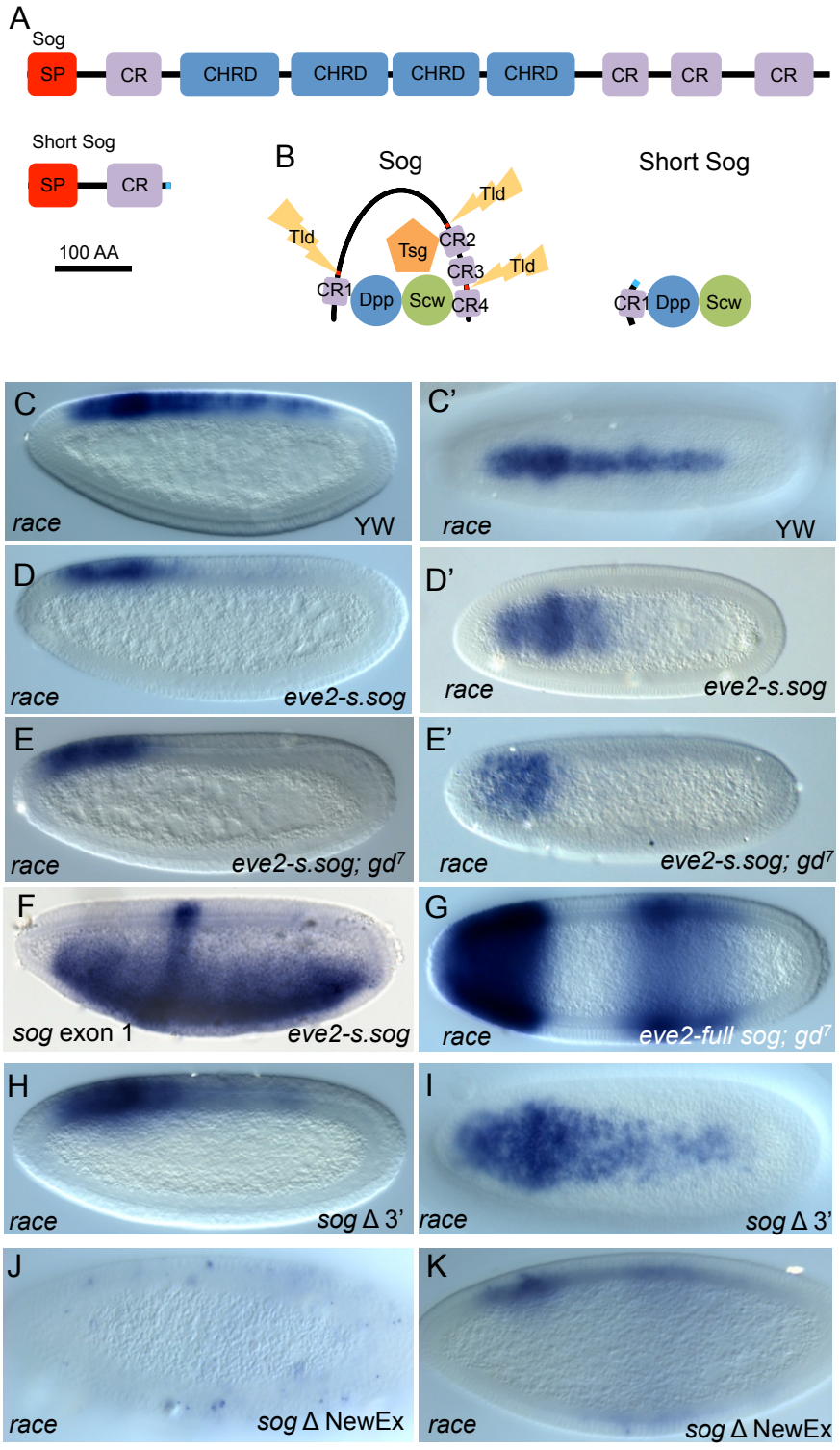


Figure 4.3. Short Sog is a functional dominant negative protein. (A) Full-length and short Sog proteins, showing functional domains (SP= Signal Peptide, CR= Cysteine Repeat, CHRD= Chordin) encoded by long and short transcripts, respectively. New intron-derived amino acids (short Sog only) in blue. (B) A model of full-length Sog and short Sog function. Tld cleavage sites in red are absent in short Sog protein. (C, D, E, F, H, J, and K) are lateral images, (C', D',

E', G, and I) are dorsal images of embryos at NC14.(C) *race* expression in wildtype embryos. (D) *race* expression in embryos expressing short sog cDNA under the control of *eve* stripe 2 enhancer. (E) Same as in (D), except in a gd^7 mutant background, lacking a Dorsal gradient concomitant with expanded Dpp domain throughout the embryo.(F) Expression of short *sog* cDNA ("s. sog") under the control of *eve* stripe 2 enhancer as well as endogenous *sog*. (G) *race* expression in embryos with full-length *sog* cDNA ("full sog") under the control of *eve* stripe 2 enhancer in a gd^7 mutant background. (H and I) In situ staining for *race* in a P-element insertion line that mimics the short *sog* truncation ("sogΔ3'"). (J and K) In situ staining of separate embryos for *race* in a line where the novel coding sequence of short *sog* (NewEx) was deleted using CRISPR ("sogΔNewEx"). Stage of embryo in (J) is early NC14 before dorsal *race* expression appears in wildtype. Embryo in (K) is in mid NC14, when full-length *sog* can support *race* expression. (See also Figures S4.2, S4.4)

We also expressed *eve* stripe 2-short *sog* in a *gastrulation-defective (gd)* background, which lacks a Dorsal nuclear gradient due to defective Toll signaling and expresses expanded domains of the TGF-β ligand Dpp throughout the embryo (KONRAD *et al.* 1998). In these embryos, as shown previously (ASHE AND LEVINE 1999), when the full-length cleavable Sog peptide is expressed in the *eve* stripe 2 domain, robust *race* expression is observed as stripes both in the anterior and mid-trunk regions (Figure 4.3G). In the case of short *sog* expressed in the *eve* stripe 2 domain, *race* expression in most embryos was limited to a broad anterior patch, the expression domain that is TGF-β independent (Figure 4.3E,E'), but was absent from the trunk. Tld cleavage of full-length Sog is concomitant with release of ligands at a distance from the source of Sog ectopic expression and expression of *race* 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 4.3E,E') suggests short Sog cannot be cleaved by Tld to support activation of signaling, and that binding of short Sog to Dpp and Scw is irreversible.

In mutant backgrounds affecting the *sog* locus by P-element insertion, which causes loss of full-length Sog but retention of short Sog (effectively "short Sog only"),

we detected expansion of *race* (Figure 4.3 H and I), presumably because of an inability of short Sog to concentrate ligands at the dorsal-most position, which is usually facilitated by Tld cleavage of full Sog in the dorsal ectoderm (ASHE AND LEVINE 1999). On the other hand, precocious and sporadic activation of *race* throughout the embryo is associated with deletion via CRISPR specifically of the novel short Sog coding sequence in the *sog* intron (effectively "long Sog only"), suggesting expression of this dominant negative version is important to keep cell signaling in check before cellularization, when the ligands Dpp and Scw are widely expressed throughout the embryo and free from Short Sog sequestration to activate signaling in the mutant (Figure 4.3 J and K).

When the short peptides for NetA, Sca, and Pka-C3 were compared with full-length forms, only a subset of functional domains were encoded in a similar fashion to short Sog, suggesting the short forms of these genes, also, correspond to functional truncated proteins, as identified for Sog (HU *et al.* 1995; YU *et al.* 2000; SCHNEIDERS *et al.* 2007; MILOUDI *et al.* 2016) (Figure 4.3A, Figure S4.2). We hypothesize that short forms made at NC13 correspond to dominant negative (or possibly constitutively active) forms of signaling molecules that generally impact signaling, and investigated a possible programmatic truncation of long genes during NC13 using a global approach to map 3' ends of transcripts.

Global 3' RNA-Seq Identifies Additional Truncated Transcripts

To provide insight, 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.* 2013b)]. When comparing both short and long genes, there is a marked difference between the two classes. While there is little difference in 3' transcript ends between NCs 13 and 14 for short genes, such as *knirps* (*kni*) (Figure 4.4A), long genes show large differences in 3' transcript abundances between the two NCs (Figure 4.4A). Our previous study using NanoString to quantify transcripts in the early embryo (SANDLER AND STATHOPOULOS 2016), including *sog* and *NetA*, also showed a difference in 5' vs 3' transcript abundance before NC14, confirming the results from 3' RNA-seq (Figure S4.3). In addition to confirming the lack of full-length NC13 expression for long genes already characterized, 3' RNA-seq provides evidence for additional long genes with truncations during NC13. For example, the embryonic gene *grainy head* (*grh*) is 38kb long, but has a peak of mapped reads less than 15kb from the transcription start site during NC13, with proximal Sxl binding sites located just downstream (Figure 4.4A and B).

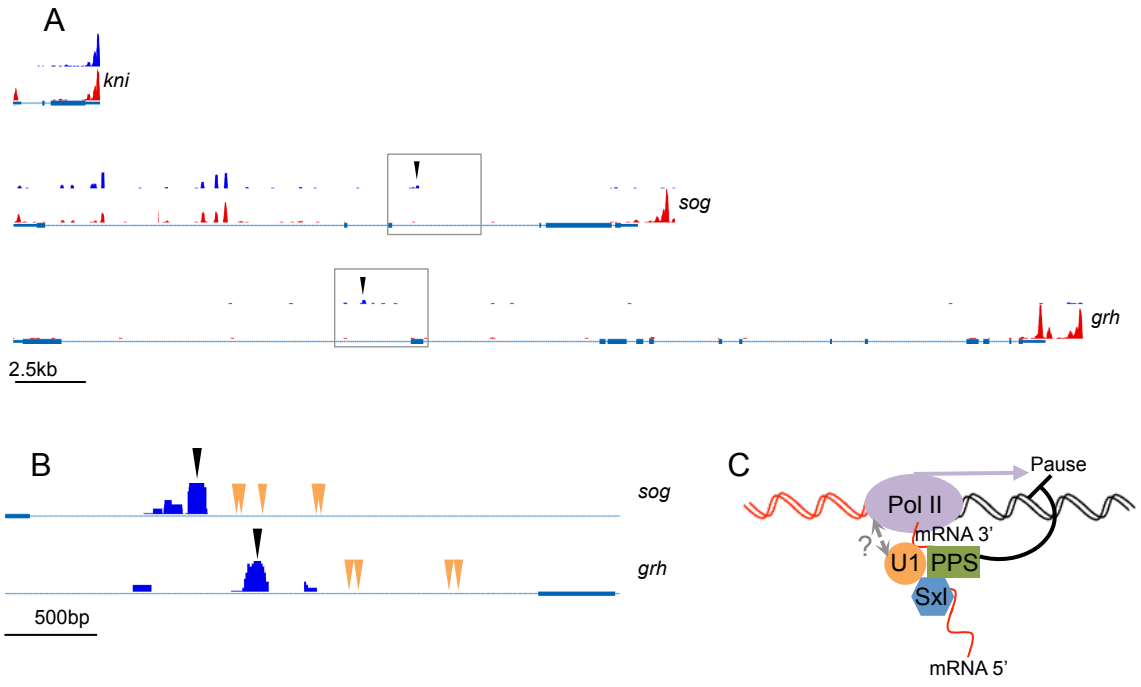


Figure 4.4. 3' RNA-seq and model for transcript truncation. (A) 3' RNA-seq showing NC13 (blue) and NC14 (red) for genes *kni* (3kb), *sog* (23kb), and *grh* (38kb). (B) Detailed browser tracks showing 3' RNA-seq short forms for *sog* and *grh*, from boxed regions in (A). Black arrows indicate novel truncated 3' transcript ends, and orange arrows indicate location of Sxl binding sites, defined as stretches of eight Us in a row in the resulting transcript. (C) A model of the Sxl-PPS-U1 snRNP complex bound to an actively transcribed mRNA and closely associated with RNA Pol II. Sxl binds directly to its RNA binding site, U_8, and is joined by U1 snRNP, which has been shown to also bind RNA Pol II, and PPS. This spliceosome-Pol II interaction may cause a short pause that allows truncation of mRNA. RNAi against any one member of this complex may break this interaction and allow for read-through of a transcript, now bound to be degraded.

