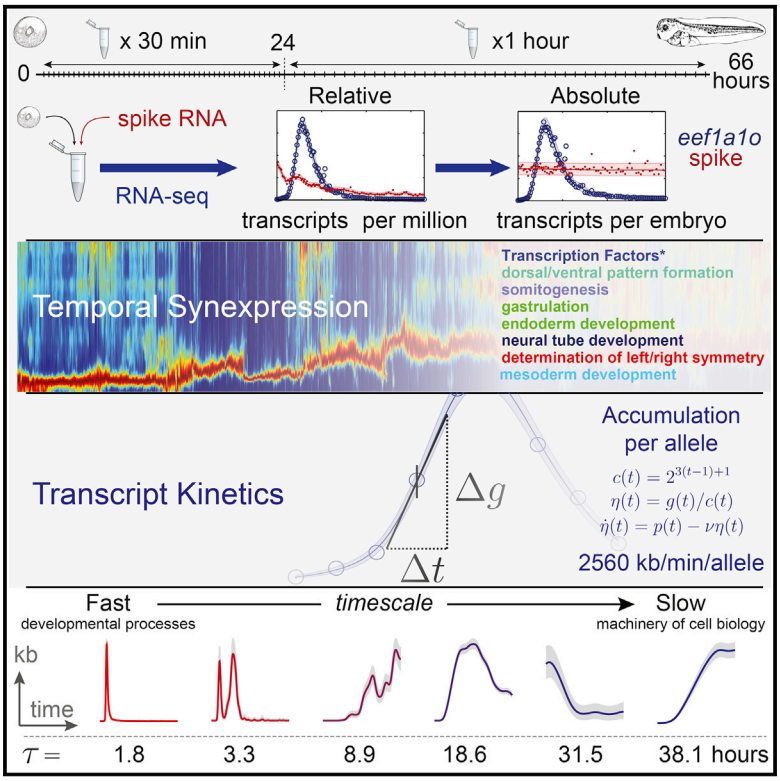
Resource

Cell Reports

Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development

Graphical Abstract



Highlights

- Ultra-high temporal resolution RNA-seq reveals smooth expression trajectories
- Absolute transcript measurements enable calculation of transcriptome kinetics
- Transcript accumulation rates vary over eight orders of magnitude during development
- Temporal synexpression and timescales distinguish gene functions

Authors

Nick D.L. Owens, Ira L. Blitz, Maura A. Lane, ..., Michael J. Gilchrist, Ken W.Y. Cho, Mustafa K. Khokha

Correspondence

mike.gilchrist@crick.ac.uk (M.J.G.), kwcho@uci.edu (K.W.Y.C.), mustafa.khokha@yale.edu (M.K.K.)

In Brief

Owens et al. present an ultra-high-resolution RNA-seq time course during embryonic development. Using spike-ins to calibrate transcript reads, they measure absolute mRNA transcript numbers and calculate transcriptome kinetics, in some cases, in kilobases per minute per allele. With these data, they find that temporal synexpression and characteristic timescales distinguish gene functions across the transcriptome.

Accession Numbers

GSE65785

Owens et al., 2016, Cell Reports *14*, 632–647
January 26, 2016
http://dx.doi.org/10.1016/j.celrep.2015.12.050

CellPress

OPEN ACCESS
CellPress

Cell Reports
Resource

Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development

Nick D.L. Owens,[1,6] Ira L. Blitz,[2,6] Maura A. Lane,[3,4] Ilya Patrushev,[1] John D. Overton,[4,5] Michael J. Gilchrist,[1,*] Ken W.Y. Cho,[2,*] and Mustafa K. Khokha[3,4,*]
[1]The Francis Crick Institute, Mill Hill Laboratory, The Ridgeway Mill Hill, London NW7 1AA, UK
[2]Department of Developmental and Cell Biology, University of California, Irvine, CA 92697 USA
[3]Program in Vertebrate Developmental Biology, Department of Pediatrics
[4]Department of Genetics
[5]Yale Center for Genome Analysis
Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520, USA
[6]Co-first author
*Correspondence: mike.gilchrist@crick.ac.uk (M.J.G.), kwcho@uci.edu (K.W.Y.C.), mustafa.khokha@yale.edu (M.K.K.)
http://dx.doi.org/10.1016/j.celrep.2015.12.050

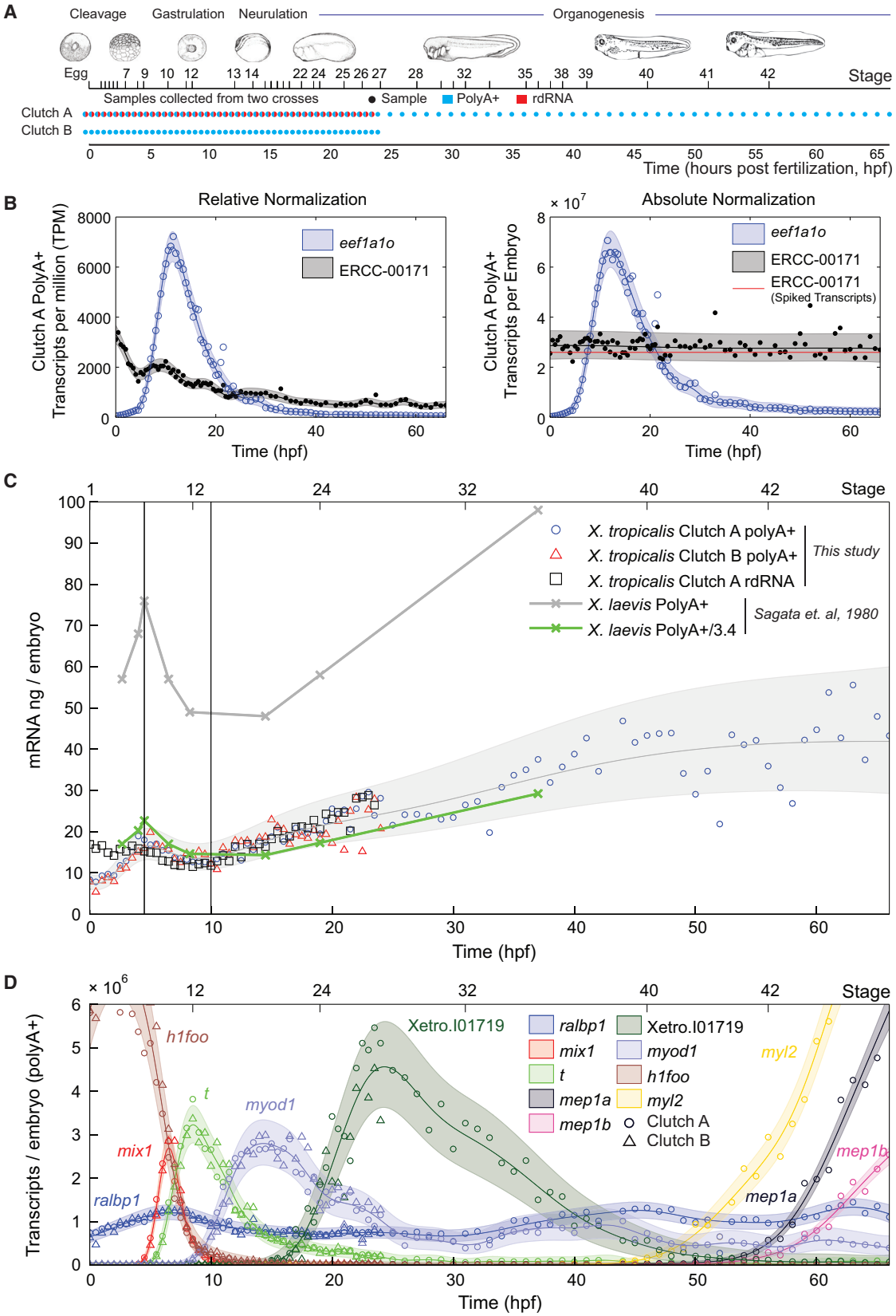
SUMMARY

Transcript regulation is essential for cell function, and misregulation can lead to disease. Despite technologies to survey the transcriptome, we lack a comprehensive understanding of transcript kinetics, which limits quantitative biology. This is an acute challenge in embryonic development, where rapid changes in gene expression dictate cell fate decisions. By ultra-high-frequency sampling of *Xenopus* embryos and absolute normalization of sequence reads, we present smooth gene expression trajectories in absolute transcript numbers. During a developmental period approximating the first 8 weeks of human gestation, transcript kinetics vary by eight orders of magnitude. Ordering genes by expression dynamics, we find that "temporal synexpression" predicts common gene function. Remarkably, a single parameter, the characteristic timescale, can classify transcript kinetics globally and distinguish genes regulating development from those involved in cellular metabolism. Overall, our analysis provides unprecedented insight into the reorganization of maternal and embryonic transcripts and redefines our ability to perform quantitative biology.

INTRODUCTION

Gene expression is dynamic and tightly regulated. To build a quantitative understanding of gene regulation, direct measurement of transcript kinetics is necessary. Transcript kinetics describe the rate of change of transcript copy numbers with time. In developing systems such as the embryo, dynamic transcript expression precisely coordinates a sequence of stereotypical events that occur in rapid succession. A quantitative understanding of the transcriptome during embryogenesis will have important applications for congenital malformation research and regenerative medicine (Fakhro et al., 2011; Tebbenkamp et al., 2014; Zaidi et al., 2013). With the ability to measure global transcript kinetics, we can effectively study the impact of different transcript regulation strategies on gene expression; for example, dynamics of chromatin modifications, utilization of *cis*-regulatory sequences, kinetics of transcription factor binding, and transcript stability. Because the study of gene regulatory networks is limited by the lack of genome-wide kinetic data (Karlebach and Shamir, 2008), a kinetic transcriptome dataset will be transformative to build and test gene regulatory network models in development.

The measurement of transcript kinetics has two requirements: (1) measurements of gene expression in absolute numbers of transcripts per cell or embryo and (2) sampling at a sufficiently high temporal resolution to properly calculate rates of change of transcript numbers. While there is great potential to use RNA sequencing (RNA-seq) to study transcript kinetics, the datasets that are currently available lack one or both of these attributes, and suitable methodologies have yet to be developed (Stegle et al., 2015). In most datasets, sampling time points are too widely spaced. Data points are effectively isolated "snapshots" in developmental time, and analysis is restricted to pairwise comparisons of differences in relative gene expression (Aanes et al., 2011; Fang et al., 2010; Kalinka et al., 2010; Paranjpe et al., 2013; Spencer et al., 2011; Tan et al., 2013; Vesterlund et al., 2011; Wang et al., 2004; Yanai et al., 2011). In addition, in each of these datasets, relative normalization precludes kinetic calculations. Relative normalization is ubiquitous within the field and provides expression levels in arbitrary units related to the number of reads that map to a transcript relative to the total number of reads from the library. These relative values do not relate directly to transcript numbers and are not reliable indicators of either the magnitude or the direction of change in gene expression between samples when RNA content changes. For



(legend on next page)

example, as the total amount of RNA increases in the developing embryo, transcripts with a constant copy number will appear to decrease in expression under relative normalization. Therefore, even in those studies in which sampling rates are high (Collart et al., 2014; Tomancak et al., 2002), relative normalization combined with varying total mRNA levels confound comparisons so that kinetic calculations cannot be made even in simple fold change terms.

