Quantitative Single-Embryo Profile of *Drosophila* Genome Activation and the Dorsal-Ventral Patterning Network

Jeremy E. Sandler¹ and Angelike Stathopoulos¹,*

¹Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125 USA

*Author for correspondence: angelike@caltech.edu

Running title: Gene Expression Dynamics Identified by Single-Embryo Profiling

Keywords: *Drosophila melanogaster*; gene expression; transcription; dorsal-ventral patterning; maternal to zygotic transition; NanoString nCounter

Copyright 2016.
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, have 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.

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 patternning 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. 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. 2008). 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

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 RNAsese 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.

**Data Availability.** Table S2 lists probe sequences used for NanoString code set. Table S3 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 (Fig. 1A). 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 (Fig. 1B).

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 $232.84 \pm 11.52$ (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 Fig. 1C.

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 $7.97 \times 10^7$ copies of *dhd* and $2.63 \times 10^3$ copies of *pntr*, a fold-difference of over 30,000 (Fig. 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 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. 1G). 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 (Fig. 1F). 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 (Figs. 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 (Fig. 2C). Genes that rapidly decrease between NCs 14A and 14B are maternally deposited transcripts or are zygotic genes refined from broad to
narrow patterns (Fig. 2D). 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 (Fig. 3A). 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. In situ hybridizations using intronic probes show that there is in fact zygotic transcription of *bcd* detected as early NC11 (Fig. S2), 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. 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 (Figs. 3A-C). Both dual switching AP genes
included, bcd and hb, switch at NC12 (Fig. 3A) 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, snai, 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 (Fig. 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 (Fig. 4C’-F’), but have different boundaries on the AP axis (Fig. 4C-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. 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 (Figs. 4A and 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, 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 (Fig 4 G-J). 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 \textit{dpp} and \textit{scw} (Fig. 5A). Furthermore, both \textit{scw} and \textit{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-\(\beta\) targets in the study, and found that they are all strongly activated beginning in NC14. We separated TGF-\(\beta\) targets into two classes, based on how they are activated in NC14. Genes \textit{pnr}, \textit{hnt}, and \textit{Doc1} are expressed in a gradually increasing manner throughout NC14, until gastrulation when the rate of transcription levels off (Fig. 5B). In contrast, \textit{Race}, \textit{tup}, and \textit{ush} increase very quickly at the beginning of NC14 and reach a plateau as early as NC14B or 14C (Fig. 5B).

The TGF-\(\beta\) 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 \textit{ush}, \textit{Race}, and \textit{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-\(\beta\) signaling, when the genes are expressed in their final domains. TGF-\(\beta\) 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, \textit{ush} is always the most abundant, with \textit{Race} at around 60\% of the \textit{ush} levels and \textit{hnt} at around 22\% of \textit{ush} levels (Fig. 5F). However when the number of transcripts per each cell is calculated based on expression domain (Fig. 5C-E), the results change drastically. The difference between \textit{Race} and \textit{ush} drops to 1\%-4\% depending on the time point, and the difference between \textit{hnt} and \textit{Race} and \textit{ush} drops to 6\%-15\% depending on time point (Fig. 5G). 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 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 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.* 2008; Materna *et al.* 2010).

The diversity observed for both the maternal to zygotic switch point and the degradation rate for *can* 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 \textit{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, \textit{Race}, \textit{ush}, and \textit{hnt} are expressed only in the dorsal ectoderm, while \textit{pnr}, \textit{tup}, and \textit{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 \textit{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 \textit{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.

ACKNOWLEDGEMENTS

We thank and dedicate this paper to Eric Davidson (1937-2105), an inventor of NanoString technology, for helpful discussions and guidance, and who generously provided use of his NanoString instrument for this study. We also thank past and present Davidson Lab members Julius Barsi, Roberto Feuda, and Stefan Materna for technical assistance, advice using
the NanoString instrument and in data analysis, and comments on the manuscript. This work was funded by grant R01GM077668 from the National Institute of Health to A.S.
Figure 1. Timing of Maternal to Zygotic Transition in the syncytial blastoderm, experimental methods, and controls
(A) A diagram of the degradation of maternal transcripts and the accumulation of zygotic transcripts. Embryo age in minutes after egg laying and corresponding nuclear cycle are displayed. (B) Confocal time series of *Drosophila* embryos expressing Histone2Av-RFP fusion. Nuclear density is used to determine nuclear cycle for NCs 10-14, and nuclear elongation (expanded images) is used to stage embryos within NC14. (C) 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. (D) 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*≤0.001. In this and other figures, number of transcripts refers to counts measured from single embryos, done in triplicate and averaged. (E) 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*≤0.001. (F,G) qPCR comparing the abundance of *bcd* (F) and *sna* (G) 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.
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-b 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).
Figure 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≤0.001.
Figure 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.
Figure 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 $p \leq 0.001$. 
REFERENCES