DISCUSSION

The need to temporally regulate the initiation of signaling pathways in the embryo is critical for proper development; when members of signaling pathways are either ectopically expressed or knocked out, there are serious and often fatal developmental defects in the embryo (DECOTTO AND FERGUSON 2001). Rapid nuclear divisions limit transcript length of key signaling pathway members, but the truncation of these transcripts to produce short proteins is a mechanism used to resolve this temporal challenge. Sxl, which we demonstrated is essential for the truncation of short transcripts, is an RNA splicing factor and the master gender regulator in *Drosophila* (CLINE 1979; BELL *et al.* 1988). It is differentially expressed in males and females, and is gender-specifically spliced in females to begin a cascade of alternate splicing that determines the gender of the embryo (BOPP *et al.* 1991). Sxl has an ancestral splicing function that is not involved in gender determination, but this is overshadowed by its gender determination function (CLINE *et al.* 2010; EVANS AND CLINE 2013). This ancestral role may be further masked by the fact that *sxl* mutations are lethal or give rise to sterile females (BERNSTEIN AND CLINE 1994), making the observation of early embryonic phenotypes very difficult.

Furthermore, the shortening of transcripts and 3' UTRs has been implicated in the activation of oncogenes and the progression of cancer (MAYR AND BARTEL 2009). Our study provides new insights into these processes and suggests targets for future research in the understanding of alternate transcript truncations.

Even with the recently published fastest rate for RNA Pol II in *Drosophila* embryos of 2.4 kb/min from an analysis of heterologous reporter genes of ~5kb in length (FUKAYA *et al.* 2017), transcription of a 38kb gene during NC13 within 15 min would be challenging, while expression of the short transcripts (<15kb) we have identified would be easily achievable. Our study elucidates a short transcript program that is regulated by Sxl during early NCs, and these short transcripts likely support important developmental roles, which are even more important if they still function in the presence of full-length counterparts to balance cell signaling pathway activation. Targeted PCR results support the view that both short and long forms of genes may be co-expressed in the early embryo at NC14 (Figure S4); however, the predominant transcript formed corresponds to long form at this later stage. The presence and function of short forms when long forms are also expressed and not time-restricted provides more evidence that the balance of short and long forms is important for proper regulation of cell signaling.

In conclusion, this study provides evidence for two previously unidentified processes taking place during development of *Drosophila* embryos: first, the expression of short transcripts that produce functional proteins during short nuclear cycles; second, Sxl functions more broadly than previously understood in the regulation, and importantly, termination of transcription and production of short transcripts. We suggest that this program of short transcripts and resulting proteins is essential for the timing and

coordination of signaling pathways during development, forming an additional regulatory level incorporated into the maternal to zygotic transition during embryogenesis. As an example, truncated Sog serves to regulate both spatial and temporal activation of the TGF-β signaling pathway when the broad expression of ligands cannot achieve this task. Furthermore, our data show that Sxl has an additional, general role outside of specifying gender in the early embryo. This can be thought of as rescuing or protecting essential short transcripts from degradation to ensure the presence of short proteins. 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 subsequently evolved to balance fast development with proper activation of cell signaling.

AUTHOR CONTRIBUTIONS

J.E.S. and A.S. designed the experiments. J.E.S. performed RACE, in situ hybridizations, RNAi, qPCR, RNA-seq, and transgenic embryo experiments. J.I. performed CRISPR and established CRISPR fly stocks, and V.S. performed immuprecipitations. J.E.S. analyzed data and H.A. performed computational analysis on RNA-seq data. J.E.S. and A.S. wrote the manuscript with contributions from J.I., V.S., and H.A.

ACKNOWLEDGMENTS

We thank Brian Williams and Igor Antoshechkin for sequencing support through the Millard and Muriel Jacobs Genetics and Genomics Laboratory at California Institute of Technology. We thank Christine Mayr for advice in implementing the 3' RNA-seq

protocol, and Howard Lipshitz and Don Rio for helpful discussions. This study was supported by NIH grants R01GM077668 and R35GM118146 to A.S.

MATERIALS AND METHODS

Fly stocks and husbandry

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), *U1 snRNP 70k* RNAi P{TRiP.HMS00274}attP2 (BDSC#33396), *PPS* RNAi P{TRiP.GL00684}attP40 (BDSC#38912), *sog* P-element disruption w^{67c23} P{GSV2}GS51273 (Kyoto Stock Center#207284), gd^{7} (BDSC #3109), and *eve* Stripe 2: *sog,* a gift from Hilary Ashe (ASHE AND LEVINE 1999). For CRISPR-Cas9 mediated genome editing flies are described in the sections below. All flies were reared under standard conditions at 23°C. *yw* background was used as wildtype unless otherwise noted.

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 (RACE) or individually collected and staged embryos (NanoString, qPCR, 3' RNA-seq) using Trizol reagent (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 Trizol. A Histone H2Av-RFP fusion was used to stage individual embryos by nuclear cycle using an epifluorescence microscope (SANDLER AND STATHOPOULOS 2016). Individual embryos

were imaged to confirm correct nuclear cycle, snap-frozen in Trizol using liquid nitrogen, and stored at -80° C until RNA extraction. The standard Trizol protocol was followed, with the addition of a second chloroform extraction and second 70% EtOH wash.

Rapid amplification of cDNA ends (RACE) to map transcripts

RACE cDNA libraries were created using the GeneRacer kit (ThermoFisher). Standard protocol was followed, and reverse transcription was done 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 alternate 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 2016). 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 each embryo before extraction at a dilution of 1:10000. 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 following published methods (KOSMAN *et al.* 2004) but omitting Proteinase K treatment. 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.

RNA binding protein site search

We used two search algorithms to find RNA binding protein (RBP) recognition sites in introns downstream of truncations identified by 3' RACE (RAY *et al.* 2013; PAZ *et al.* 2014). A list of binding sites was compiled with the criteria that sites must be found within 1kb of the truncation point for all four genes and the RBPs must be present in the early embryo, by presence of maternally deposited transcripts (GRAVELEY *et al.* 2011a).

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 (SCHUPBACH AND WIESCHAUS 1991; JOHNSON *et al.* 2010). Therefore, we employed combined heat-shock Gal4 driver with UAS-RNAi lines to generate female flies primed for RNAi (STALLER *et al.* 2013). Once a stock with both components was generated, flies were allowed to mate, then females were heat-shocked three days in a row at 37°C for one hour, 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.

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 and dechorionated for 3 minutes according to standard protocols in 50% bleach, washed

with water, followed by a Triton-NaCl embryo wash, then rinsed briefly 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 MgCl2, 0.2mM EDTA, and 350mM sucrose supplemented with 1x Complete protease inhibitors and PhosStop (Roche)), at a ratio of 2ml buffer to 1g embryos. Extract was filtered through miracloth to remove debris. Nuclei were collected at 3000xg for 10 minutes, then washed 2X 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 α-Sxl (DSHB M114) or α-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), and 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). 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.* 2013a) with modifications. 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 β-Agarase I (NEB), and purified with a phenol:chloform 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).

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).

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 *$y^2cho^2v^1$; 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) (GRATZ *et al.* 2014). Integration in the genome at this position was confirmed by PCR/sequencing.

To delete the new exon, non-homologous end joining (NHEJ) mediated by the CRISPR-Cas9 genome editing system was utilized (BASSETT AND LIU 2014). *$y^2cho^2v^1$;sp/CyO;P {nos-Cas9,y^+,v^+} 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.

To delete the region including Sxl binding sites at the *sog* locus, homology directed repair (HDR) mediated CRISPR-Cas9 system was utilized (BASSETT AND LIU 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.

$y^2 cho^2 v^1$;sp/CyO;P {nos-Cas9,y^+,v^+} 2A (NIG-Fly #Cas-0004) virgin flies were collected and crossed with gRNA transgenic male flies. Embryos were collected and injected with 300ng/λ 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 for the presence of the 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.

SUPPLEMENTAL MATERIAL

FISH Probe Primers	
Sog 5' F	TCAGGTTCAGTCGCTCTTGA
Sog 5' R	GTGTCGGACTCCTCGAACA
sog In 3 F	ACAACAGCAGCAACAATGCGAGTAC
sog In 3 R	ACAATCTTCACTCGGGCAACGAAAATTC
Sog 3' F	GATGCACAGAATCCACATGCCGTC
Sog 3' R	GCCGTTCTCGTACACCTTGTTGAC
NetA 5' F	ATGATCCGTGGAATCTTGCTCCTGC
NetA 5' R	CTTTGCACTCATTGGCTTCCTTGGC
NetA 3' F	CCTGCGATTGCCATCCGATCG
NetA 3' R	ACACATAAATTATTGCTTACTGGCGCTCTG
sca 5' F	TTCAGTCCTTCGAAGAACGCCGC
sca 5' R	CTGAGTTGGTTCAGTTCGCTGCG
sca 3' F	CAAACCGGACTTGAAACCACTGCTG
sca 3' R	TTACGCCTGCCCGGCGGCC
Pka-C3 5' F	GAAAAGCGACAGGAGACAACGGG
Pka-C3 5' R	CGCACTCCCTTCGCGGTC
Pka-C3 3' F	GCACGGGCACCTTTGGAC
Pka-C3 3' R	CTTCATGTTGCCCAGTCGTTTGGT

qPCR Primers	
sog Ex1 F	GGAGCCCCAAGCGAGCAAAA
sog Ex1 R	CGCCAAGCAGACGATCAGCA
sog NewEx F	TGTGATGCCCCCAAACACCG
sog NewEx R	GCGCTGACCTTCATTTCGGG
sog In1 F	TGCTGCCTTGTTGCCAGTTG
sog In1 R	ATTCCGCCCGCCCATCAC
sog In2 F	GGTGGAAGCGAGAGGTGGAAAT
sog In2 R	CTGAATGGACGAATGCCAGGGG
sog Ex 5 F	CGGTTTGGCGTGGGTCTACT
sog Ex5 R	CGCTTGCCCTGCTCCTCAA
NetA F	GGACTTTGTGAACGCCGCCTA
NetA R	GTCGCAGGTGTGGCAGGAG
sca F	CGAGGATAGCGAGGACATCAGC
sca R	CGCAGCATCAGGGCGTTG
Pka-C3 F	GCCTCAAGCGAGTCATCCGA
Pka-C3 R	GTGGTGGTGGTGGCGGTG
sxl F	AACAACGACAGCAGCAGGC
sxl R	AGGATGATGAGGTGAGTTGCAGT
msl-2 F	TTCGCAGGATTCGGGGCAAG
msl-2 R	CGGCAGGTGGTGAGGGTATT
msl-2 3' F	GCTTCGGTTCCCTTCCCCAG
msl-2 3' R	CGGTGGCTCGATGACTTCCC
tra F	GCCTCAAGCGAGTCATCCGA
tra R	TGCGTCTGGTGGATTGGTGC
twi F	AGACGGAGGAGACGGACGAG
twi R	GGGCAGCGTGGGGATGAT
sna F	GCGACGAGGAGACCCAGGA
sna R	GCTCCAACTCCTGCCTGCTG