Absolute transcript measurements can be achieved by spiking in RNAs at known abundances for use as quantification standards. Native transcripts are calibrated against RNA standards to calculate absolute transcript copy number estimates. Previous reports have used spike-ins to normalize for sequencing depth, to control for technical variability, and to calculate estimates of absolute transcript numbers (Brennecke et al., 2013; Islam et al., 2014; Junker et al., 2014; Risso et al., 2014). However, recent reports have suggested that such a direct absolute normalization with currently available RNA standards is infeasible due to significant sequencing biases (Risso et al., 2014; SEQC/MAQC-III Consortium, 2014). To date, a few RNA-seq datasets have been calibrated indirectly, where a small number of transcripts are first quantified using a non-sequencing technology that, in turn, is used to normalize RNA-seq data (Marguerat et al., 2012; Tu et al., 2014). Here, we show that direct absolute quantification using sequence data alone is feasible and yields accurate measurements of embryo RNA content.

Using RNA-seq and spike-in RNAs for absolute normalization, we created a genome-wide dataset to precisely measure transcript kinetics, which can be linked to biological processes occurring during embryogenesis (i.e., gastrulation or organogenesis such as for eye, muscle, and heart). To accomplish this, we sampled rapidly developing embryos at such high temporal resolution that data from neighboring time points were found to be as similar as biological replicates, ensuring that we captured precise transcript dynamics.

This rich dataset enables the visualization of gene dynamics of the entire transcriptome over an extended developmental period. In the following text, we outline a number of discoveries enabled by combining absolute normalization and high temporal resolution sampling.

RESULTS

Data Collection, Quality, and Overview of Analysis

To profile gene expression at high time resolution in the human disease model, *Xenopus tropicalis*, we collected eggs (time 0) and then synchronously developing embryos from an in vitro fertilization at 30-min intervals for the first 24 hr of development, followed by hourly sampling thereafter to 66 hr (Figure 1A). We sampled RNA from two crosses collected in parallel (Figure 1A), Clutch A and Clutch B. For Clutch A, we sequenced poly(A)+ RNA for the full 66 hr and also total RNA depleted of rRNA (rdRNA), containing both poly(A)+ and poly(A)− RNAs, for the first 24 hr. For Clutch B, we sequenced poly(A)+ RNA for the first 24 hr only.

To calculate absolute transcript numbers and kinetics, we calibrated read counts of all transcripts using spike-in RNAs as quantification standards. After embryo homogenization and prior to RNA extraction, we added a known amount of spike-in RNAs per embryo to each sample independently. This ensured that RNA standards underwent the same variations in recovery as the endogenous embryonic transcripts during RNA purification, library preparation, and sequencing. To Clutch A and Clutch B, we added the 92 synthetic RNAs available from the External RNA Control Consortium (ERCC) (Baker et al., 2005). In addition, we independently added three of the *E. coli*-derived ArrayControl spike-ins to Clutch B.

We aligned reads to the *X. tropicalis* genome, known off-genome expressed sequence tags (ESTs) sequences and spike RNA sequences, and calculated relative transcript abundances of an augmented version of *X. tropicalis* v7.2 models in transcripts per million (TPM). The overall quality of our datasets was excellent, according to a number of metrics (Figures S1A and S1B). Importantly, for capturing transcript kinetics, transcriptome-wide comparisons of neighboring time points were as correlated as biological replicates, as measured by transcriptome-wide Spearman correlation (Figure S1A). Four time points (among 139) and one RNA standard that performed poorly (Qing et al., 2013; SEQC/MAQC-III Consortium, 2014) were excluded from all subsequent analysis (Figures S1C–S1E; Supplemental Experimental Procedures). We also found that a small fraction of non-rRNA transcripts are depleted in the rdRNA dataset (Figure S2F; Table S2), indicating that our ribosomal-depletion protocol may have a small number of off-target effects.

Commonly, in large-scale genomics studies, unsupervised machine learning tools such as clustering methods (e.g., weighted correlation network analysis; WGCNA) (Langfelder and Horvath, 2008) or dimensionality reduction methods (e.g., principal-component analysis; PCA) (Jolliffe, 2002) can discover relationships within the data. In our case, PCA confirmed a strong temporal correlation inherent in our experimental design (Figure 1A; Figure S1F). Therefore, we opted for a data analysis framework that can explicitly describe these temporal correlations. Gaussian processes are used commonly in the physical sciences

Figure 1. Data Generation, mRNA Content of the Embryo, and Gene Expression Dynamics

(A) Experimental design and sample collection.

(B) Left: *eef1a1o* and ERCC-00171 abundance in relative normalization (TPM), Clutch A poly(A)+. RNA standard decreases with time. Right: absolute normalization of left panel. Red line indicates ERCC-00171 transcripts added based on the manufacturer's data sheet. The RNA standard is constant at the correct abundance.

(C) Total mRNA (poly(A)+ and rdRNA) in the embryo in nanograms with time. Gray region and line, respectively, mark Gaussian process 95% confidence intervals (CI) and median of Clutches A and B poly(A)+ data.

(D) Dynamics of gene expression of poly(A)+ RNA in transcripts per embryo. Circles mark Clutch A poly(A)+, and triangles mark Clutch B poly(A)+. The horizontal axis marks time (bottom) and Nieuwkoop and Faber (NF) developmental stage (top). Shaded regions mark Gaussian process 95% CIs of the data.

See also Figures S1 and S2.

and have received significant attention as general tools for machine learning (Rasmussen and Williams, 2006). They offer a statistical framework for representing and examining the strong temporal correlations present in our data. Gaussian processes have been successfully applied to a range of biological data (Äijö et al., 2014; Heinonen et al., 2014; Honkela et al., 2010; Stegle et al., 2010; wa Maina et al., 2014). Here, Gaussian processes allow us to address multiple topics within a single framework to elucidate transcript dynamics. We use them in absolute normalization, regression, estimation of confidence intervals, identification of differential expression, and the calculation of kinetics.

Direct Absolute Normalization of RNA-Seq Data

We achieved absolute normalization with a two-step procedure. Our first step is relative normalization for sequencing depth, resulting in transcript abundances in TPM. We note that the relative normalized abundances of our RNA standards decreased with time due to accumulating RNA in the embryo (Figure 1B left; Figure S2A). In our second step, we transformed these RNA standard abundances to be constant with time at the known amount spiked into each sample. To minimize the effects of technical noise of the RNA standards, we use Gaussian processes to learn a function that varies smoothly with time to transform the RNA standards. This smooth mapping is critical, as it ensures that absolute normalized copy numbers are not contaminated by sample-to-sample technical noise in the RNA standards.

Recent work has demonstrated that ERCC spikes exhibit significant sequencing biases, and their appropriateness for direct absolute normalization in RNA-seq has been questioned (Risso et al., 2014; SEQC/MAQC-III Consortium, 2014). To investigate these concerns, we evaluated the consistency of our proposed absolute normalization on our three datasets independently: Clutch A poly(A)+, Clutch A rdRNA, and Clutch B poly(A)+ (Figure S2B). In agreement with previous observations (Qing et al., 2013; Risso et al., 2014; SEQC/MAQC-III Consortium, 2014), we find that (1) ERCC spikes perform poorly in poly(A)+ sequencing when compared to rdRNA sequencing (Figures S1F and S2B), (2) spike-in variation between the two independently spiked poly(A)+ datasets (Clutches A and B) is smaller than between Clutch A poly(A)+ and Clutch A rdRNA (Figure S2B), and (3) spike-in performance varies across spike-in species but is consistent over replicates (Figure S2C). We apply corrections for the first two sources of variation (Figure S2B; Supplemental Experimental Procedures) and then calculate absolute transcripts per embryo for native transcripts and RNA standards. We capture the Clutch A/B variation and the spike-in species sequencing biases in a single uncertainty model for the absolute normalization (Figure S2D). Our absolute normalization performs well with $R^2 = 0.97–0.98$ (Figure S2C) and has a 1.11- to1.25-fold error when comparing the ERCC and ArrayControl spike-ins that were added independently to Clutch B samples (Figure S2C). We discuss the limitations of this normalization, the validity of our Gaussian process models, and influences of gene model quality in Supplemental Experimental Procedures and Table S2.

The detection limit, which is the number of transcripts required to produce a single read on average, increases with time as mRNA accumulates in the embryo (Figure S2E). Averaged over all samples, the detection limit is ~1,300 transcripts per embryo, which is less than 1 transcript per cell once the embryo has attained the cell number present in the blastula.

Total mRNA Content of the Embryo Validates Absolute Normalization

We investigated the levels of total mRNA in the embryo during development. We summed the absolute abundance of all transcripts measured to derive the total mRNA in nanograms per embryo at each time point for both poly(A)+ and rdRNA (Figure 1C). Although there is a notable wave between 0 and 10 hours post-fertilization (hpf) (discussed later), the total amount of poly(A)+ mRNA increases with time from 10–15 ng of mRNA per embryo at fertilization to 30–50 ng of mRNA per embryo at 66 hpf (swimming tadpole, stage 42). These values validate our absolute normalization, as they agree well with total RNA yields after tissue homogenization (~1.3 μg per embryo and ~2.4 μg per embryo at the earliest and latest stages of our time course, respectively), where ~1%–2% of total RNA is polyadenylated (Davidson, 1986). As further validation of the absolute normalization, we compare our *X. tropicalis* mRNA predictions to experimentally measured poly(A)+ mRNA yields from *X. laevis* (Sagata et al., 1980). Accounting for the uncertainty in our absolute normalization, the mean ratio between these *X. laevis* mRNA yields and our data is 3.38 ± 0.02 (mean ± SD) (Figure 1C). This is in good agreement with the ratio in volume between *X. laevis* and *X. tropicalis* eggs; their respective diameters are 1.19 ± 0.07 mm and 0.80 ± 0.05 mm (Crowder et al., 2015), giving a volume ratio of 3.31 ± 0.76. Therefore, our absolute quantification agrees well with measurements of in vivo mRNA levels determined by an independent experimental method.

Transcript Dynamics in the Developing Embryo

Examining the expression of individual genes, we observe smooth transitions in gene expression (Figure 1D). Our sampling rate is sufficient to capture the dynamics of the expressed genes (see timescale analysis given later, Figure 7). We find genes whose expression peaks at different times in development, demonstrating transcript kinetics that are highly dynamic and suggesting specific roles for these genes during defined developmental periods (Figure 1D). For example, in the case of muscle induction, a series of genes critical for mesoderm and then somite specification accumulate and extinguish at different times as development progresses (Figure S3A). The temporal expression profiles of these genes agree with RNA detection by in situ hybridization in whole-mount embryos (Figure S3A). We further examined genes expressed later in development, which also correlated with in situ hybridization data (Figures S3B and S3C).

There is utility in identifying genes with constant expression, as these may serve as optimal loading controls in various experimental regimens (Figure S4). We found 109 genes (<0.7%) in poly(A)+ (0–66 hpf) and 1,078 genes (6.9%) in rdRNA (0–24 hpf) that had less than a 2-fold change (Figure S4A; Table S3). Commonly used loading controls (*gapdh*, *odc1*, *eef1a1*) have more variable expression (Figure S4B). Our data would suggest that the RNA helicase *ddx3x* has the best loading control performance across developmental time and RNA-seq preparations (Figure S4C).