CRISPR Primers	
gRNA sog.del.sxl f	TATATAGGAAAGATATCCGGGTGAACTTCgatagttagaaggcacgggcgGTTTTAGAGCTAGAAATAGCAAG
gRNA sog.del.sxl r	ATTTTAACTTGCTATTTCTAGCTCTAAAACtggtccactacttcggataacGACGTTAAATTGAAAATAGGTC
conf gRNA	GACACAGCGCGTACGTCCTTCG
HDR.LA.sog.del.sxl f	GTACGTCCCGGGAGCAGCCACCAAAGTGTTCT
HDR.LA.sog.del.sxl r	CTAGCGGCTAGCCGATTTGGATTCGGAATAGG
HDR.RA.sog.del.sxl f	GTACGTGGCGCGCCGCGCGGCTTTTCCAG
HDR.RA.sog.del.sxl r	CTAGCGCTCGAGCGAGTCGATGGAATCGAAA
sog.del.newexon.conf f	GCGGCGACAGACATAAAAAC
sog.del.newexon.conf r	CCAATGGGGCATAAATCAGT
gRNA.sog.del.newexon f	TATATATAGGAAAGATATCCGGGTGAACTTCGGCATTTGTTGGTATCGATAGTTTTAGAGCTAGAAATAGCAAG
gRNA.sog.del.newexon r	GACCTATTTTCAATTTAACGTCGTAATTATGCGTAGCGTCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAAT

Supplemental Table 4.1. Primer list

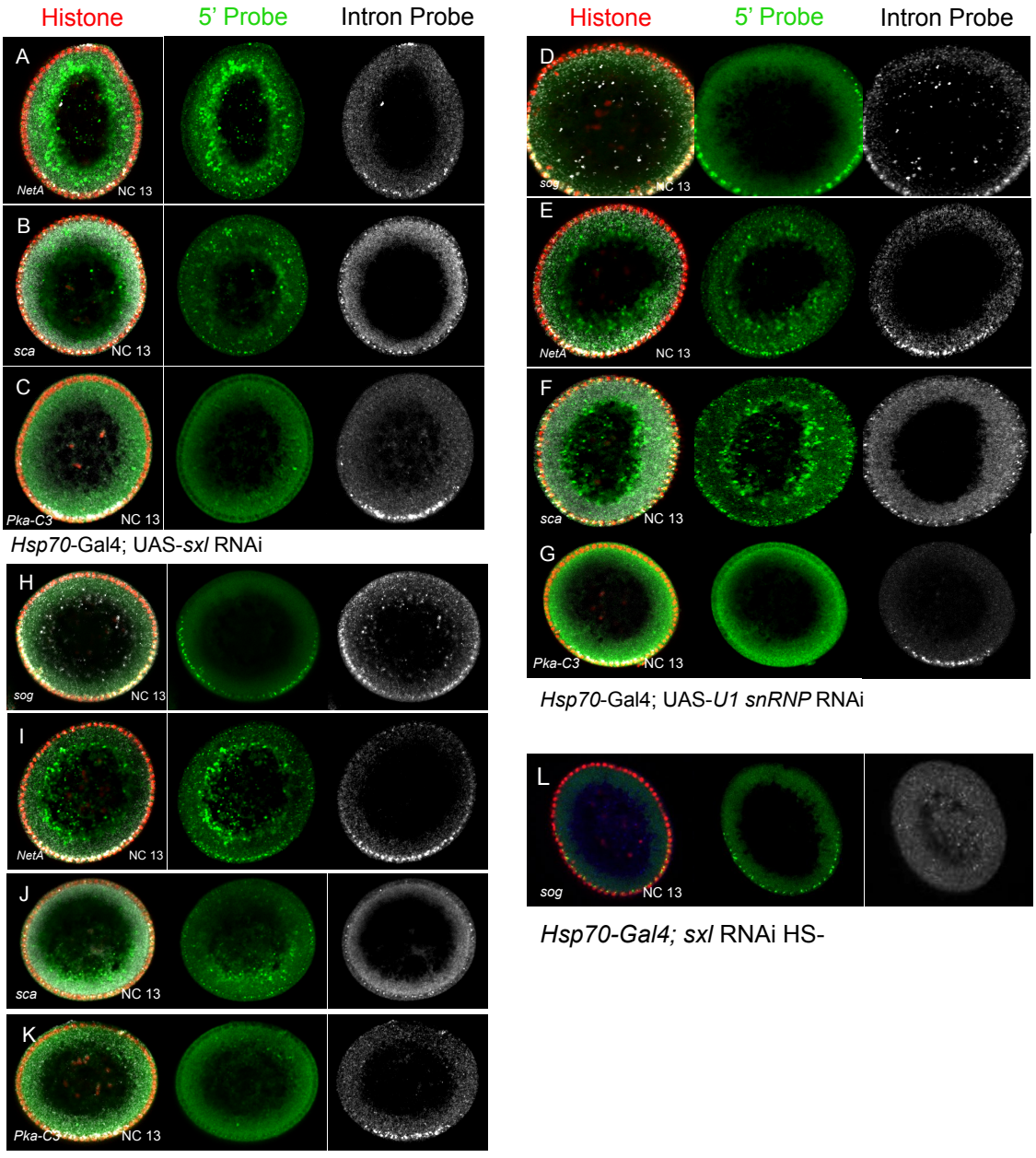


Figure S4.1. Transcriptional read-through in RNAi. Relates to Figure 2. (A-C) RNAi against *sxl* for genes *NetA, sca,* and *Pka-C3* leads to transcriptional read-through past truncation point. (D-G) RNAi against *U1 snRNP* for genes *sog, NetA, sca,* and *Pka-C3* leads to transcriptional read-through past truncation point. (H-K) RNAi against *PPS* for genes *sog, NetA, sca,* and *Pka-C3* leads to transcriptional read-through past truncation point. All images are chopped embryo sections as described in supplemental methods. (L) Heat shock negative RNAi control for *sog* using 5' exon and intronic probes. Images are chopped embryo sections as described in supplemental methods.

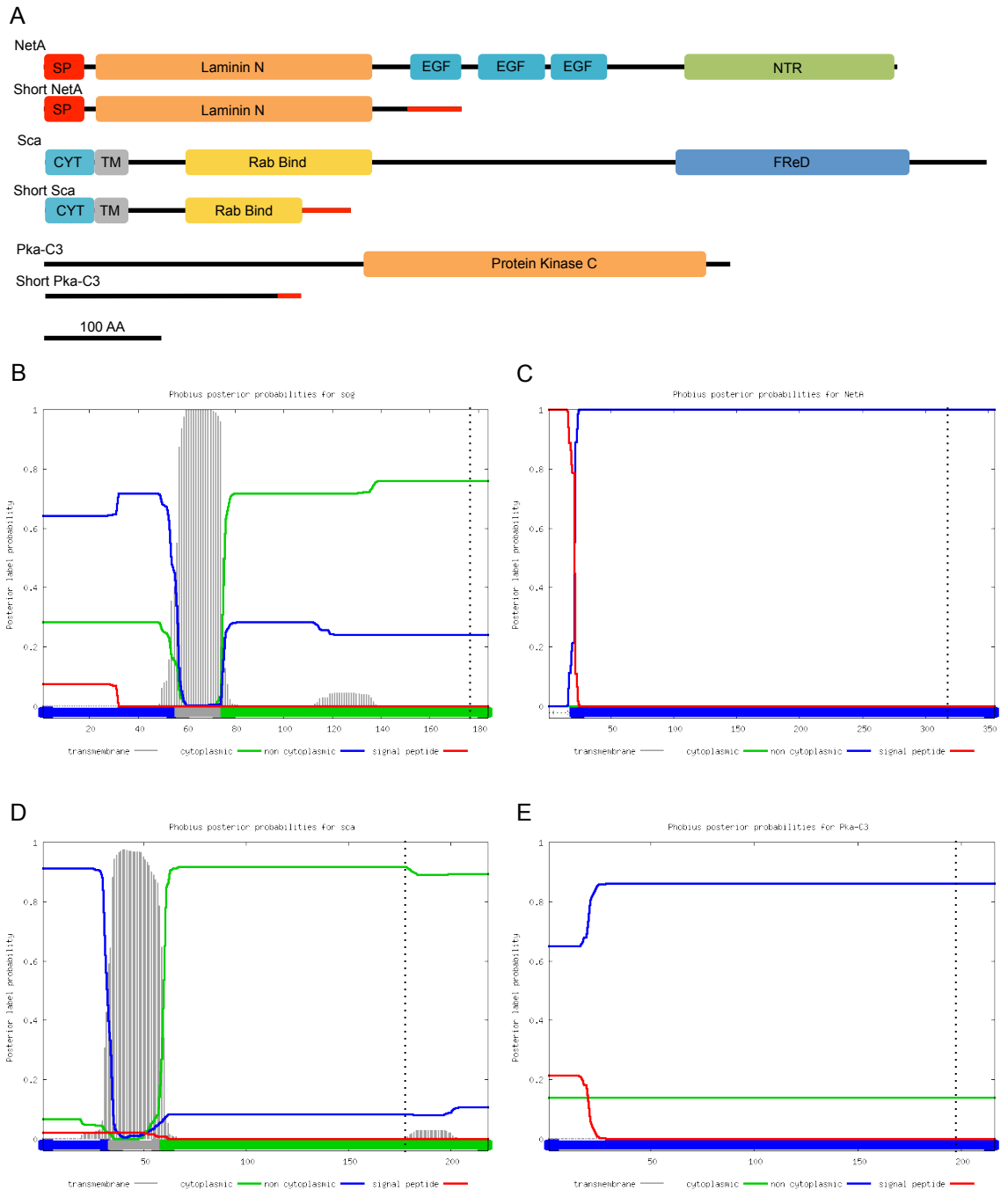


Figure S4.2. Protein products of short transcripts. Relates to Figure 3. (A) Full-length proteins and proteins resulting from truncated transcripts. SP= Signal Peptide, NTR= Netrin 1, CYT= Cytoplasmic, TM= Transmembrane, Rab Bind= Rab GTPase binding, FRed= Fibrinogen-related Domain. (B-E) Plots showing amino acid properties for short form proteins Sog, NetA, sca, and Pka-C3 respectively, with novel amino acids after the dashed line. Novel amino acids

retain the same or highly similar properties of preceding canonical sequence. Plots were generated using the Phobius tool for amino acid property prediction (KALL *et al.* 2004).