Differential Expression between Biologically Independent Replicates Reveals Precise Regulation of Transcripts

Our dense temporal sampling offers a unique ability to detect differential expression over the entire time course. We began by exploring differential expression between our Clutch A/B biological replicates (Figure 2; Supplemental Experimental Procedures). We found that 12,062 of 16,914 genes (71%) had identical expression dynamics between the two clutches. Additionally, 4,561 genes (27%) showed differential expression with a small statistical effect (Figures 2A and 2B); their expression profiles show only minor differences (Figure 2C). Thus, expression dynamics of 98% of the transcriptome are similar between two independent biological replicates. This highlights both the high degree of reproducibility in our dataset and the exceptionally robust control of gene expression during development. The remaining 291 genes (1.7%) had a large statistical effect, indicating strong differential expression. The majority of these genes have differential maternal expression, but most Clutch A and Clutch B differences converge by 10–12 hpf, displaying identical expression later in development (e.g., *dhx32*; Figure 2D). Interestingly, this set of genes is enriched for Gene Ontology (GO) terms relating to cell division (Table S4), suggesting roles in the rapid cell division of the early embryo. Notably, a smaller number of genes have expression profiles that are divergent between replicates (Figure 2D). From these data, we conclude that gene expression is tightly regulated throughout embryogenesis and highly reproducible when measured using high-throughput sequencing.

Differential Expression and Switching of Transcript Isoforms during Development

Next, we analyzed the differential expression of alternative transcript isoforms. We can classify differential isoform expression as either "differential abundance" or "differential temporal dynamics." Transcripts are defined as having differential abundance if one transcript is expressed at a different level than the other, with a constant ratio between them, or as having differential dynamics if the ratio varies with time (Figure S5). Here, we restrict our attention to those isoforms that have differential temporal dynamics (Figure 3; Figure S5). We found 761 genes with isoforms showing differential dynamics, with differences in promoters, exons, and 3′ UTRs. Isoform expression profiles for 147 genes show a switchpoint where their expression profiles cross, marking a change in which isoform is most abundant (Figure 3A). The timing of these switchpoints was not evenly distributed over development; two thirds are transitions from a maternal to a zygotic isoform concentrated around 10 hpf (Figure 3A). Interestingly, we found examples of isoform switching due to post-transcriptional regulation; two maternal isoforms of *bicd2* exhibit differential polyadenylation (Figure 3B). One isoform *bicd2(1)* has a long 3′ UTR and is polyadenylated upon fertilization, the other *bicd2(2)* has a short 3′ UTR and is deadenylated upon fertilization (Figure 3C). The 3′ UTR of *bicd2(1)* must contain control elements to remain deadenylated during oocyte maturation and become polyadenylated after fertilization.

We predicted that isoforms with different temporal dynamics may show different spatial as well as temporal expression, implying different function. We selected three genes where the sequence differences between isoforms were long enough to generate effective in situ hybridization probes. Remarkably, all three showed different spatial expression domains (Figure 3D). This demonstrates the biological importance of isoform dynamics during embryogenesis and suggests specific mechanisms to regulate isoforms in different tissues. Moreover, we demonstrate the power of high-resolution sampling for transcript discovery and gene modeling. The small fragment sizes associated with high-throughput sequencing present challenges for identifying the co-expression of distant exons in long transcripts. The temporal statistics here can aid in the resolution of long-range order; distant exons that are expressed in the same isoform will exhibit related temporal dynamics (Figure 3C).

Temporal Synexpression Predicts Common Gene Function

To explore the expression dynamics of the entire dataset, we created a heatmap of transcripts organized according to a hierarchical clustering of their normalized expression profiles (Figure 4A). The heatmap is ordered so that nearest neighbors have similar expression profiles, which we term "temporal synexpression." Synexpression suggests that genes are controlled by shared regulatory networks and may have common function (Niehrs and Pollet, 1999). To investigate temporal synexpression, we used a sliding window to assess local GO enrichment across the heatmap (Figure S6A; Table S4). Interestingly, we found many blocks of genes enriched for specific functions ranging from cellular biology to developmental patterning events (Figure 4A). In the set of genes showing early transient expression, we found GO terms associated with early developmental steps, including germ layer specification, mesoderm/endoderm development, and axis patterning (Figure 4A, right, top of heatmap). Genes transiently expressed later in development are enriched for GO terms associated with organogenesis (Figure 4A, right, bottom of heatmap). Therefore, proximity within the heatmap enriches for GO terms that are consistent with developmental events, demonstrating that temporal synexpression is a powerful predictor of gene function (see below).

While exploring temporal synexpression, we found two distinct sets of genes (S1 and S2 in Figures 4A and 4B, left; Supplemental Experimental Procedures) that have similar oscillatory behaviors between 34 and 66 hr of development. These sets are not immediately adjacent in the overall heatmap due to different expression prior to 34 hr. Postulating that there may be more genes across the heatmap with similar oscillations and, hence, similar function, we used one of these genes, *ckb* (Figure 3D, right), as a "seed" to identify other similarly oscillating genes and discovered a much larger set (Table S5, 150 genes). To determine whether this *local* temporal synexpression identified common functions, we examined spatial expression patterns. Our seed gene, *ckb*, has pronounced expression in the somitic mesoderm and lies within a block of genes enriched for muscle contraction GO terms. Remarkably, our larger local temporal synexpression set also has significant enrichment in somitic mesoderm. 38 of these genes have expression patterns available in public databases (Bowes et al., 2010), and 34/38 (89%) have somitic expression (Figure S6B). These genes appear to oscillate with a period of approximately 20 hr, which is significantly longer than the somitic

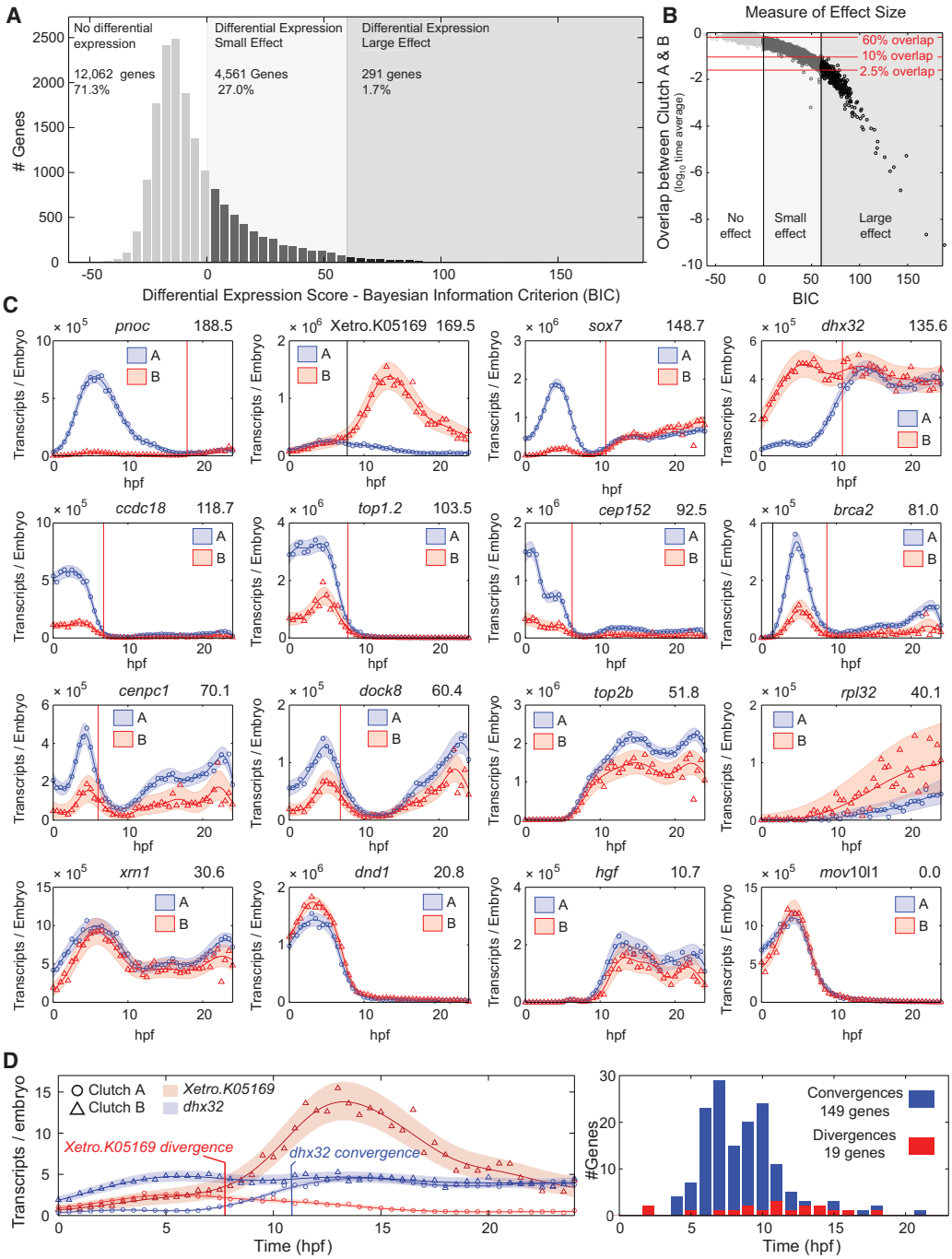


Figure 2. Clutch A versus Clutch B Poly(A)+ Differential Expression

(A) Clutch A versus Clutch B histogram of differential expression scores as measured by the Bayesian information criterion (BIC); larger scores indicate greater differential expression (Supplemental Experimental Procedures). Three regions marked: (1) no differential expression, BIC $\leq$ 0; (2) differential expression small effect, 0 < BIC $\leq$ 60, and (3) differential expression large effect, BIC > 60. See (B) for explanation of thresholds.

(B) Differential expression effect size as measured by the $\log_{10}$ mean overlap between Clutch A and Clutch B Gaussian process models (Supplemental Experimental Procedures). Mean overlap decreases with increasing BIC. Thresholds at BIC = 0, 60 in addition to mean overlaps of 60%, 10%, and 2.5% are marked. At BIC = 60, approximately all genes have less than 10% overlap, and all genes with less than 2.5% overlap have a BIC > 60.

(C) Differential expression examples with decreasing BIC (top right). Genes on the boundary for strong differential expression have highly correlated expression profiles in (A) and (B). Vertical lines mark convergences (red) or divergences (black).

(D) Left: examples of convergence (*dhx32*) and divergence (*Xetro.K05169*). Right: histogram of convergence and divergence times.

See also Supplemental Experimental Procedures.