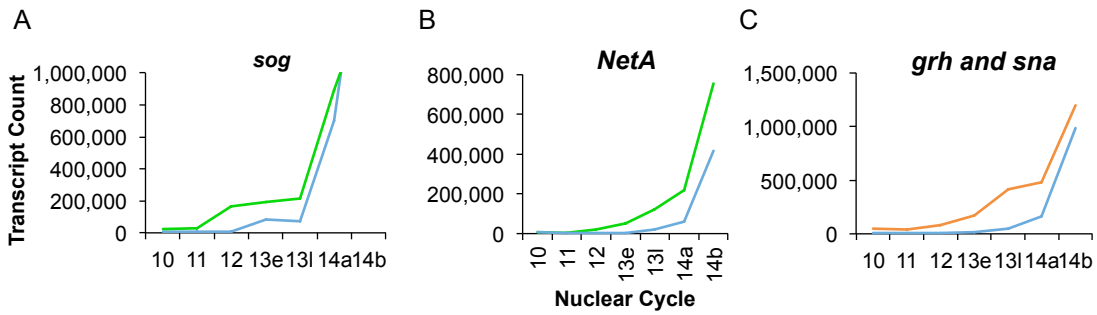


Figure S3. NanoString quantification of long transcripts. Relates to Figure 1. (A-C) Quantification of transcripts (y-axis individual transcript counts) within single *Drosophila* embryos of the indicated stages (x-axis) using NanoString (see supplemental methods, (SANDLER AND STATHOPOULOS 2016)). Levels of expression for long genes *sog* (A) and *NetA* (B) were probed at the 5' (blue) and 3' (green) end of genes. *grh* 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). 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)

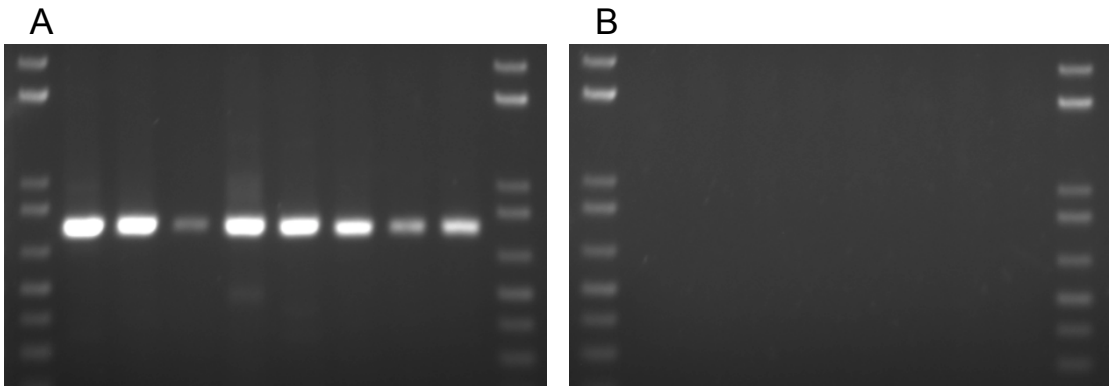


Figure S4. Truncated *sog* is present in NC14. **Relates to Figure 3.**
(A) RT-PCR using cDNA from eight individual NC14 embryos transcribed with an oligo-dT primer showing the short *sog* transcript is present. Reverse primer located in novel coding region of short *sog*. (B) Reverse Transcriptase negative control on RNA from the same embryos, demonstrating that signal is cDNA-dependent. DNA Ladder is 1kb Plus Ladder (ThermoFisher).

REFRENCES

Ali-Murthy, Z., S. E. Lott, M. B. Eisen and T. B. Kornberg, 2013 An Essential Role for Zygotic Expression in the Pre-Cellular Drosophila Embryo. Plos Genetics 9.

Ardehali, M. B., and J. T. Lis, 2009 Tracking rates of transcription and splicing in vivo. Nature Structural & Molecular Biology 16**:** 1123-1124.

Artieri, C. G., and H. B. Fraser, 2014 Transcript Length Mediates Developmental Timing of Gene Expression Across Drosophila. Molecular Biology and Evolution 31**:** 2879-2889.

Ashe, H. L., and M. Levine, 1999 Local inhibition and long-range enhancement of Dpp signal transduction by Sog. Nature 398**:** 427-431.

Bassett, A. R., and J. L. Liu, 2014 CRISPR/Cas9 and Genome Editing in Drosophila. Journal of Genetics and Genomics 41**:** 7-19.

Bell, L. R., E. M. Maine, P. Schedl and T. W. Cline, 1988 Sex-Lethal, a Drosophila Sex Determination Switch Gene, Exhibits Sex-Specific Rna Splicing and Sequence Similarity to Rna-Binding Proteins. Cell 55**:** 1037-1046.

Bernstein, M., and T. W. Cline, 1994 Differential-Effects of Sex-Lethal Mutations on Dosage Compensation Early in Drosophila Development. Genetics 136**:** 1051-1061.

Bopp, D., L. R. Bell, T. W. Cline and P. Schedl, 1991 Developmental Distribution of Female-Specific Sex-Lethal Proteins in Drosophila-Melanogaster. Genes & Development 5**:** 403-415.

Cline, T. W., 1979 Male-Specific Lethal Mutation in Drosophila-Melanogaster That Transforms Sex. Developmental Biology 72**:** 266-275.

Cline, T. W., M. Dorsett, S. Sun, M. M. Harrison, J. Dines *et al.*, 2010 Evolution of the Drosophila Feminizing Switch Gene Sex-lethal. Genetics 186**:** 1321-U1402.

Decotto, E., and E. L. Ferguson, 2001 A positive role for short gastrulation in modulating BMP signaling during dorsoventral patterning in the Drosophila embryo. Development 128**:** 3831-3841.

Evans, D. S., and T. W. Cline, 2013 Drosophila switch gene Sex-lethal can bypass its switch-gene target transformer to regulate aspects of female behavior. Proceedings of the National Academy of Sciences of the United States of America 110**:** E4474-E4481.

Foe, V. E., and B. M. Alberts, 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., B. Lim and M. Levine, 2017 Rapid Rates of Pol II Elongation in the Drosophila Embryo. Curr Biol.

Garcia, H. G., M. Tikhonov, A. Lin and T. Gregor, 2013 Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. Curr Biol 23**:** 2140-2145.

Geiss, G. K., R. E. Bumgarner, B. Birditt, T. Dahl, N. Dowidar *et al.*, 2008 Direct multiplexed measurement of gene expression with color-coded probe pairs. Nature Biotechnology 26**:** 317-325.

Gratz, S. J., F. P. Ukken, C. D. Rubinstein, G. Thiede, L. K. Donohue *et al.*, 2014 Highly Specific and Efficient CRISPR/Cas9-Catalyzed Homology-Directed Repair in Drosophila. Genetics 196**:** 961-+.

Graveley, B. R., A. N. Brooks, J. Carlson, M. O. Duff, J. M. Landolin *et al.*, 2011a The developmental transcriptome of Drosophila melanogaster. Nature 471**:** 473-479.

Graveley, B. R., A. N. Brooks, J. W. Carlson, M. O. Duff, J. M. Landolin *et al.*, 2011b The developmental transcriptome of Drosophila melanogaster. Nature 471**:** 473-479.

Hoskins, R. A., J. W. Carlson, K. H. Wan, S. Park, I. Mendez *et al.*, 2015 The Release 6 reference sequence of the Drosophila melanogaster genome. Genome Research 25**:** 445-458.

Hu, X., E. C. Lee and N. E. Baker, 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., A. A. Nagengast and H. K. Salz, 2010 PPS, a Large Multidomain Protein, Functions with Sex-Lethal to Regulate Alternative Splicing in Drosophila. Plos Genetics 6.

Kall, L., A. Krogh and E. L. L. Sonnhammer, 2004 A combined transmembrane topology and signal peptide prediction method. Journal of Molecular Biology 338**:** 1027-1036.

Kamakaka, R. T., C. M. Tyree and J. T. Kadonaga, 1991 Accurate and Efficient Rna Polymerase-Ii Transcription with a Soluble Nuclear Fraction Derived from Drosophila Embryos. Proceedings of the National Academy of Sciences of the United States of America 88**:** 1024-1028.

Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley *et al.*, 2013 TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14.

Konrad, K. D., T. J. Goralski, A. P. Mahowald and J. L. Marsh, 1998 The gastrulation defective gene of Drosophila melanogaster is a member of the serine protease superfamily. Proceedings of the National Academy of Sciences of the United States of America 95**:** 6819-6824.

Kosman, D., C. M. Mizutani, D. Lemons, W. G. Cox, W. McGinnis *et al.*, 2004 Multiplex detection of RNA expression in Drosophila embryos. Science 305**:** 846-846.

Lianoglou, S., V. Garg, J. L. Yang, C. S. Leslie and C. Mayr, 2013a Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes & Development 27**:** 2380-2396.

Lianoglou, S., V. Garg, J. L. Yang, C. S. Leslie and C. Mayr, 2013b Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. Genes Dev 27**:** 2380-2396.

Livak, K. J., and T. D. Schmittgen, 2001 Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. Methods 25**:** 402-408.

Lott, S. E., J. E. Villalta, G. P. Schroth, S. Luo, L. A. Tonkin *et al.*, 2011 Noncanonical compensation of zygotic X transcription in early Drosophila melanogaster development revealed through single-embryo RNA-seq. PLoS Biol 9**:** e1000590.

Marques, G., M. Musacchio, M. J. Shimell, K. WunnenbergStapleton, K. W. Y. Cho *et al.*, 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 D. P. Bartel, 2009 Widespread Shortening of 3 ' UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. Cell 138**:** 673-684.

Miloudi, K., F. Binet, A. Wilson, A. Cerani, M. Oubaha *et al.*, 2016 Truncated netrin-1 contributes to pathological vascular permeability in diabetic retinopathy. Journal of Clinical Investigation 126**:** 3006-3022.

Morris, D. P., and A. L. Greenleaf, 2000 The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. Journal of Biological Chemistry 275**:** 39935-39943.

Ofarrell, P. H., 1992 Big Genes and Little Genes and Deadlines for Transcription. Nature 359**:** 366-367.

Paz, I., I. Kosti, M. Ares, M. Cline and Y. Mandel-Gutfreund, 2014 RBPmap: a web server for mapping binding sites of RNA-binding proteins. Nucleic Acids Research 42**:** W361-W367.

Peluso, C. E., D. Umulis, Y. J. Kim, M. B. O'Connor and M. Serpe, 2011 Shaping BMP Morphogen Gradients through Enzyme-Substrate Interactions. Developmental Cell 21**:** 375-383.

Penalva, L. O., and L. Sanchez, 2003 RNA binding protein sex-lethal (Sxl) and control of Drosophila sex determination and dosage compensation. Microbiol Mol Biol Rev 67**:** 343-359, table of contents.

Pritchard, D. K., and G. Schubiger, 1996 Activation of transcription in Drosophila embryos is a gradual process mediated by the nucleocytoplasmic ratio. Genes Dev 10**:** 1131-1142.

Quinlan, A. R., 2014 BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics 47**:** 11 12 11-34.

Ray, D., H. Kazan, K. B. Cook, M. T. Weirauch, H. S. Najafabadi *et al.*, 2013 A compendium of RNA-binding motifs for decoding gene regulation. Nature 499**:** 172-177.

Reeves, G. T., N. Trisnadi, T. V. Truong, M. Nahmad, S. Katz *et al.*, 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., M. Pehl, H. Taubert and H. Jackle, 1992 Loss of Gene-Function through Rapid Mitotic-Cycles in the Drosophila Embryo. Nature 359**:** 156-159.

Rusch, J., and M. Levine, 1997 Regulation of a dpp target gene in the Drosophila embryo. Development 124**:** 303-311.

Sandler, J. E., and A. Stathopoulos, 2016 Quantitative Single-Embryo Profile of Drosophila Genome Activation and the Dorsal-Ventral Patterning Network. Genetics.

Schneiders, F. I., B. Maertens, K. Bose, Y. Li, W. J. Brunken *et al.*, 2007 Binding of netrin-4 to laminin short arms regulates basement membrane assembly. Journal of Biological Chemistry 282**:** 23750-23758.

Schupbach, T., and E. Wieschaus, 1991 Female Sterile Mutations on the 2nd Chromosome of Drosophila-Melanogaster .2. Mutations Blocking Oogenesis or Altering Egg Morphology. Genetics 129**:** 1119-1136.

Shermoen, A. W., and P. H. Ofarrell, 1991 Progression of the Cell-Cycle through Mitosis Leads to Abortion of Nascent Transcripts. Cell 67**:** 303-310.

Staller, M. V., D. Yan, S. Randklev, M. D. Bragdon, Z. B. Wunderlich *et al.*, 2013 Depleting Gene Activities in Early Drosophila Embryos with the "Maternal-Gal4-shRNA" System. Genetics 193**:** 51-61.

Tadros, W., and H. D. Lipshitz, 2009 The maternal-to-zygotic transition: a play in two acts. Development 136**:** 3033-3042.

Xu, M., N. Kirov and C. Rushlow, 2005 Peak levels of BMP in the Drosophila embryo control target genes by a feed-forward mechanism. Development 132**:** 1637-1647.

Yu, K., S. Srinivasan, O. Shimmi, B. Biehs, K. E. Rashka *et al.*, 2000 Processing of the Drosophila Sog protein creates a novel BMP inhibitory activity. Development 127**:** 2143-2154.

Chapter 5

DISCUSSION

The experiments and analysis described in this thesis all revolve around a unique developmental strategy found in the *Drosophila* genus: a rapidly dividing syncytium of nuclei that reaches gastrulation in around 3 hours (FOE AND ALBERTS 1983). While many insects, and even vertebrates (CARVALHO AND HEISENBERG 2010), include a complete or partial syncytial phase in development, *Drosophila* is unique in the speed and dynamic changes that occur during this time. The mosquitoes *Aedes aegypti* and *Anopheles gambiae*, which diverged from *Drosophila* ~250 million years ago (MYA) (BOLSHAKOV *et al.* 2002; SEVERSON *et al.* 2004), also undergo a syncytial blastoderm stage in early embryogenesis, but nuclear divisions are slower and cellularization occurs at ~8 hours after egg laying (GOLTSEV *et al.* 2004; CLEMONS *et al.* 2010), compared to ~2.5 in *Drosophila.* When comparing the time delay between cellularization and gastrulation, *Drosophila* again exhibit very rapid development, with the two events occurring less than 30 minutes apart, while the delay in mosquitoes is ~4 hours, gastrulation finally occurring ~12 hours after egg laying.

The flour beetle *Tribolium castaneum*, which diverged from *Drosophila* ~300 MYA, is also a syncytium during early embryogenesis (BROWN *et al.* 1994), and has developmental timing similar to mosquitoes, with cellularization occurring at ~8 hours after egg laying and gastrulation at ~12 hours (HANDEL *et al.* 2000). Lastly, the honeybee *Apis millifera*, ~300 MYA, diverged from *Drosophila (SAVARD et al. 2006)*, has a syncytial stage

with timing of cellularization and gastrulation matching *Tribolium* and mosquitoes *(PIRES et al. 2016)*. All three insects, mosquitoes*, Tribolium,* and *Apis*, are common genetic outgroups used in comparison to *Drosophila,* and provide good points of reference for examining evolution after divergence.

On the other hand, the rapid program is common to *Drosophila* species, with development uniformly rapid across the 12 commonly studied *Drosophila* species, especially in the early stages before gastrulation (KUNTZ AND EISEN 2014). The many *Drosophila* species vary in their time of divergence from each other, but some of the most distantly diverged species, *melanogaster* and *virilis*, are almost 40 million years apart (RUSSO *et al.* 1995). Based on the lack of rapid syncytial nuclear divisions and progression to cellularization and gastrulation in mosquitoes*, Tribolium,* and *Apis*, and its commonality among *Drosophila* species, it appears that this development program is a derived trait in *Drosophila*.

The question is then raised: How has *Drosophila* addressed the challenges associated with rapid embryonic development, namely, the time constraints on transcript length in the early embryo before cellularization? We took a multi-layered approach to address this question, first creating a fine time scale profile of the activation of the DV GRN in the embryo. This analysis allowed us to monitor the transcriptional dynamics of ~70 genes at a 10-15 minute resolution, giving insight into how the genome is rapidly activated and how the GRN is able to pattern the entire embryo and prime it for gastrulation in ~2.5 hours. Next, we focused on the presence of long transcripts during the short NC 13, when transcript length is limited by interphase time before mitosis. We identified truncated transcripts of several genes in these time points, characterized a mechanism responsible for

their truncation, and showed that the short transcripts produce functional proteins to regulate the spatiotemporal activation of signaling pathways in the early embryo. With the data and insights gained from these approaches, a picture has emerged that describes the evolution of a unique developmental program and strategies needed to address challenges along the way.

The developmental time course created using NanoString provides the highest temporal resolution data to date describing *Drosophila* development, and the first highly quantitative data of ~70 genes in parallel from a single sample. The genes we analyzed were all chosen from the previously characterized DV GRN (LEVINE AND DAVIDSON 2005; STATHOPOULOS AND LEVINE 2005) (Figure 5.1).

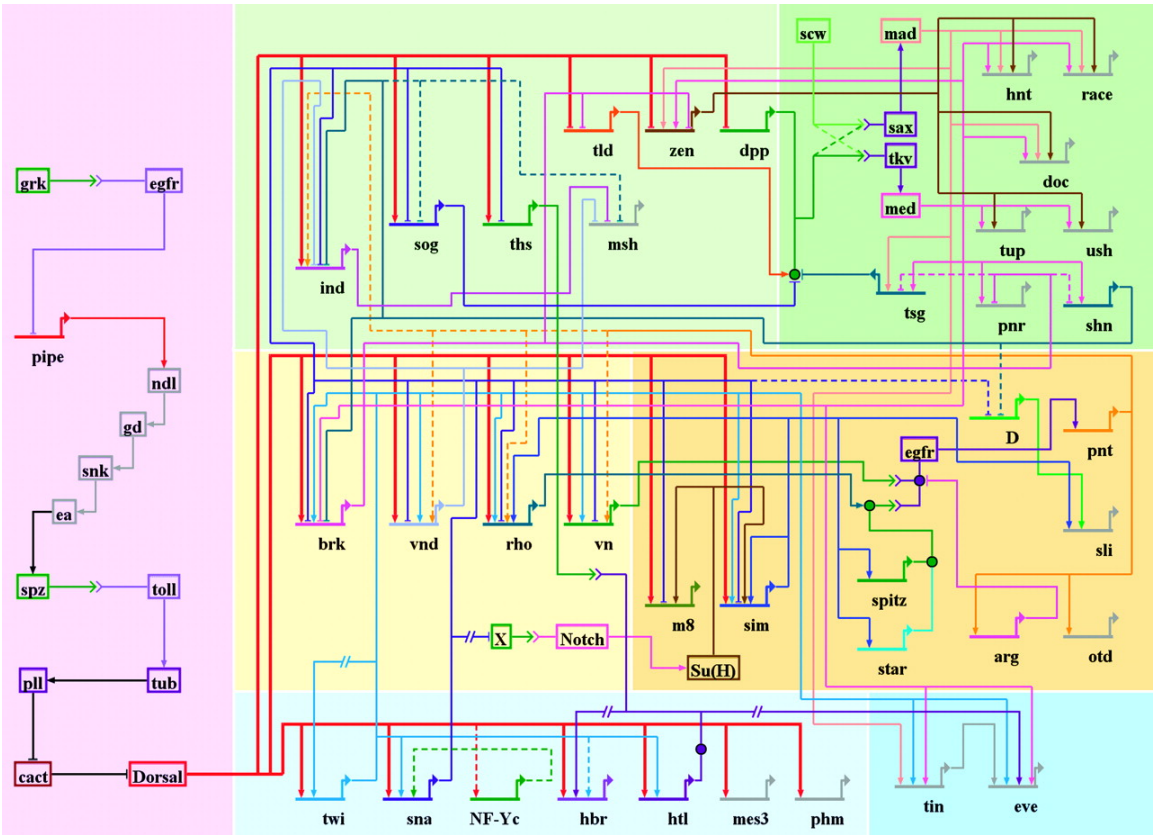


Figure 5.1 *Drosophila* DV GRN. The GRN describing genes action on the DV axis of the embryo. Genes in purple are maternally deposited. Blue domain indicates mesoderm, orange indicated

neurogenic ectoderm, and green indicates dorsal ectoderm. Lighter shades are genes acting before cellularization, and darker shades after cellularization. Horizontal lines above gene symbols represent enhancers, with inputs from other genes as arrows (activating) or blocks (repressing). Signaling nodes are represented by black circles. Figure from (LEVINE AND DAVIDSON 2005).