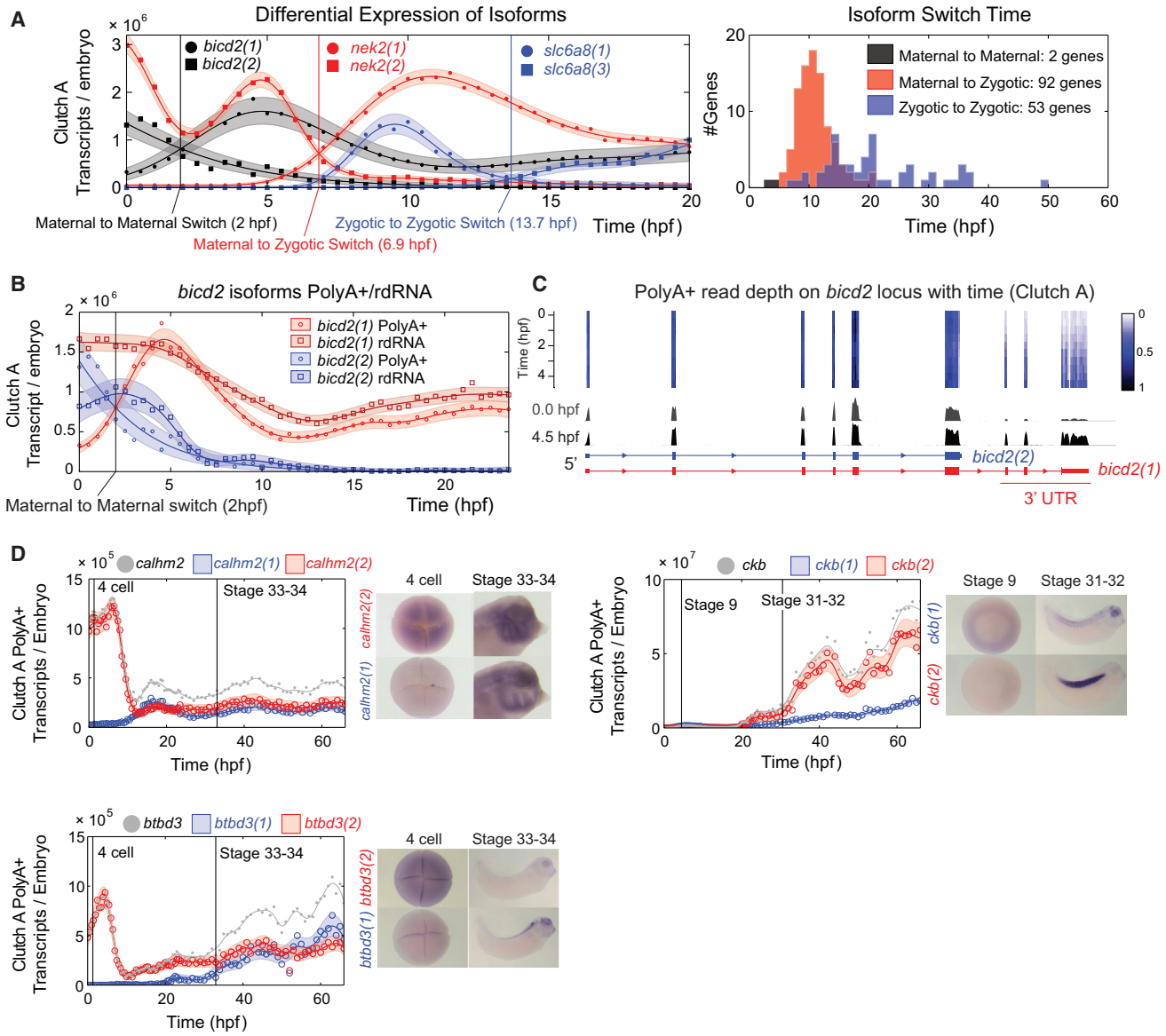


Figure 3. Isoform Differential Expression in Clutch A

(A) Isoform switching event examples: maternal to maternal; maternal to zygotic; and zygotic to zygotic switching events (left). Histogram of isoform switching events by category and time (right).

(B) *bicd2* isoforms switch in poly(A)+ data (circles) but do not switch in rdRNA data (squares), indicating differential polyadenylation. Poly(A)+ and rdRNA abundances agree within uncertainty of the absolute normalization (Supplemental Experimental Procedures).

(C) Normalized read depth on *bicd2* locus in Clutch A poly(A)+ between 0 and 4.5 hpf. Note temporal dynamics shared by the final three *bicd2(1)* exons. Heatmap and depth of pile-ups corrected for changing total mRNA.

(D) Spatial expression of three genes with isoforms with differential dynamics. All isoforms show different temporal and spatial expression domains. Gray lines and points mark total expression for each gene (sum of the red and blue isoforms).

See also Figure S5 and Supplemental Experimental Procedures.

clock (Pourquié, 2003) and more reminiscent of the circadian rhythm. Interestingly, a few circadian clock genes are expressed in the somites (Curran et al., 2008, 2014), and our analysis indicates that there may be additional unexplored links between clock genes and somite regulatory networks.

In another case, we found two sets of genes (V1 and V2; Supplemental Experimental Procedures) with similar temporal expression patterns from 40 to 66 hpf that enrich for visual perception (GO: 0007601). With the exception of *tmem145*, all V1 and V2 genes are well characterized as having roles in rod and cone cells and are associated with human retinal diseases (Table S5). Therefore, we predicted that the uncharacterized *tmem145* also plays a role in vision; and, indeed, *tmem145* does have spatial expression in the eye (XDB3 [Bowes et al.,

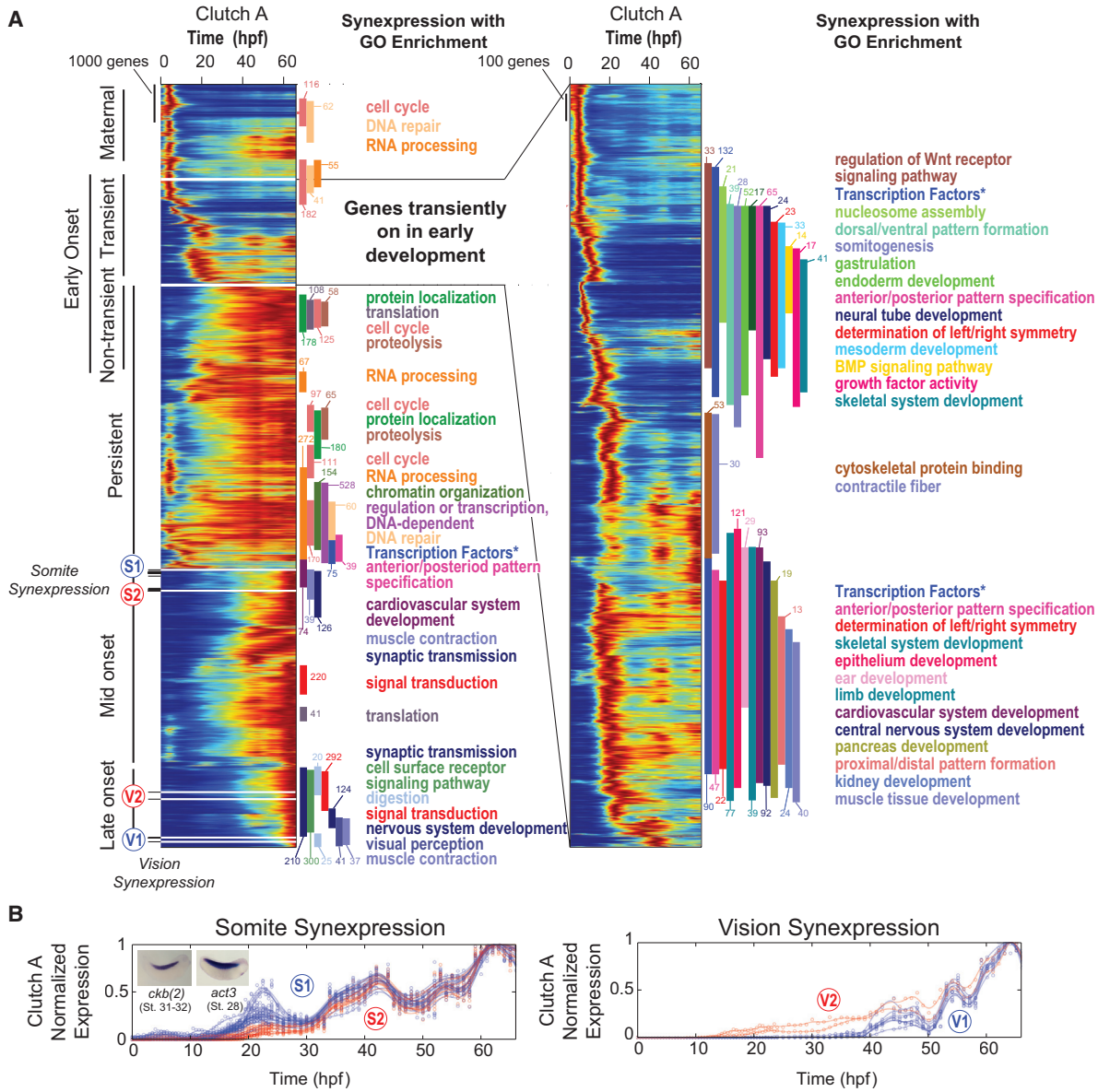


Figure 4. Temporal Synexpression in Clutch A Poly(A)+

(A) Temporal map of the transcriptome: an enumeration of all gene expression dynamics in the embryo. Heatmaps display all genes normalized by maximum expression and ordered by similarity. Inset (right) expands on genes transiently expressed with early onset. Vertical bars mark GO enrichment (colors correspond to GO terms), numbers appended to GO bars indicate the number of genes of given category. All reported GO blocks are enriched with $p < 2 \times 10^{-4}$ (Fisher's exact test) for entire block. S1, S2, V1, and V2 mark the locations of somite and vision temporal synexpression genes, respectively, in (B). "Transcription Factors*" labels an annotation of transcription factors separate to GO (Supplemental Experimental Procedures).

(B) Somite (left) and vision (right) synexpression groups.

See also Figure S6 and Supplemental Experimental Procedures.

2010] and XGC EST database [Klein et al., 2002; Morin et al., 2006]).

We conclude that temporal synexpression can be used effectively to predict gene function and predictions can be refined depending on the local developmental period of interest. Remarkably, we uncovered a large cohort of genes expressed in the somites, simply by ranking their similarity in expression to a gene with known somite expression. We anticipate that future investigations of temporal synexpression in this data may be able to assign putative functions to many of the 5,718 genes (~28%) that remain unnamed.

The Reorganization of Maternal and Zygotic Transcriptomes

In our analysis of total mRNA during development (Figure 1C), we noted a difference between total rdRNA levels and poly(A)+ mRNA