The GRN represents a compilation of years of experimental observations, from mutating, overexpressing and, ectopically expressing genes, and observing the changes in expression of known and hypothesized interacting genes. GRNs can be displayed in a variety of ways, but a commonly used tool to input network information and create a graphical representation of genetic interactions is through the use of the software package BioTapestry (LONGABAUGH *et al.* 2005). The information contained in the graphical GRN shows many aspects of cis-regulatory circuitry, with enhancers of constituent genes depicted and transcriptional inputs from other network genes (transcription factors) directly contacting enhancers (DAVIDSON *et al.* 2002) (Figure 5.1). Secreted signaling molecules, cell membrane receptors, and signaling transducers are represented as nodes of interaction that then connect to enhancers (Figure 5.1). The network view of a developing embryo provides a new way of thinking about development itself, shifting from studying single genes that are used to make individual network connections, to a view that incorporates all the genes at once to describe the progression of development.

The combined transcriptional inputs present in any tissue or cell type, represented by presence in a GRN model, can be thought of as the regulatory state, which controls the overall output of the GRN. The regulatory state is responsible for an instantaneous snapshot of GRN activity, but the regulatory state also is constantly in flux as genes are turned on or off (DAVIDSON 2006). Because the regulatory state changes every time a new gene is expressed or a gene stops being expressed, the GRN model can provide a prediction

of the new resulting regulatory state, but the GRN model itself must also be revised. Even highly detailed GRNs compress a 4-dimensional embryo (three spatial dimensions plus time) into a two-dimensional depiction. This compression results in a loss of data from the time dimension, which in some cases was never included in the GRN in the first place. Single gene perturbation experiments also provide a snapshot of gene interactions at a specific developmental stage, rather than in a dynamic and developing embryo where regulatory state is in constant flux. To answer questions and provide insight into the GRN details that are are lost in temporal compression, we used NanoString to create a dynamic GRN, with details on gene expression at high time resolution through development.

One question that often arises with the use of NanoString is whether or not the data can give spatial information on gene expression in the embryo. As the embryos are homogenized during RNA extraction, spatial information is lost. That is why targeted in situ hybridizations have become a powerful tool to investigate changes in expression. NanoString Technologies is developing a method to address this question and provide spatially resolved information about gene expression. This method uses UV-cleavable probes and microcapillaries to sample transcripts from single cells in fixed tissues (DUNAWAY *et al.* 2016). Different cell types to be sampled are stained with cell type specific markers, and UV light is used to cleave the NanoString probes from specific cells. Cleaved probes representing transcripts present in specific cells are collected in a microcapillary and quantified using the NanoString instrument. Currently, the method has only been tested on FFPE tissue slides, but NanoString is developing a method for sampling 3D tissues which will allow transcript quantification using fixed *Drosophila* embryos.

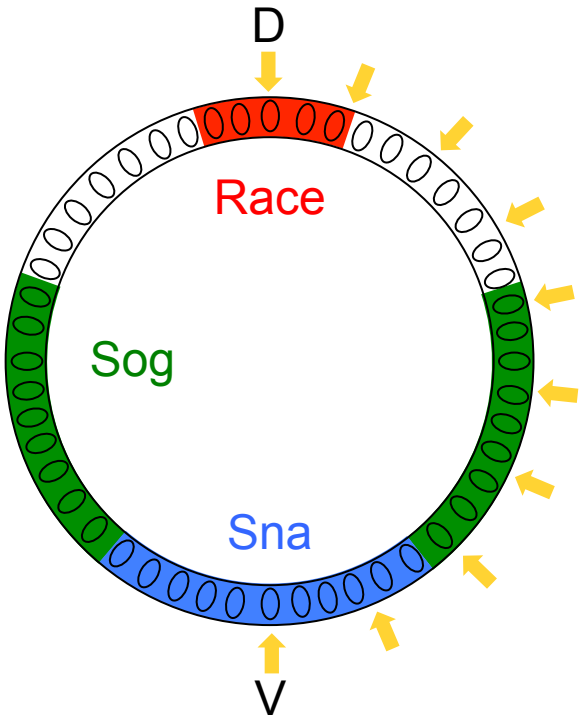


Figure 5.2. Sampling the embryo along the DV axis for NanoString. An illustration of an embryo labeled with antibodies for Sna (blue), Sog (green), and Race (red), providing positional information for the DV axis and used to orient the embryo for sampling. Yellow arrows indicate 10 points along the DV axis where focused UV light would free NanoString probes bound to RNA for sampling and transcriptional profiling.

For example, an antibody against Snail could be used to mark the mesoderm, an antibody against Sog could be used to mark the neurogenic ectoderm, and an antibody against Race could mark the dorsal ectoderm, providing coordinates and embryo orientation so a tissue-specific transcriptional profile for multiple sampling points along the DV axis can be created (Figure 5.2). With the dorsal and ventral midpoints marked, the same absolute position along the embryo could be sampled in wild type and mutant embryos, allowing the Dorsal gradient along the DV axis to be measured and providing a look at how regulatory states change in mutants. Some genes, such as *Neu2,* have no place in the DV GRN despite evidence that they are in fact expressed and regulated long the DV

axis (STATHOPOULOS *et al.* 2002). Using NanoString, *Neu2*$^{-}$ embryos can be compared to WT, precisely quantifying the expansion of the Dl nuclear gradient along fixed points using UV-linked probes (Figure 5.2).

An additional experiment that would shed light on the role of *Neu2* would be to overexpress *Neu2* throughout the embryo, using a UAS-*Neu2* construct with a ubiquitous Gal4 driver. If the model of Neu2 binding to Weckle to balance or titrate the concentration of Dl is accurate, then an overexpression should have a mostly opposite effect on gene expression compared to the *Neu2*$^{-}$ embryos observed with NanoString. Genes that rely on high concentrations of Dl, such as *sna* and *twi,* would likely be expressed in a narrower domain, while genes that rely on a lower concentration, such as *brinker* (*brk)* or *intermediate neuroblasts defective* (*ind),* would be shifted ventrally. There may be a ventral expansion of *sog* or *brk* as the *sna* domain shrinks, but lower concentration of Dl in lateral regions may shift the dorsal boundaries of those genes towards the ventral pole.

The effect of ubiquitous *Neu2* expression on the TGF-β signaling pathway is slightly more complicated to predict, as there is no direct Dl binding to the enhancers of TGF-β output genes. If *brk* and *sog* shift ventrally and overall expression decreases, the expression of *race* could expand, due to an expansion of *zerknullt* (*zen), dpp,* and *tld (JAZWINSKA et al. 1999; RUSHLOW et al. 2001),* all three of which are directly repressed by *brk* and have an activating input into the domain of TGF-β target genes. On the other hand, with less *sog* present, its ability to transport Dpp and Scw dorsally to initiate TGF-β signaling may be reduced, leading to a weaker expression of TGF-β target genes. As has been noted before, while NanoString provides very accurate quantification of transcription,

follow-up in situ hybridizations can help clarify changes in expression and determine if they are due to changes in domain or expression level.

In the case of *twi*$^{-}$ embryos, even though over 25 years of molecular research has been done on interactions of the transcription factor, it is difficult to find any studies measuring change in *race* expression in *twi* mutant backgrounds. This lack of experiments could be due to two major factors: the fact that *twi* and *race* are expressed in opposite sides of the DV axis and Twist does not diffuse extracellularly (LEPTIN 1991), and that there are no *twi* binding sites in the *race* enhancer (RUSCH AND LEVINE 1997). For these reasons, single gene *twi* perturbation experiments did not include *race* as a target gene with expected changes in expression. Only when the network view of development was used and the entire DV GRN was assayed in parallel did all of the changes needed to alter *race* expression become evident. This is the true strength of the NanoString data: the ability to monitor dynamic changes in GRN regulatory state and precisely quantify the expression of an entire netwrok to see how all the genes react to perturbation. Relatively simple follow-up experiments can fill in the gap of the missing spatial information and complete the picture.

The more difficult question to address is that of the seemingly contradictory presence of long transcripts observed during NC 13 of the syncytial blastoderm stage. The problem of limited time available for transcription has been understood for over 25 years (SHERMOEN AND OFARRELL 1991; OFARRELL 1992; ROTHE *et al.* 1992), with the prevailing assumption that if transcription was started on long genes during NC 13, mitosis would occur before the mRNA could be completed, and the transcript would be degraded during nuclear division. Even recently, researchers dismissed the presence of long transcripts from

Drosophila embryos staged prior to NC 14 as sequencing error due to this limitation, and dismissed the reads (ALI-MURTHY *et al.* 2013).

We took the approach that recent evidence for transcription of long genes during NC 13 might represent a meaningful biological process, not nascent transcription destined to be degraded or an experimental error (REEVES *et al.* 2012; SANDLER AND STATHOPOULOS 2016), and investigated the phenomenon further. In brief, we uncovered a developmental program to truncate the transcripts of long genes during the short NC 13 so they produce functional proteins that regulate the timing and spatial activation of signaling pathways. The RBP Sxl binds directly to these transcripts and is responsible for their truncation, working with the proteins PPS and U1 snRNP. The findings of the investigation are discussed in detail in chapter 4, but there are unanswered questions and several lines of research that warrant further investigation.

The area of research with several unanswered questions is that of the protein interactions responsible for truncation of the short transcripts. Results from RNAi experiments and immunoprecipitations (IPs) show that Sxl, PPS, and U1 snRNP are all critical for the truncation of the transcripts, and that Sxl binds directly to the RNA. The nature of the interactions of PPS and U1 snRNP, however, both with the RNA transcript and RNA Polymerase II (RNA Pol II) and the rest of the transcriptional complex is unclear. PPS has been described in one previous study (JOHNSON *et al.* 2010), and has been shown to physically interact with Sxl and U1 snRNP to participate in Sxl-mediated alternative splicing. This interaction is partially dependent on the RNA being spliced itself, as PPS-Sxl binding decreases with the addition of RNase to the reaction. An unexplored avenue of the role of PPS is predicated on the presence of several plant homeodomain (PHD finger)

motifs, which have been shown to bind to methylated H3K4 (H3K4me3), an epigenetic marker of active transcription (LIANG *et al.* 2004; SHI *et al.* 2007). It should be noted that H3K4me3 is enriched primarily at the promoter, so possible PPS-H3K4me3 interactions downstream of the promoter at truncation points may be a rare occurrence or may not be as relevant as other interactions with the Sxl splicing complex. In addition, Prp40, a component of the U1 snRNP complex, has been shown to bind to RNA Pol II itself (MORRIS AND GREENLEAF 2000).

The binding interactions described above paint a picture of a possible mechanism for transcript truncation. First, Sxl binds to its RNA binding site, U_8. Next, PPS and U1 snRNP are recruited to bind to Sxl. Lastly, the PHD fingers of PPS bind to H3K4me3 and Prp40 binds to RNA Pol II. This complex bound to the nascent RNA strand, RNA Pol II, and H3K4me3 may be strong enough to pause or stall RNA Pol II so the polyadenylation complex can terminate transcription and polyadenylate the transcript.

Much investigation is needed to confirm or refute this model, and it centers around exploring the protein-protein interactions between the Sxl-PPS-U1 snRNP complex and transcriptional machinery and chromatin proteins. These experiments would involve an IP of Sxl and a western blot to confirm the binding of PPS and U1 snRNP. A previously described antibody against PPS no longer exists, so assaying interactions may be difficult. There are other options, since PPS is an ortholog of the human protein DIDO1, and shares regions of very close homology, and antibodies against DIDO1 can be tested for recognition of PPS.