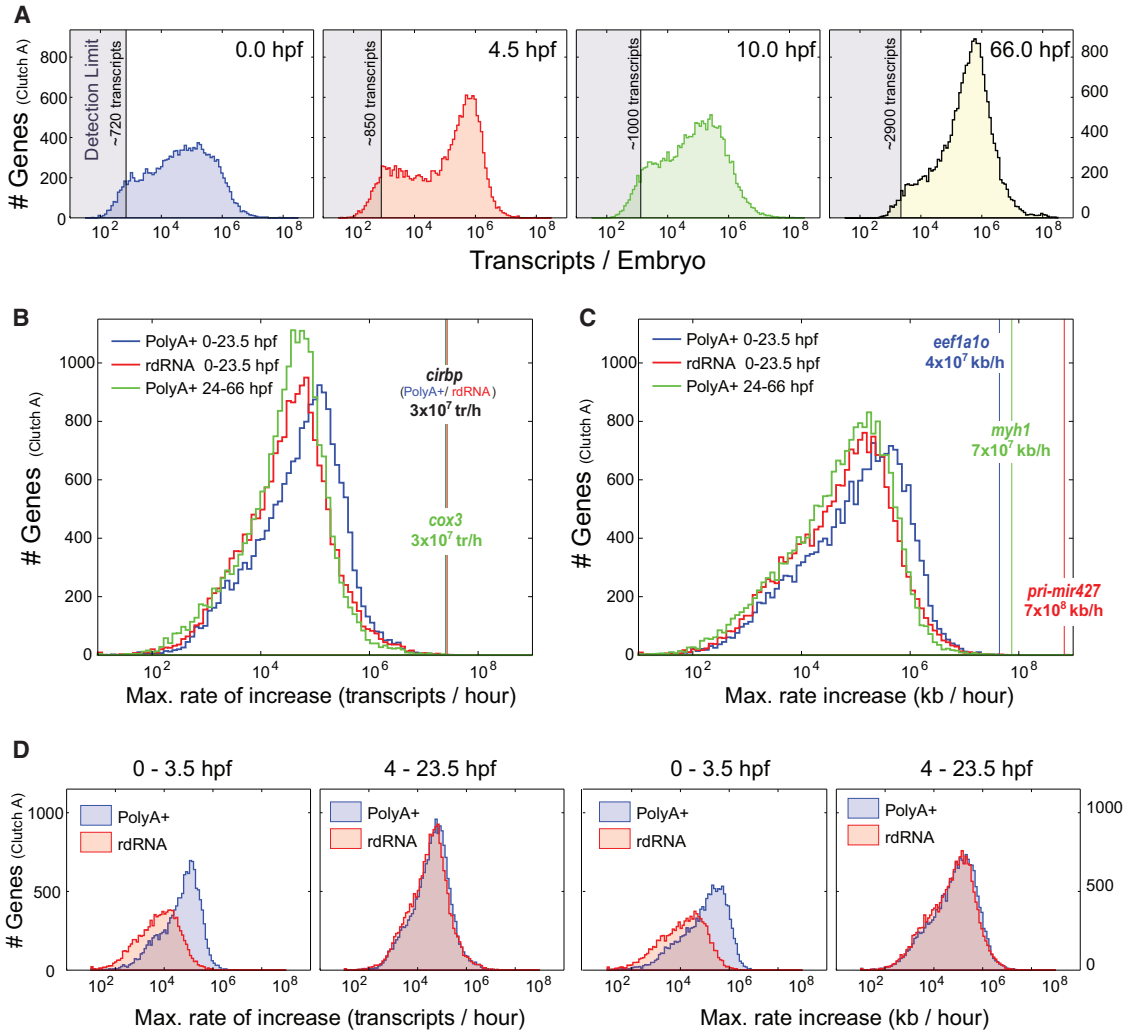


Figure 5. Transcript Copy Numbers and Kinetics per Embryo in Clutch A

(A) Transcript copy number histogram in poly(A)+ sequencing at time 0.0 (egg), 4.5 hpf (early stage 9, blastula), 10 hpf (stage 12.5, end of gastrulation), and 66 hpf (tadpole stage 42). Detection limits give number of transcripts required to generate a single read averaged over all transcript lengths (Figure S2E). Transcript copy number distributions transition smoothly between these time points (data not shown).

(B) Histogram of maximum rates of increase in transcripts per hour between any consecutive measurement points calculated from Gaussian process medians for each gene. Blue indicates poly(A)+ 0–23.5 hpf; red indicates rdRNA 0–23.5 hpf; green indicates poly(A)+ 24–66 hpf. Vertical lines mark the gene with maximum rate of increase for each category.

(C) As in (B), but in kilobases per hour.

(D) Maximum rate of increase in transcripts per hour (left) and kilobases per hour (right) for 0–3.5 hpf and 4–23.5 hpf time intervals. Poly(A)+ and rdRNA distributions are discrepant for 0–3.5 hpf, reflecting polyadenylation of maternal RNAs. Distributions are identical over 4–23.5 hpf.

See also Figure S2E.

levels prior to 4.5 hpf (early stage 9 blastula), the point at which zygotic transcription becomes widespread. From 0 to 4.5 hpf, there is an increase in poly(A)+ transcripts, while rdRNA levels show little change, reflecting the polyadenylation of maternal transcripts (Collart et al., 2014). Then, from 4.5 to 10 hpf, both poly(A)+ and rdRNA levels fall, as the clearance of maternal mRNA exceeds nascent zygotic transcription. The fall in RNA constitutes an ~20% loss of the embryo's mRNA content, supporting the notion that very early embryonic development relies heavily on maternally stored transcripts. By 10 hpf (late gastrula stage 12), those maternal mRNAs targeted for clearance are largely extinguished (see top left, Figure 4A), and zygotic transcription causes mRNA levels to increase for the remainder of our time course.

To examine these transitions more closely, we considered the distribution of the number of transcripts per gene at different time points (Figure 5A). Initially, in the egg, the distribution has a broad peak at ~100,000 transcripts, which changes dramatically at 4.5 hpf into a bimodal distribution with a marked peak at ~600,000 transcripts. This change in distribution reflects the rapid increase in polyadenylated maternal transcripts. The

distribution is again reorganized at 10 hpf, losing the bimodal appearance and shifting leftward toward fewer transcripts as maternal transcripts are eliminated. By the end of our time course, the peak at ~600,000 transcripts is re-established without the bimodal appearance seen at 4.5 hpf. The coincidence of a mode at ~600,000 transcripts at 4.5 hpf and 66 hpf is unexpected. The tadpole at 66 hpf has twice the mRNA content and is considerably more complex than the 4.5 hpf blastula-stage embryo. Therefore, the doubling of the RNA content is achieved by having roughly twice as many genes with at least ~600,000 transcripts per embryo rather than simply shifting the entire distribution (rightward) so that there is twice the copy number of each gene overall.

Transcript Kinetics Reveal a Rapid Deployment of Maternal Transcripts

To examine these changes in the distribution of transcript numbers in more detail, we calculated transcript kinetics. We determined the maximum rate of increase (per embryo) in transcripts per hour and kilobases per hour and examined the distributions of all genes (Figures 5B and 5C). We compared these maximum rate distributions between rdRNA and poly(A)+ over 0–24 hpf and then between rdRNA and poly(A)+ over 24–66 hpf. Strikingly, the maximum rates of accumulation vary over eight orders of magnitude, suggesting varied mechanisms for the regulation of transcript accumulation. The distributions for transcripts per hour and kilobases per hour are similar, indicating that the rate distribution is not influenced by transcript length (Pearson correlation of 0.28 between maximum accumulation rate and transcript length in Clutch A poly(A)+). Both the distributions for transcripts per hour and kilobases per hour are negatively skewed, producing the sharp rise on the right tail of the distribution. This indicates that many genes experience a rate of accumulation that is close to the maximal and that most genes experience a rate of accumulation that is greater than the mean.

Interestingly, genes measured from 0 to 24 hpf in poly(A)+ preparations show faster rates (blue line shifted rightward, Figures 5B and 5C) than in both 0–24 hpf rdRNA and 24–66 hpf poly(A)+, which are aligned. Given the reorganization of transcripts from 0 to 4.5 hpf (Figure 5A), we compared maximum accumulation rates in poly(A)+ and rdRNA at this time interval. At 0–3.5 hpf, the maximum accumulation rates for poly(A)+ are faster than for the rdRNA; in contrast, this difference is lost at 4–24 hpf, where the poly(A)+ and rdRNA distributions are coincident (Figure 5D). We conclude that the polyadenylation of maternal transcripts offers the cleavage-stage embryo a mechanism by which a large number of transcripts can be deployed for translation very rapidly, overcoming the embryo's limited capacity for transcription (due to very rapid cell cycling and relatively few nuclei per embryo).

Rapid Kinetics of *pri-mir427* and the Clearance of Maternal Transcripts

The clearance of maternal RNAs is one of the major embryo-wide transcriptome changes, resulting in a net loss of ~20% of the total mRNA content (Figure 1D). The mir427/430 family, in *Xenopus* and zebrafish, respectively, plays a critical role in clearing maternal RNAs at the onset of zygotic transcription (Giraldez et al., 2006; Lund et al., 2009). Early embryonic phenotypes due to the loss of *dicer*—and, therefore, all miRNAs—can be rescued in zebrafish embryos by mir430, highlighting the importance of this microRNA (Giraldez et al., 2005). Interestingly, we find that the *pri-mir427* transcript has the fastest kinetics of any transcript in the early frog embryo, at 7×10^8 kb/hr per embryo (Figure 5C), an order of magnitude faster than the very abundantly expressed *ef1a1* (Figure 5C). In addition, we found *pri-mir427* transcripts to be ubiquitously expressed, and we first detected their transcription at the eight-cell stage (2 hpf), significantly earlier than the previously identified onset of *nodals 3*, *5*, and *6* and well before the classic midblastula transition (Figure 6A, inset; Figure S7) (Kimelman et al., 1987; Newport and Kirschner, 1982; Skirkanich et al., 2011; Yang et al., 2002). Therefore, *pri-mir427* is not only transcribed at a prodigious rate but is also the earliest detected zygotic transcript in *Xenopus*.

As *pri-mir427* is expressed during a period when precise cell numbers are known, we can calculate the average rate of transcription per allele from rdRNA data (Supplemental Experimental Procedures). The *mir427* locus contains numerous copies of the *mir427* hairpin covering ~55 kb of transcribed genome (Figure 6D; Data S2). Here, we report the aggregate RNA abundance and accumulation over the entire locus.

The abundance of *pri-mir427* transcripts per cell peaks at 4.4 hpf (130 Mb per cell), with the peak accumulation rate per allele occurring at 4.2 hpf (2.6 Mb/min per allele; Figures 6B and 6C). Performing a similar analysis for other early transcribed genes (Figure S7B), we found that *pri-mir427* accumulates nearly 1,000-fold more rapidly. We demonstrate that it is the size of the *mir427* locus combined with ubiquitous expression that renders such a dramatic accumulation rate possible. We note that *pri-mir427* is RNA polymerase II (pol II) transcribed (Lund et al., 2009). The maximum RNA pol II elongation rate has been estimated at ~4 kb/min (Ardehali and Lis, 2009). At this elongation rate, the locus requires a polymerase density of 11–12 pol II/kb to achieve the measured 2,560 (95% CI: 2,280 to 2,840) kb/min per allele. This is in good agreement with the maximum polymerase densities of ~10 pol II/kb observed on rapidly transcribed genes of amphibian oocyte lampbrush chromosomes (Miller and Hamkalo, 1972). Corroborating this prediction, we examined published *X. tropicalis* blastula stage 9 RNA pol II chromatin immunoprecipitation sequencing (ChIP-seq) data (van Heeringen et al., 2014) and found high RNA pol II occupancy over the *mir427* locus (Figure S7A). In summary, we conclude that the *pri-mir427* achieves the highest known rate of transcript production during the very early rapid cell divisions of the frog embryo. To achieve this brisk rate, the size of the locus and the ubiquitous expression (supported by its spatial expression) are essential; however, the RNA pol II density and transcription rate need not have extraordinary values. This does imply that *pri-mir427* is transcribed homogenously by all cells; any significant heterogeneity would require some cells to transcribe at rates beyond what has been observed.