Next, the hypothetical PPS-H3K4me3 interaction and U1 snRNP-RNA Pol II interaction need to be tested. These experiments will be similar to the Sxl IPs, with PPS and

U1 snRNP pulled down and western blots performed using H3K4me3 and RNA Pol II antibodies. If one or both of these interactions is confirmed, a further Sxl IP or western blot would provide evidence that Sxl is also acting, especially since U1 snRNP is a general splicing factor, and the U1 snRNP-RNA Pol II interaction may be common without Sxl.

A series of Chromatin IP sequencing (ChIP-seq) experiments would also shed light on Sxl binding to long genes and a possible stalled RNA Pol II. Sxl ChIP-seq would be relatively straightforward, with well-characterized target genes such as *sxl, msl-2,* and *tra* serving as positive controls for Sxl binding to RNA. Furthermore, RNA Pol II ChIP-seq could identify pause points in long genes that correlate to truncation points (FUSBY *et al.* 2016). This method has identified ChIP-seq signatures of paused RNA Pol II specifically associated with alternative polyadenylation, which would be very helpful to future experiments.

While rapid nuclear cycles and progression to gastrulation are derived and unique to *Drosophila*, trandscript truncation is common among many organisms and model systems. The alternate truncation of transcripts and shortening of 3' UTRs has been identified as an important change that leads to the activation of oncogenes and the progression of cancer (MAYR AND BARTEL 2009), and is observed in activated immune and neuronal cells, and stem cells (BERG *et al.* 2012). The shortening of 3' UTRs in particular can lead to rapid de-regulation of transcripts, since most miRNAs bind to 3' UTRs of transcripts (GRIMSON *et al.* 2007). With new short 3' UTRs, oncogenes are free from repression and can lead to progression of cancer. In addition, poly-U sequences are enriched in close proximity to alternate polyadenylation sites (PROUDFOOT 1991;

KUERSTEN AND GOODWIN 2003; LEGENDRE AND GAUTHERET 2003; PROUDFOOT 2011), just as was found in the binding site search of truncated *Drosophila* genes.

The human orthologs of Sxl are found in the ELAV-like gene family, comprised of four genes. ELAVL1 is enriched in blood and immune cells and germ cells, while ELAVL2, ELAVL3, and ELAVL4 are enriched in neuronal cells (SU *et al.* 2004), consistent with previous observations about cell types with frequent alternative polyadenylation (BERG *et al.* 2012). ELAVL1 is a target for study in the well-characterized system of activated B cells. When B cells are activated, an alternative splicing event takes place in immunoglobulin genes where membrane-bound regions are swapped for regions promoting the secretion of immunoglobulin (EARLY *et al.* 1980; ROGERS *et al.* 1980; PETERSON *et al.* 1991). Experiments in cell cultures of B cells could provide insight into the action of ELAVL1, alternative splicing, and possibly the function of Sxl in *Drosophila*, with the alternate splicing of genes or secretion state of immunoglobulin as an assay. *Drosophila* Sxl could also be expressed in B cell lines lacking ELAVL1 (through RNAi or in specific cell lines that do not express the gene) to assay the ability of Sxl to alternatively splice and rescue immunoglobulin secretion, and the expression of ELAVL1 in *Drosophila* to attempt a rescue of *sxl* mutations could also be done.

Another research direction stemming from the truncation of transcripts involves the evolutionary aspects of the program itself. Since the rapid nuclear cycles are a derived trait in *Drosophila*, there were likely some challenges that the embryos faced when the rapid nuclear cycles evolved. First, as has been previously shown, mitosis between nuclear cycles truncates active transcription and nascent transcripts are degraded. The loss of transcripts from long genes probably resulted in the deregulation of signaling pathways, as was shown

for the TGF-β signaling pathway with the *sog* mutants and resulting changes in *race* expression. How did the embryos adapt to this change and possible loss of signaling regulation? The truncation of long transcripts is a solution to the problem, but how was this solution reached? Is the *sxl*-mediated truncation an ancestral or derived function?

A possible answer lies in the duplication of ancestral Sxl and evolution of new functions for the gene. The Mediterranean fruit fly, *Ceratitis capitata,* diverged from *Drosophila* ~100 Mya (AYALA *et al.* 1996), much more recently than mosquitoes*, Tribolium,* and *Apis*. *Ceratitis* shares the same developmental timing as mosquitoes*, Tribolium,* and *Apis*, with syncytial nuclear cycles lasting for about 10 hours before cellularization (GABRIELI *et al.* 2010). Also, *Ceratitis* relies on the gene *transformer (tra)* as the master regulator of sex determination (GABRIELI *et al.* 2010), unlike in *Drosophila*, where *tra* can not splice itself and is downstream of *sxl*. Based on the timing of the *Ceratitis*-*Drosophila* divergence and the appearance of the duplication of *sxl* to create *sister of sxl (ssx)* in *Drosophila*, the gender-determining role of *sxl* must have evolved in a time span of around 10 million years, a relatively short amount of time in evolutionary terms (CLINE *et al.* 2010).

Two distantly related *Drosophila* species, *melanogaster* and *virilis,* can be compared, and it is evident that the role of *sxl* evolved before their divergence ~40 Mya, as the gene maintains the same gender determining functions in both species (CLINE *et al.* 2010). Furthermore, when *ssx* is deleted, there is no observable phenotype or effect on fitness, even in the presence of *sxl* mutations, suggesting that *sxl* did not merely evolve new functions and leave *ssx* as a functional ancestral form, but *sxl* must still maintain its ancestral function in some capacity (CLINE *et al.* 2010). Another piece of evidence

supporting *sxl* maintaining an ancestral non-gender determining function in *Drosophila* is that flies make an almost full-length and non-gender specific *sxl* transcript and protein at 20-40 fold less than the female-specific form, contrary to the popular perception that the male form of the transcript is short and non-functional (BOPP *et al.* 1991; CLINE *et al.* 2010). This non-gender dependent *sxl* isoform is transcribed starting with a small exon termed "exon Z" (CLINE *et al.* 2010), which is upstream of the gender-spliced exon 3 that contains the male stop codon (Figure 5.3).

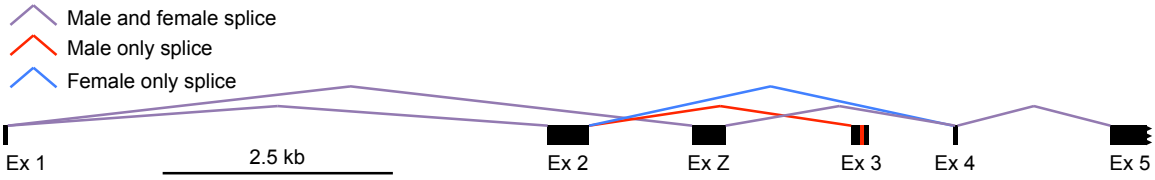


Figure 5.3. Sxl splicing. A schematic of the *sxl* locus showing gender-based splicing. Female only splicing is shown blue, male only splicing in red, and non-gender specific splicing in purple. Exon 3. The stop codon in exon 3 is represented by a red mark. Exon 1 splices to exon 2 in gender-based splicing, but exon 1 splices directly to exon Z in gender-independent splicing. Adapted from (CLINE *et al.* 2010).

Exon Z splices directly to exon 4, and produces an almost full-length *sxl* transcript and Sxl protein. In addition, when full-length *sxl* is ectopically expressed in neurons of adult *Drosophila*, it exhibits a subtle non-gender specific alternative splicing function, suggesting it can still function in an ancestral role (EVANS AND CLINE 2013).

When these observations about the function and evolution of *sxl* are combined with the observations of *sxl*-mediated truncation of short transcripts made in chapter 4, many questions can be answered. It appears likely that the action of Sxl truncating transcripts during NC 13 is a function of the non-gender specific full-length *sxl* observed in low levels, and is in fact separate from the gender-determining role. One major roadblock to studying

sxl is that *sxl* mutants are either sterile, the mutation itself is lethal, or mutants produce only male offspring (BERNSTEIN AND CLINE 1994). All three of these cases make studying *sxl* in the early embryo a challenge, as sterile females, dead flies, and males cannot produce embryos. It is likely that these challenges have masked the novel function of *sxl* we describe in chapter 4, and the strong and obvious gender-determining mutant phenotypes have overshadowed the subtle function we describe.

Taking into account the aforementioned challenges, we used a different approach to studying the loss of *sxl* in the early embryo. We used a precisely timed heat shock Gal4 driver to deliver RNAi against sxl in the female ovaries, after ovaries developed, while the developing oocytes were being loaded with maternal RNAs (STALLER *et al.* 2013). This allowed us to remove *sxl* transcripts from early embryos while still maintaining fertile females to produce the embryos. Only using this approach were we able to collect *sxl*$^{-}$ embryos to study the effects resulting from the loss of the gene.

With the perspective of the of evolutionary insights into the origin and function of *sxl,* we can examine the four long genes shown to be truncated and search for conserved *sxl* binding sites, both among *Drosophila* species and in the outgroup insects mosquitoes, *Tribolium,* and *Apis.* In the genes *sog* and *NetA,* and *grh*, the Sxl binding sites are highly conserved among *Drosophila* species, but not the outgroups (Figure 5.4), with the exception of a singe Sxl binding site in the intron of *Apis grh* (Figure 5.4 C). In fact, intron sequences of the outgroups are divergent, while the coding sequence is still highly conserved, strengthening the observation that the intronic Sxl binding sites evolved only when the short nuclear cycles presented a challenge in *Drosophila,* while Sxl binding sites in outgroup introns evolved after the divergence and independently of *Drosophila* short

nuclear cycles. Interestingly, *Ceratitis* has a single Sxl binding site in the intron corresponding to the *sog* intron with Sxl binding sites, although slightly closer to the end of the coding exon than in *Drosophila sog*.

Figure 5.4. Conserved Sxl binding sites among *Drosophila* species. Conservation tracks for the genes *sog* (A), *NetA* (B), and *grh* (C) with genome sequences of 12 commonly studied *Drosophila* species and *Anopheles, Apis,* and *Tribolium*. Sxl binding sites, boxed in red, are conserved among the vast majority of *Drosophila* species, but not in the outgroups. Sxl binding sites in the genes are highlighted in black text.

The recent identification of exon Z in *sxl* (CLINE *et al.* 2010) (Figure 5.3) will lead to interesting and insightful studies of its action in the early embryo. A fly stock exists that

deletes *sxl* exon Z, along with exons 1 and 2, the male promoter, and the promoter responsible for the first burst of *sxl* expression, but leaves the female promoter whole (EVANS AND CLINE 2013). This stock may provide interesting results when *sog, NetA, sca,* and *Pka-C3* are evaluated for transcriptional read-through, just as in *sxl* RNAi or CRISPR deletion of the Sxl binding sites. An alternative experiment would be to use CRISPR/Cas9 to delete only exon Z, leaving exons 1, 2, and the earlier promoters intact so *sxl* can be transcribed early without interference. This more precise method would allow for greater confidence in results.

The experiments on *sxl* in relation to its truncating role in the early embryo, combined with the evolutionary analysis and conservation of binding sites across *Drosophila,* paints a picture of how a species can adapt to and overcome a radical change in its developmental program. In the case of *Drosophila,* the rapid nuclear cycles presented a need to truncate essential transcripts in order to maintain the proper timing of signaling pathways. The near concurrent duplication of *sxl* and its cooption as a master gender regulator provided a means to solve the time constraint problem. The conservation of Sxl binding sites across many *Drosophila* species points to the rapid evolution of these binding sites as a solution, as they exist across ~40 million years of divergence and evolution, along with the novel function of Sxl. The recent discovery of exon Z now provides a specific hypothesis to test about the previously undescribed developmental program described in this thesis.

REFERENCES

Ali-Murthy, Z., S. E. Lott, M. B. Eisen and T. B. Kornberg, 2013 An Essential Role for Zygotic Expression in the Pre-Cellular Drosophila Embryo. Plos Genetics 9.

Ayala, F. J., E. Barrio and J. Kwiatowski, 1996 Molecular clock or erratic evolution? A tale of two genes. Proc Natl Acad Sci U S A 93**:** 11729-11734.

Berg, M. G., L. N. Singh, I. Younis, Q. Liu, A. M. Pinto *et al.*, 2012 U1 snRNP Determines mRNA Length and Regulates Isoform Expression. Cell 150**:** 53-64.

Bernstein, M., and T. W. Cline, 1994 Differential-Effects of Sex-Lethal Mutations on Dosage Compensation Early in Drosophila Development. Genetics 136**:** 1051-1061.

Bolshakov, V. N., P. Topalis, C. Blass, E. Kokoza, A. della Torre *et al.*, 2002 A comparative genomic analysis of two distant diptera, the fruit fly, Drosophila melanogaster, and the malaria mosquito, Anopheles gambiae. Genome Research 12**:** 57-66.

Bopp, D., L. R. Bell, T. W. Cline and P. Schedl, 1991 Developmental Distribution of Female-Specific Sex-Lethal Proteins in Drosophila-Melanogaster. Genes & Development 5**:** 403-415.

Brown, S. J., J. K. Parrish, R. E. Denell and R. W. Beeman, 1994 Genetic-Control of Early Embryogenesis in the Red Flour Beetle, Tribolium-Castaneum. American Zoologist 34**:** 343-352.

Carvalho, L., and C. P. Heisenberg, 2010 The yolk syncytial layer in early zebrafish development. Trends Cell Biol 20**:** 586-592.

Clemons, A., D. Severson and M. D. Scheel, 2010 Functional analysis of developmental genes in Aedes aegypti, an emerging model for vector mosquito development. Developmental Biology 344**:** 528-528.

Cline, T. W., M. Dorsett, S. Sun, M. M. Harrison, J. Dines *et al.*, 2010 Evolution of the Drosophila Feminizing Switch Gene Sex-lethal. Genetics 186**:** 1321-U1402.

Davidson, E. H., 2006 *The regulatory genome : gene regulatory networks in development and evolution*. Academic, Burlington, MA ; San Diego.

Davidson, E. H., J. P. Rast, P. Oliveri, A. Ransick, C. Calestani *et al.*, 2002 A genomic regulatory network for development. Science 295**:** 1669-1678.

Dunaway, D., C. Merritt, J. Jung, P. Webster, C. Ngouenet *et al.*, 2016 Spatially Resolved, Multiplexed Digital Characterization of Protein and RNA Expression in FFPE Tissue Sections. Journal of Molecular Diagnostics 18**:** 1008-1008.

Early, P., J. Rogers, M. Davis, K. Calame, M. Bond *et al.*, 1980 Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. Cell 20**:** 313-319.

Evans, D. S., and T. W. Cline, 2013 Drosophila switch gene Sex-lethal can bypass its switch-gene target transformer to regulate aspects of female behavior. Proceedings of the National Academy of Sciences of the United States of America 110**:** E4474-E4481.

Foe, V. E., and B. M. Alberts, 1983 Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis. J Cell Sci 61**:** 31-70.

Fusby, B., S. Kim, B. Erickson, H. Kim, M. L. Peterson *et al.*, 2016 Coordination of RNA Polymerase II Pausing and 3 ' End Processing Factor Recruitment with Alternative Polyadenylation. Molecular and Cellular Biology 36**:** 295-303.

Gabrieli, P., A. Falaguerra, P. Siciliano, L. M. Gomulski, F. Scolari *et al.*, 2010 Sex and the single embryo: early deveopment in the Mediterranean fruit fly, Ceratitis capitata. Bmc Developmental Biology 10.

Goltsev, Y., W. Hsiong, G. Lanzaro and M. Levine, 2004 Different combinations of gap repressors for common stripes in Anopheles and Drosophila embryos. Developmental Biology 275**:** 435-446.

Grimson, A., K. K. H. Farh, W. K. Johnston, P. Garrett-Engele, L. P. Lim *et al.*, 2007 MicroRNA targeting specificity in mammals: Determinants beyond seed pairing. Molecular Cell 27**:** 91-105.

Handel, K., C. G. Grunfelder, S. Roth and K. Sander, 2000 Tribolium embryogenesis: a SEM study of cell shapes and movements from blastoderm to serosal closure. Development Genes and Evolution 210**:** 167-179.

Jazwinska, A., C. Rushlow and S. Roth, 1999 The role of brinker in mediating the graded response to Dpp in early Drosophila embryos. Development 126**:** 3323-3334.

Johnson, M. L., A. A. Nagengast and H. K. Salz, 2010 PPS, a Large Multidomain Protein, Functions with Sex-Lethal to Regulate Alternative Splicing in Drosophila. Plos Genetics 6.

Kuersten, S., and E. B. Goodwin, 2003 The power of the 3' UTR: translational control and development. Nat Rev Genet 4**:** 626-637.

Kuntz, S. G., and M. B. Eisen, 2014 Drosophila embryogenesis scales uniformly across temperature in developmentally diverse species. PLoS Genet 10**:** e1004293.

Legendre, M., and D. Gautheret, 2003 Sequence determinants in human polyadenylation site selection. Bmc Genomics 4.

Leptin, M., 1991 twist and snail as positive and negative regulators during Drosophila mesoderm development. Genes Dev 5**:** 1568-1576.

Levine, M., and E. H. Davidson, 2005 Gene regulatory networks for development. Proceedings of the National Academy of Sciences of the United States of America 102**:** 4936-4942.

Liang, G. N., J. C. Y. Lin, V. V. Wei, C. Yoo, J. C. Cheng *et al.*, 2004 Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. Proceedings of the National Academy of Sciences of the United States of America 101**:** 7357-7362.

Longabaugh, W. J. R., E. H. Davidson and H. Bolouri, 2005 Computational representation of developmental genetic regulatory networks. Developmental Biology 283**:** 1-16.

Mayr, C., and D. P. Bartel, 2009 Widespread Shortening of 3 ' UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. Cell 138**:** 673-684.

Morris, D. P., and A. L. Greenleaf, 2000 The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. Journal of Biological Chemistry 275**:** 39935-39943.

Ofarrell, P. H., 1992 Big Genes and Little Genes and Deadlines for Transcription. Nature 359**:** 366-367.

Peterson, M. L., E. R. Gimmi and R. P. Perry, 1991 The Developmentally Regulated Shift from Membrane to Secreted Mu Messenger-Rna Production Is Accompanied by an Increase in Cleavage-Polyadenylation Efficiency but No Measurable Change in Splicing Efficiency. Molecular and Cellular Biology 11**:** 2324-2327.

Pires, C. V., F. C. D. Freitas, A. S. Cristino, P. K. Dearden and Z. L. P. Simoes, 2016 Transcriptome Analysis of Honeybee (Apis Mellifera) Haploid and Diploid Embryos Reveals Early Zygotic Transcription during Cleavage. Plos One 11.

Proudfoot, N., 1991 Poly(a) Signals. Cell 64**:** 671-674.

Proudfoot, N. J., 2011 Ending the message: poly(A) signals then and now. Genes & Development 25**:** 1770-1782.

Reeves, G. T., N. Trisnadi, T. V. Truong, M. Nahmad, S. Katz *et al.*, 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, J., P. Early, C. Carter, K. Calame, M. Bond *et al.*, 1980 Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin mu chain. Cell 20**:** 303-312.

Rothe, M., M. Pehl, H. Taubert and H. Jackle, 1992 Loss of Gene-Function through Rapid Mitotic-Cycles in the Drosophila Embryo. Nature 359**:** 156-159.

Rusch, J., and M. Levine, 1997 Regulation of a dpp target gene in the Drosophila embryo. Development 124**:** 303-311.

Rushlow, C., P. F. Colosimo, M. C. Lin, M. Xu and N. Kirov, 2001 Transcriptional regulation of the Drosophila gene zen by competing Smad and Brinker inputs. Genes & Development 15**:** 340-351.

Russo, C. A. M., N. Takezaki and M. Nei, 1995 Molecular Phylogeny and Divergence Times of Drosophilid Species. Molecular Biology and Evolution 12**:** 391-404.

Sandler, J. E., and A. Stathopoulos, 2016 Quantitative Single-Embryo Profile of Drosophila Genome Activation and the Dorsal-Ventral Patterning Network. Genetics.

Savard, J., D. Tautz, S. Richards, G. M. Weinstock, R. A. Gibbs *et al.*, 2006 Phylogenomic analysis reveals bees and wasps (Hymenoptera) at the base of the radiation of Holometabolous insects. Genome Research 16**:** 1334-1338.

Severson, D. W., B. deBruyn, D. D. Lovin, S. E. Brown, D. L. Knudson *et al.*, 2004 Comparative genome analysis of the yellow fever mosquito Aedes aegypti with Drosophila melanogaster and the malaria vector mosquito Anopheles gambiae. Journal of Heredity 95**:** 103-113.

Shermoen, A. W., and P. H. Ofarrell, 1991 Progression of the Cell-Cycle through Mitosis Leads to Abortion of Nascent Transcripts. Cell 67**:** 303-310.

Shi, X. B., I. Kachirskaia, K. L. Walter, J. H. A. Kuo, A. Lake *et al.*, 2007 Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. Journal of Biological Chemistry 282**:** 2450-2455.

Staller, M. V., D. Yan, S. Randklev, M. D. Bragdon, Z. B. Wunderlich *et al.*, 2013 Depleting Gene Activities in Early Drosophila Embryos with the "Maternal-Gal4-shRNA" System. Genetics 193**:** 51-61.

Stathopoulos, A., and M. Levine, 2005 Genomic regulatory networks and animal development. Developmental Cell 9**:** 449-462.

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

Su, A. I., T. Wiltshire, S. Batalov, H. Lapp, K. A. Ching *et al.*, 2004 A gene atlas of the mouse and human protein-encoding transcriptomes. Proceedings of the National Academy of Sciences of the United States of America 101**:** 6062-6067.