We speculate that numerous copies are present in *Xenopus* and zebrafish to maximize production of mature miR-427/miR-430 microRNAs during the rapid development of

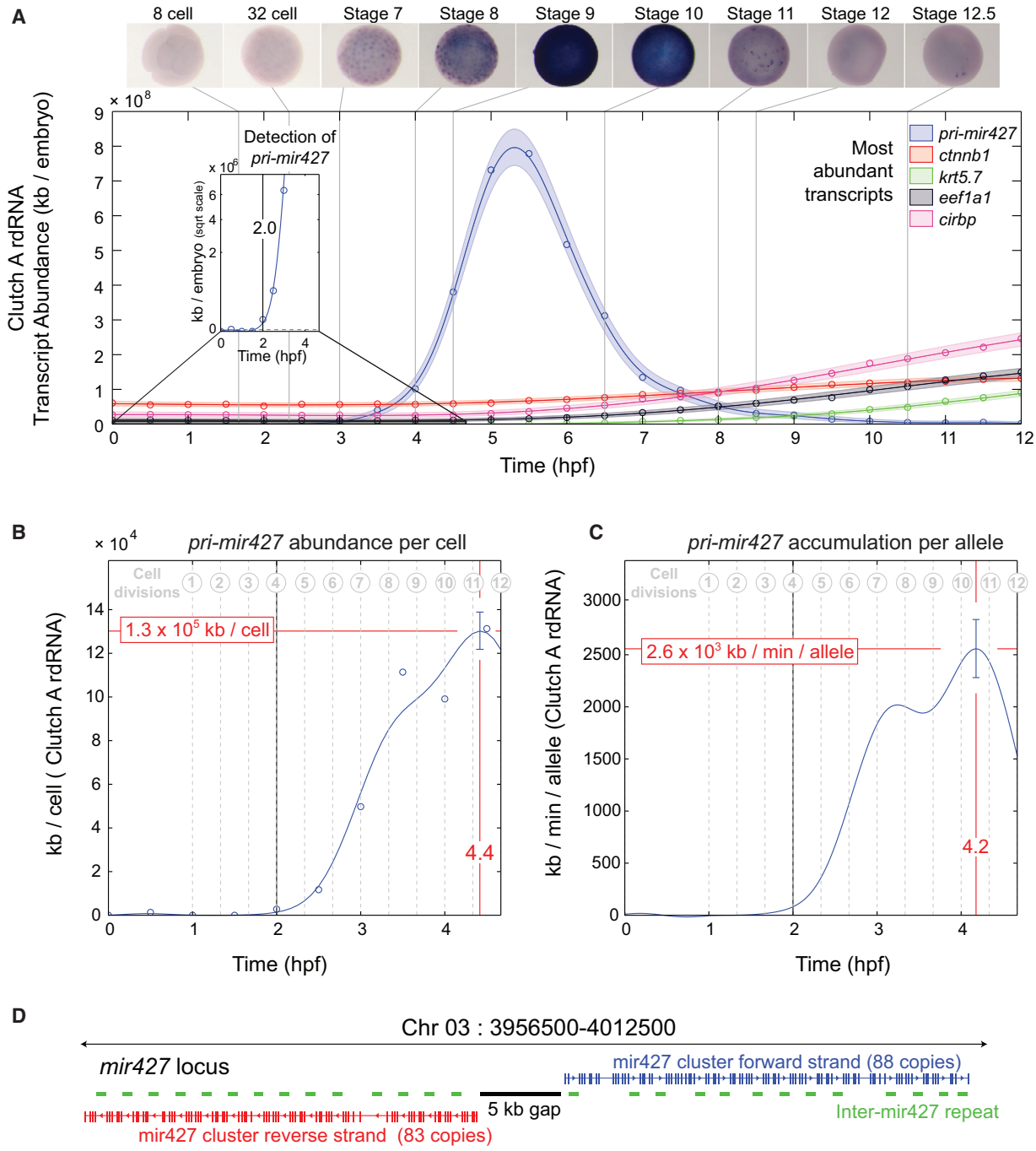


Figure 6. Accumulation Kinetics of *pri-mir427* per Allele in Clutch A rdRNA

(A) *pri-mir427* expression over first 12 hpf in rdRNA data and in situ hybridizations. *pri-mir427* show ubiquitous expression matching RNA-seq temporal data. Examples of most abundant transcripts shown for comparison. Inset: first detection of *pri-mir427* above detection limit at 2.0 hpf (8- to 16-cell transition).

(B) *pri-mir427* abundance in kilobases per cell. Line marks Gaussian process median; error bar indicates 95% CI (Supplemental Experimental Procedures). Peak per cell abundance occurs after 11th division at 4.4 hpf.

(C) *pri-mir427* accumulation rate in kilobases per minute per allele with time. Line indicates median of differential of Gaussian process; error bar indicates 95% CI (Supplemental Experimental Procedures). Peak accumulation rate is achieved at 4.2 hpf just after tenth cell division.

(D) Organization of *pri-mir427* locus on *X. tropicalis* chromosome 3 (*X. tropicalis* v8 genome; Supplemental Experimental Procedures). Clusters of *mir427* hairpins appear in tandem with a repeated sequence arranged symmetrically on opposite strands around a gap in the genome assembly.

See also Figure S7 and Supplemental Experimental Procedures.

these organisms. Interestingly, the mammalian member of the family, *mir302*, is present in far fewer copies on the genome, presumably because mammalian development is dramatically slower and the embryos are smaller (with less maternal mRNA), and thus there is ample time to generate sufficient amounts of *mir302* to clear maternal transcripts.

Interestingly, the earliest detection of *pri-mir427* occurs at the eight-cell stage, when its accumulating transcript breaks our detection limit. The next transcriptional events we detect are from loci that are, on average, 19 times smaller: *nodal3/5/6* and *siamois1/2* break our detection limit at 32- to 256-cell embryos (Figure S7B), which contain 4- to 32-fold more copies of these loci than eight-cell embryos. It is possible that these shorter loci may be transcribed from the eight-cell stage or earlier, but our expression measurements are insufficiently sensitive to detect this early transcription. Together, we conclude that we are able to detect the onset of zygotic transcription vastly earlier than previously thought.

In principle, the approach taken to quantify *pri-mir427* transcript kinetics can be applied to any gene during any time interval in our time course, once the numbers of expressing cells per embryo are ascertained.

Characteristic Timescale of Gene Expression Classifies Developmental Gene Function

While the rate of accumulation of *pri-mir427* is remarkable, transcript accumulation rates may not be optimal for identifying potent developmental regulatory genes, where small changes in transcript levels may be sufficient to alter cell fate and morphogenesis. We note that the distribution of maximum transcript accumulation rates does not differentiate transcription factors from all genes (Figure 7A), whereas ribosome-related genes accumulate significantly faster. We postulated that developmentally regulated genes would not necessarily be transcribed at maximal rates but would be tightly regulated and transiently expressed, reflecting changing biological events (i.e., their expression would turn on and off rapidly). Using Gaussian processes, we calculate a parameter to quantify such behavior, a characteristic timescale (τ) for each gene.

Formally, the timescale parameterizes the Gaussian process covariance function for each gene (Supplemental Experimental Procedures). The covariance function describes the covariance between a gene's transcript levels at two time points (Figure 7B). Measurements of a gene's expression that are closer in time will share a greater covariance (be more correlated) than those further apart in time. The timescale reflects how rapidly the covariance decays when considering expression at increasing time intervals (Figure 7B). Genes with rapidly changing expression kinetics lose covariance quickly and have short timescales (Figures 7B and 7C, left). Conversely, genes with expression kinetics that vary more smoothly and gradually have covariances that persist in time and are described by long timescales (Figures 7B and 7C, right).

To explore the characteristic timescales relevant to development, we plotted the distribution of timescales over all genes (Figure 7C). We found the mass of the timescale distribution to be slower than our detection limit (Supplemental Experimental Procedures); in fact, our sampling rate is sufficient to reliably detect very short timescales (for example, *pri-mir427* with τ = 1.8 hr) (Figure 7C, upper left panel). We investigated the timescale distribution by GO (Figure 7C). Remarkably, GO term enrichments segregated between short and long timescales (Table S4). At short timescales, there was a striking enrichment of genes associated with cell signaling and developmental processes; these genes are involved in rapid, spatially restricted patterning events during embryogenesis. At long timescales, we found enrichment in genes necessary for cellular metabolism, many of which correlate in expression with the changes in total mRNA in the developing embryo. We conclude that timescales are a powerful metric for investigating gene function and that they are able to classify novel genes whose kinetics are similar to those with developmental or cell signaling roles or unidentified roles in cellular metabolism.

DISCUSSION

We present a rich array of findings on transcript kinetics from our ultra-high temporal resolution expression profiling of *Xenopus*. To establish these biological observations, we overcame the challenges of direct absolute normalization of RNA-seq data using ERCC spikes (SEQC/MAQC-III Consortium, 2014). In this study, we show the following: (1) It is essential to spike RNA standards at a constant ratio to embryo/cell numbers directly into homogenates *prior* to RNA extraction. (2) Relative normalized spikes exhibit a clear decreasing trend that was the result of increasing RNA in the embryo with time. This demonstrated that the ERCC spikes were of suitable fidelity for absolute normalization and identified our absolute normalization strategy. (3) It is necessary to sequence the same RNA by multiple protocols—in this case, poly(A)+ and rdRNA—so that protocol-specific biases can be understood and minimized. (4) It is essential to estimate the error, or uncertainty, in the measurement of the RNA standards (Figures S2B–S2D) and then average over this uncertainty when making absolute transcript number predictions. We have achieved this latter point using Gaussian processes, and in doing so, we have ensured that our absolute normalization of native transcripts is not unduly influenced by technical noise or sequencing biases of the RNA standards (Figure S2D). (5) Finally, we note that high-resolution temporal sampling has a significant advantage over multiple replicates of a low temporal sampling. With high-frequency sampling, neighboring time points are still as similar as biological replicates, providing the necessary information on reproducibility, but, importantly, we also capture the transcript dynamics and so obtain more information for the same number of samples.

We have demonstrated that direct absolute normalization of RNA-seq data is feasible; the quality of the RNA standards and the understanding of their sequencing biases are the largest factors limiting progress in the absolute quantification of gene expression data and the accuracy of our predictions (see uncertainty CIs in Figure S2D). Currently, the ERCC spikes are the predominant RNA standards; they are manufactured to log-scale precision and exhibit significant sequencing biases. With log-scale precision, the error is proportional to abundance and

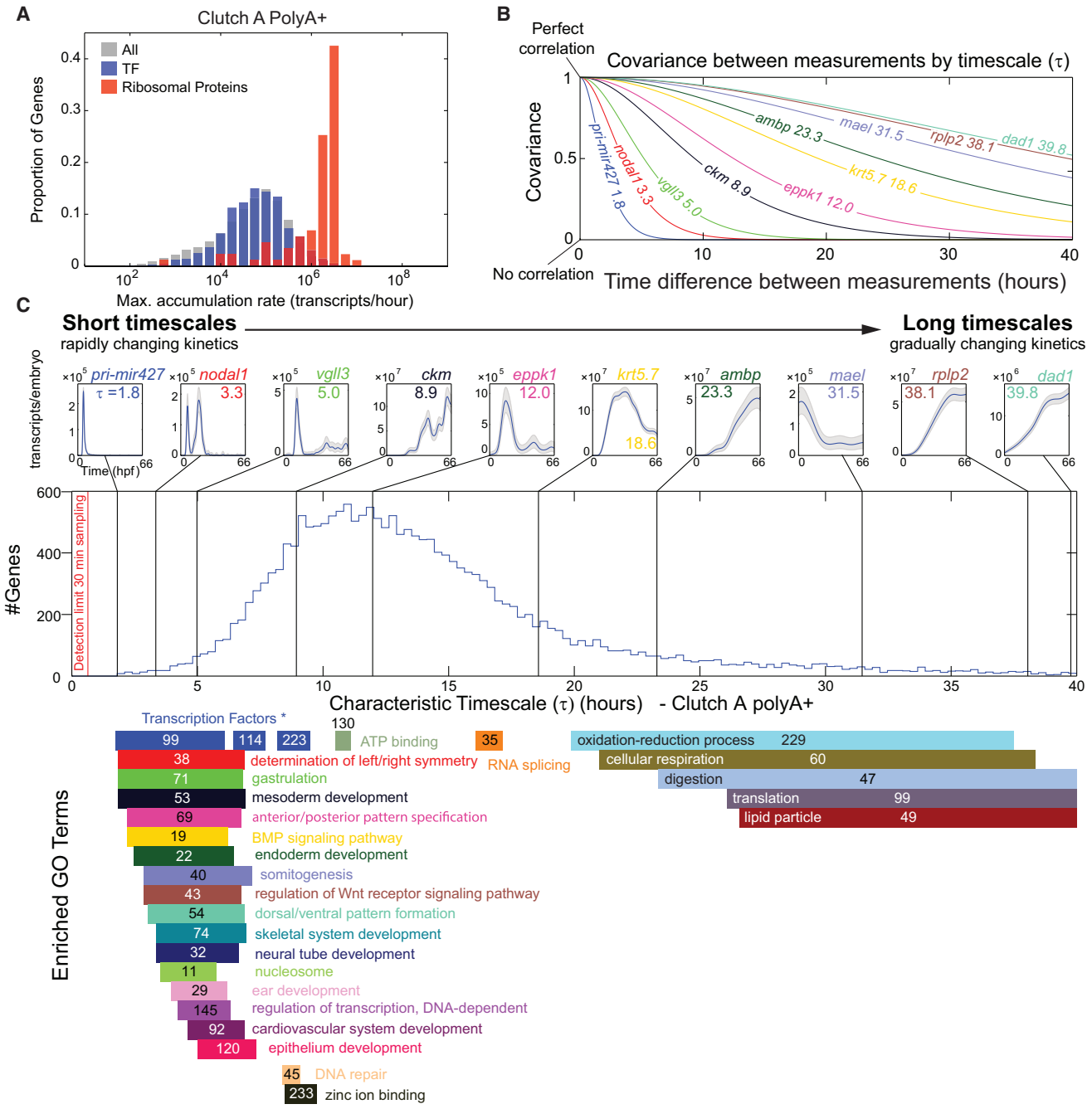


Figure 7. Characteristic Timescales of Gene Expression in Clutch A Poly(A)+

(A) Distribution of maximum rate of increase of all transcripts (gray) (as in Figure 5), transcription factors (blue), and ribosomal proteins *rps** and *rpl** (red).

(B) Theoretical covariance (Matérn covariance function; Supplemental Experimental Procedures) between two measurements by time interval for example genes in (C).

(C) Histogram of characteristic timescale over all genes in in Clutch A poly(A)+. Top row annotates example genes (timescale inset); fast/short timescales are on the left, and slow/long timescales on the right. Gaussian process 95% CI (gray) and median (blue) are marked. Bars below indicate GO enrichment (calculated as in Figure 2; Figure S6A) of histogram, and numbers indicate total genes with given category. All enrichments have $p < 2.6 \times 10^{-5}$ (Table S4). "Transcription Factors*" labels an annotation of transcription factors separate to GO (Supplemental Experimental Procedures).

See also Figure S6A, Table S4, and Supplemental Experimental Procedures.

doubles for every doubling of target transcript copy numbers. Clearly, transcripts are under much tighter regulation in the embryo, and the log-scale precision in RNA standard concentrations is far from ideal. The need for improved RNA quantification standards to advance future absolute quantification efforts is clear.

We demonstrate that high-frequency sampling is critical to define smooth expression trajectories and properly examine transcript dynamics. With such data, we can identify cohorts of temporally co-regulated genes and make accurate predictions on spatial expression and gene function. Our case studies of somite and vision temporal synexpression (Figure 4B) demonstrate the potential of this dataset not only for characterizing genes of unknown function but also for uncovering new temporal phenomena in gene expression. High-frequency sampling combined with Gaussian process analysis is a powerful tool for capturing the temporal evolution of differential expression, including identifying convergence and divergence points. Importantly, our analysis of timescales confirmed that our sampling rate was sufficient to capture the transcript behavior for even the most dynamic genes. The concept of the characteristic timescale will be critical for future transcriptome time series studies to ensure that dynamic behaviors are not missed or misinterpreted.

We also analyzed the global properties and kinetics of the transcriptome during our 66-hr time course. Interestingly, we found that much of the transcriptome accumulates to ~600,000 transcripts per embryo at different developmental stages. Examinations of transcript kinetics revealed that different kinetic signatures can be associated with the accumulation of transcripts via polyadenylation or transcription. We expect that further examination of kinetics may reveal additional signatures associated with different mechanisms of mRNA regulation.

The maternal-to-zygotic transition constitutes the most dramatic change in the transcriptome that we observe in our time course. It begins with the polyadenylation of maternal transcripts, followed by the onset of zygotic transcription and the clearance of maternal transcripts. The absolute nature of our data provides new quantitative insights into this period. For example, zygotic transcription starts early during the cleavage stages with the activation of *pri-mir427*, which plays an important role in maternal RNA clearance. The reduction in maternal mRNA exceeds transcription as the cleavage stages are completed (blastula stage 9; 4.5 hpf) and mRNA levels fall until the end of gastrulation (stage 12.5; 10 hpf). During this time window, the majority of transcript-level differences between Clutch A and Clutch B converge (Figure 2D), and many zygotic isoforms are activated whose abundance will exceed their maternal counterpart by ~10 hpf (Figure 3A).

Our data demonstrate that transcript levels during embryonic development are exquisitely controlled, with only 2% of transcripts showing large differences in expression between replicates. Although the biological importance of these differences is not certain, these 2% of genes are, nevertheless, tightly regulated within the embryos of each clutch. This differential expression appears to be interesting and worthy of further exploration. Tight control over gene expression occurs at multiple levels, including chromatin modifications, *cis*-regulatory elements, transcription factors, and transcript stability. Our experimental approach can be used to investigate the effects of perturbation of these regulatory processes on transcript kinetics, which is particularly difficult without measurements of absolute transcript levels. Such real measurements have the potential to transform our understanding of the transcriptional control that is fundamental to gene regulatory networks. Our approach affords many opportunities to build gene regulatory networks whose foundation lies on much-needed quantitative data. The potential for using Gaussian processes with time series expression data for uncovering gene regulatory interactions has been demonstrated (Honkela et al., 2010). Combining these approaches with our time course data holds exciting prospects for the elucidation of gene regulatory networks in early vertebrate development.

A major goal in developmental biology is to determine the temporal and spatial expression patterns of all genes during development. Our work is a step toward that goal. By combining absolute normalization and high-frequency sampling with the recent advances in RNA tomography (Junker et al., 2014) and/or the spatial reconstruction of single-cell RNA-seq data (Satija et al., 2015), there is an exciting opportunity to create a four-dimensional (4D) atlas of developmental gene expression, especially in vertebrate human disease models such as *Xenopus*.

EXPERIMENTAL PROCEDURES

Embryo Collection

X. tropicalis were housed and cared for in our aquatics facility according to established protocols that were approved by the Yale Institutional Animal Care and Use Committee (IACUC) (Khokha et al., 2002). To obtain RNA samples for RNA-seq, we performed two parallel in vitro fertilizations of siblings of 12th-generation inbred Nigerian *Xenopus tropicalis*. All fertilizations and subsequent culturing of embryos were performed in a temperature-controlled room maintained at ~24°C, with fluctuations ±1°C over the entire time course. All samples are reported in hpf and developmental stages. Mapping from hpf to the Nieuwkoop and Faber (NF) stage is given in Table S1. We labeled the progeny of the two crosses Clutch A and Clutch B. All collections from the two clutches occurred concurrently.

RNA Isolation and Spike-Ins

For most samples, ten embryos per time point were homogenized in 200 μl Trizol and frozen at −80°C, with the exception of egg, 0.5, 1, 1.5, 2, 2.5, and 3 hpf time points, which were sampled at 25, 30, 20, 20, 20, 15, and 15 embryos, respectively. ERCC RNA Spike-In Mix 1 (Thermo Fisher Scientific) was added on a per-embryo basis to Clutch A samples. For the Clutch B time course, ERCC RNA Spike-In Mix 1 was also added, along with the independent adding of Ambion ArrayControl RNA Spikes (Thermo Fisher Scientific). See Supplemental Experimental Procedures for additional details.

Data Analysis

All libraries were sequenced with 76-bp paired-ends on an Illumina HiSeq 2000. Read pairs were aligned using TopHat v2.0.10 (Trapnell et al., 2009) to the *X. tropicalis* version 7.1 genome (Hellsten et al., 2010; Karpinka et al., 2015), along with our ERCC and ArrayControl exogenous spike sequences and known off-assembly gene sequences used in a previous study (Collart et al., 2014). Alignment was guided by the v7.2 *X. tropicalis* gene models, to which we made improvements and corrections (Data S1). We estimated the relative abundance of all transcripts in fragments per kilobase per million (FPKM) with Cufflinks v2.1.1 (Trapnell et al., 2010) and converted to TPM. Absolute normalization was achieved by calibrating relative normalized transcript abundances against spikes to achieve absolute normalization. Corrections for poly(A)+ bias and Clutch A/B variation were applied to absolute normalization of native transcripts. All subsequent analysis used Gaussian processes. See Supplemental Experimental Procedures for details.

ACCESSION NUMBERS

The accession number for the data reported in this paper is GEO: GSE65785. Gene expression profiles may be visualized at http://genomics.crick.ac.uk/apps/profiles/.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, seven tables, and two data files and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2015.12.050.

AUTHOR CONTRIBUTIONS

I.L.B. and M.K.K. collected the embryos for the time course, and I.L.B. prepared all RNAs. J.D.O. supervised and provided critical expertise on RNA-seq at the Yale Center for Genome Analysis. N.D.L.O. performed all of the bioinformatics analysis. M.A.L. did all of the in situ hybridizations. I.P. built the gene expression visualization website. N.D.L.O., I.L.B., M.J.G., K.W.Y.C., and M.K.K. conceived the work, designed the experiments, and interpreted all experimental data. All authors contributed to the writing of the manuscript.

ACKNOWLEDGMENTS

We thank Sarah Kubek and Michael Slocum for animal husbandry. We are indebted to the support and sequencing at the Yale Center for Genome Analysis, especially the assistance of Shrikant Mane. The authors also thank Joe Gall and Margaret Fish for stimulating discussions. The authors thank James Noonan, Xiaohui Xie, Sarah Teichmann, and George Gentsch and the members of the M.J.G., M.K.K., and K.W.Y.C. labs for helpful discussion of the manuscript. This work was supported by NIH/NIGMS grant 5R01GM099149 to M.J.G., K.W.Y.C., and M.K.K. M.K.K. is a Mallinckrodt Scholar. This work was supported in part by the Francis Crick Institute, which receives its core funding from Cancer Research UK, the UK Medical Research Council, and the Wellcome Trust.

Received: June 25, 2015
Revised: November 2, 2015
Accepted: December 7, 2015
Published: January 7, 2016

REFERENCES

Aanes, H., Winata, C.L., Lin, C.H., Chen, J.P., Srinivasan, K.G., Lee, S.G., Lim, A.Y., Hajan, H.S., Collas, P., Bourque, G., et al. (2011). Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. Genome Res. *21*, 1328–1338.

Äijö, T., Butty, V., Chen, Z., Salo, V., Tripathi, S., Burge, C.B., Lahesmaa, R., and Lähdesmäki, H. (2014). Methods for time series analysis of RNA-seq data with application to human Th17 cell differentiation. Bioinformatics *30*, i113–i120.

Ardehali, M.B., and Lis, J.T. (2009). Tracking rates of transcription and splicing in vivo. Nat. Struct. Mol. Biol. *16*, 1123–1124.

Baker, S.C., Bauer, S.R., Beyer, R.P., Brenton, J.D., Bromley, B., Burrill, J., Causton, H., Conley, M.P., Elespuru, R., Fero, M., et al.; External RNA Controls Consortium (2005). The External RNA Controls Consortium: a progress report. Nat. Methods *2*, 731–734.

Bowes, J.B., Snyder, K.A., Segerdell, E., Jarabek, C.J., Azam, K., Zorn, A.M., and Vize, P.D. (2010). Xenbase: gene expression and improved integration. Nucleic Acids Res. *38*, D607–D612.

Brennecke, P., Anders, S., Kim, J.K., Kołodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C., and Heisler, M.G. (2013). Accounting for technical noise in single-cell RNA-seq experiments. Nat. Methods *10*, 1093–1095.

Collart, C., Owens, N.D., Bhaw-Rosun, L., Cooper, B., De Domenico, E., Patrushev, I., Sesay, A.K., Smith, J.N., Smith, J.C., and Gilchrist, M.J. (2014). High-resolution analysis of gene activity during the Xenopus mid-blastula transition. Development *141*, 1927–1939.

Crowder, M.E., Strzelecka, M., Wilbur, J.D., Good, M.C., von Dassow, G., and Heald, R. (2015). A comparative analysis of spindle morphometrics across metazoans. Curr. Biol. *25*, 1542–1550.

Curran, K.L., LaRue, S., Bronson, B., Solis, J., Trow, A., Sarver, N., and Zhu, H. (2008). Circadian genes are expressed during early development in Xenopus laevis. PLoS ONE *3*, e2749.

Curran, K.L., Allen, L., Porter, B.B., Dodge, J., Lope, C., Willadsen, G., Fisher, R., Johnson, N., Campbell, E., VonBergen, B., et al. (2014). Circadian genes, xBmal1 and xNocturnin, modulate the timing and differentiation of somites in Xenopus laevis. PLoS ONE *9*, e108266.

Davidson, E.H. (1986). Gene Activity in Early Development, Third Edition (Orlando, FL: Academic Press).

Fakhro, K.A., Choi, M., Ware, S.M., Belmont, J.W., Towbin, J.A., Lifton, R.P., Khokha, M.K., and Brueckner, M. (2011). Rare copy number variations in congenital heart disease patients identify unique genes in left-right patterning. Proc. Natl. Acad. Sci. USA *108*, 2915–2920.

Fang, H., Yang, Y., Li, C., Fu, S., Yang, Z., Jin, G., Wang, K., Zhang, J., and Jin, Y. (2010). Transcriptome analysis of early organogenesis in human embryos. Dev. Cell *19*, 174–184.

Giraldez, A.J., Cinalli, R.M., Glasner, M.E., Enright, A.J., Thomson, J.M., Baskerville, S., Hammond, S.M., Bartel, D.P., and Schier, A.F. (2005). MicroRNAs regulate brain morphogenesis in zebrafish. Science *308*, 833–838.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75–79.

Heinonen, M., Guipaud, O., Milliat, F., Buard, V., Micheau, B., Tarlet, G., Benderitter, M., Zehraoui, F., and d'Alche-Buc, F. (2014). Detecting time periods of differential gene expression using Gaussian processes: an application to endothelial cells exposed to radiotherapy dose fraction. Bioinformatics *31*, 728–735.

Hellsten, U., Harland, R.M., Gilchrist, M.J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N.H., Shu, S., Taher, L., et al. (2010). The genome of the Western clawed frog Xenopus tropicalis. Science *328*, 633–636.

Honkela, A., Girardot, C., Gustafson, E.H., Liu, Y.H., Furlong, E.E., Lawrence, N.D., and Rattray, M. (2010). Model-based method for transcription factor target identification with limited data. Proc. Natl. Acad. Sci. USA *107*, 7793–7798.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. Nat. Methods *11*, 163–166.

Jolliffe, I.T. (2002). Principal Component Analysis, Second Edition (New York, NY: Springer).

Junker, J.P., Noël, E.S., Guryev, V., Peterson, K.A., Shah, G., Huisken, J., McMahon, A.P., Berezikov, E., Bakkers, J., and van Oudenaarden, A. (2014). Genome-wide RNA Tomography in the zebrafish embryo. Cell *159*, 662–675.

Kalinka, A.T., Varga, K.M., Gerrard, D.T., Preibisch, S., Corcoran, D.L., Jarrells, J., Ohler, U., Bergman, C.M., and Tomancak, P. (2010). Gene expression divergence recapitulates the developmental hourglass model. Nature *468*, 811–814.

Karlebach, G., and Shamir, R. (2008). Modelling and analysis of gene regulatory networks. Nat. Rev. Mol. Cell Biol. *9*, 770–780.

Karpinka, J.B., Fortriede, J.D., Burns, K.A., James-Zorn, C., Ponferrada, V.G., Lee, J., Karimi, K., Zorn, A.M., and Vize, P.D. (2015). Xenbase, the Xenopus model organism database; new virtualized system, data types and genomes. Nucleic Acids Res. *43*, D756–D763.

Khokha, M.K., Chung, C., Bustamante, E.L., Gaw, L.W., Trott, K.A., Yeh, J., Lim, N., Lin, J.C., Taverner, N., Amaya, E., Papalopulu, N., Smith, J.C., Zorn, A.M., Harland, R.M., and Grammer, T.C. (2002). Techniques and probes for the study of Xenopus tropicalis development. Dev Dyn. *225*, 499–510.

Kimelman, D., Kirschner, M., and Scherson, T. (1987). The events of the midblastula transition in Xenopus are regulated by changes in the cell cycle. Cell *48*, 399–407.

Klein, S.L., Strausberg, R.L., Wagner, L., Pontius, J., Clifton, S.W., and Richardson, P. (2002). Genetic and genomic tools for Xenopus research: The NIH Xenopus initiative. Dev. Dyn. *225*, 384–391.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Lund, E., Liu, M., Hartley, R.S., Sheets, M.D., and Dahlberg, J.E. (2009). Deadenylation of maternal mRNAs mediated by miR-427 in Xenopus laevis embryos. RNA *15*, 2351–2363.

Marguerat, S., Schmidt, A., Codlin, S., Chen, W., Aebersold, R., and Bähler, J. (2012). Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. Cell *151*, 671–683.

Miller, O.L., Jr., and Hamkalo, B.A. (1972). Visualization of RNA synthesis on chromosomes. Int. Rev. Cytol. *33*, 1–25.

Morin, R.D., Chang, E., Petrescu, A., Liao, N., Griffith, M., Chow, W., Kirkpatrick, R., Butterfield, Y.S., Young, A.C., Stott, J., et al. (2006). Sequencing and analysis of 10,967 full-length cDNA clones from Xenopus laevis and Xenopus tropicalis reveals post-tetraploidization transcriptome remodeling. Genome Res. *16*, 796–803.

Newport, J., and Kirschner, M. (1982). A major developmental transition in early Xenopus embryos: II. Control of the onset of transcription. Cell *30*, 687–696.

Niehrs, C., and Pollet, N. (1999). Synexpression groups in eukaryotes. Nature *402*, 483–487.

Paranjpe, S.S., Jacobi, U.G., van Heeringen, S.J., and Veenstra, G.J. (2013). A genome-wide survey of maternal and embryonic transcripts during Xenopus tropicalis development. BMC Genomics *14*, 762.

Pourquié, O. (2003). Vertebrate somitogenesis: a novel paradigm for animal segmentation? Int. J. Dev. Biol. *47*, 597–603.

Qing, T., Yu, Y., Du, T., and Shi, L. (2013). mRNA enrichment protocols determine the quantification characteristics of external RNA spike-in controls in RNA-Seq studies. Sci. China Life Sci. *56*, 134–142.

Rasmussen, C.E., and Williams, C.K.I. (2006). Gaussian Processes for Machine Learning (Cambridge, MA: MIT Press).

Risso, D., Ngai, J., Speed, T.P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. Nat. Biotechnol. *32*, 896–902.

Sagata, N., Shiokawa, K., and Yamana, K. (1980). A study on the steady-state population of poly(A)+RNA during early development of Xenopus laevis. Dev. Biol. *77*, 431–448.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.

SEQC/MAQC-III Consortium (2014). A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nat. Biotechnol. *32*, 903–914.

Skirkanich, J., Luxardi, G., Yang, J., Kodjabachian, L., and Klein, P.S. (2011). An essential role for transcription before the MBT in Xenopus laevis. Dev. Biol. *357*, 478–491.

Spencer, W.C., Zeller, G., Watson, J.D., Henz, S.R., Watkins, K.L., McWhirter, R.D., Petersen, S., Sreedharan, V.T., Widmer, C., Jo, J., et al. (2011). A spatial and temporal map of C. elegans gene expression. Genome Res. *21*, 325–341.

Stegle, O., Denby, K.J., Cooke, E.J., Wild, D.L., Ghahramani, Z., and Borgwardt, K.M. (2010). A robust Bayesian two-sample test for detecting intervals of differential gene expression in microarray time series. J. Comput. Biol. *17*, 355–367.

Stegle, O., Teichmann, S.A., and Marioni, J.C. (2015). Computational and analytical challenges in single-cell transcriptomics. Nat. Rev. Genet. *16*, 133–145.

Tan, M.H., Au, K.F., Yablonovitch, A.L., Wills, A.E., Chuang, J., Baker, J.C., Wong, W.H., and Li, J.B. (2013). RNA sequencing reveals a diverse and dynamic repertoire of the Xenopus tropicalis transcriptome over development. Genome Res. *23*, 201–216.

Tebbenkamp, A.T., Willsey, A.J., State, M.W., and Sestan, N. (2014). The developmental transcriptome of the human brain: implications for neurodevelopmental disorders. Curr. Opin. Neurol. *27*, 149–156.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S.E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S.E., et al. (2002). Systematic determination of patterns of gene expression during Drosophila embryogenesis. Genome Biol. *3*, RESEARCH0088.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Tu, Q., Cameron, R.A., and Davidson, E.H. (2014). Quantitative developmental transcriptomes of the sea urchin Strongylocentrotus purpuratus. Dev. Biol. *385*, 160–167.

van Heeringen, S.J., Akkers, R.C., van Kruijsbergen, I., Arif, M.A., Hanssen, L.L., Sharifi, N., and Veenstra, G.J. (2014). Principles of nucleation of H3K27 methylation during embryonic development. Genome Res. *24*, 401–410.

Vesterlund, L., Jiao, H., Unneberg, P., Hovatta, O., and Kere, J. (2011). The zebrafish transcriptome during early development. BMC Dev. Biol. *11*, 30.

wa Maina, C., Honkela, A., Matarese, F., Grote, K., Stunnenberg, H.G., Reid, G., Lawrence, N.D., and Rattray, M. (2014). Inference of RNA polymerase II transcription dynamics from chromatin immunoprecipitation time course data. PLoS Comput. Biol. *10*, e1003598.

Wang, Q.T., Piotrowska, K., Ciemerych, M.A., Milenkovic, L., Scott, M.P., Davis, R.W., and Zernicka-Goetz, M. (2004). A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. Dev. Cell *6*, 133–144.

Yanai, I., Peshkin, L., Jorgensen, P., and Kirschner, M.W. (2011). Mapping gene expression in two Xenopus species: evolutionary constraints and developmental flexibility. Dev. Cell *20*, 483–496.

Yang, J., Tan, C., Darken, R.S., Wilson, P.A., and Klein, P.S. (2002). Beta-catenin/Tcf-regulated transcription prior to the midblastula transition. Development *129*, 5743–5752.

Zaidi, S., Choi, M., Wakimoto, H., Ma, L., Jiang, J., Overton, J.D., Romano-Adesman, A., Bjornson, R.D., Breitbart, R.E., Brown, K.K., et al. (2013). De novo mutations in histone-modifying genes in congenital heart disease. Nature *498*, 220–223.