Fig S1

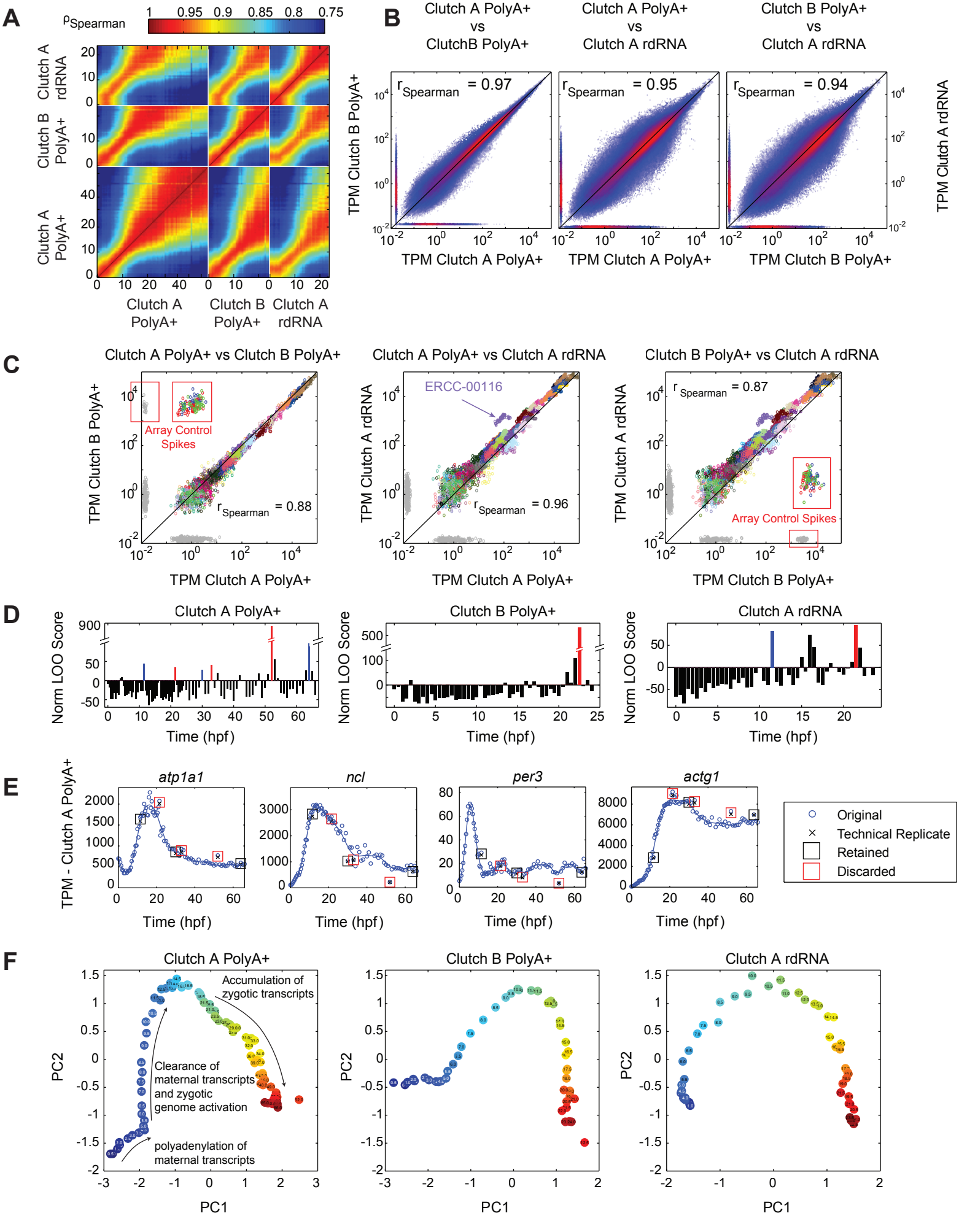
A
ρSpearman
1 0.95 0.9 0.85 0.8 0.75
Clutch A rdRNA
Clutch B PolyA+
Clutch A PolyA+
Clutch A PolyA+
Clutch B PolyA+
Clutch A rdRNA
B
Clutch A PolyA+ vs ClutchB PolyA+
Clutch A PolyA+ vs Clutch A rdRNA
Clutch B PolyA+ vs Clutch A rdRNA
rSpearman = 0.97
rSpearman = 0.95
rSpearman = 0.94
TPM Clutch B PolyA+
TPM Clutch A PolyA+
TPM Clutch A PolyA+
TPM Clutch B PolyA+
TPM Clutch A rdRNA
C
Clutch A PolyA+ vs Clutch B PolyA+
Clutch A PolyA+ vs Clutch A rdRNA
Clutch B PolyA+ vs Clutch A rdRNA
Array Control Spikes
rSpearman = 0.88
ERCC-00116
rSpearman = 0.96
rSpearman = 0.87
Array Control Spikes
TPM Clutch B PolyA+
TPM Clutch A PolyA+
TPM Clutch A rdRNA
TPM Clutch A PolyA+
TPM Clutch A rdRNA
TPM Clutch B PolyA+
D
Clutch A PolyA+
Clutch B PolyA+
Clutch A rdRNA
Norm LOO Score
Time (hpf)
E
atp1a1
ncl
per3
actg1
TPM - Clutch A PolyA+
Time (hpf)
Original
Technical Replicate
Retained
Discarded
F
Clutch A PolyA+
Clutch B PolyA+
Clutch A rdRNA
Accumulation of zygotic transcripts
Clearance of maternal transcripts and zygotic genome activation
polyadenylation of maternal transcripts
PC1
PC2

Fig S1 – RNA-seq data quality, related to Fig 1, Supplemental Experimental Procedures

(A) Spearman correlation between all pairs of all samples for relative normalized TPM abundances. Neighboring samples are as correlated as replicates.

(B) Scatter plots of all TPM abundances for all samples. Color indicates density of points. Horizontal and vertical strips indicate measurements only present in one sample.

(C) RNA standard performance. Scatter plots compare RNA standard TPM abundances between all pairs of datasets. Color indicates spike-in species. Red boxes mark ArrayControl spikes present only in Clutch B PolyA+. Although, we find small numbers of reads that align uniquely to ArrayControl sequences when ArrayControl spikes are not present in the sample. We note that this is not true for ERCC spikes, we never find ERCC reads when ERCC spikes are not present (data not shown). ERCC PolyA+ and rdRNA comparisons lie off the diagonal indicating poor PolyA+ performance. ERCC-00116 has the worst PolyA+ performance and is not used in absolute normalization.

(D) Identification of poorly performing libraries. Normalized leave-one-out score (Supplemental Experimental Procedures), higher scores indicate a poorly performing sample. Samples adjacent to a bad sample can also receive high scores. Red and blue samples in Clutch A PolyA+ are selected for repeat library construction and sequencing (E). Red samples excluded from all analysis.

(E) Technical replicates in Clutch A PolyA+ of poorly performing samples reveal that the poor performance is highly reproducible indicating that performance issues are not due to library construction and sequencing. Red boxes marked those discarded, black boxes marked those retained.

(F) Projections onto the first two principle components for Clutch A PolyA+ (left), Clutch B PolyA+ (center), and Clutch A rdRNA (right). Note strong temporal correlation between samples, major changes in the transcriptome are visible.

Fig S2

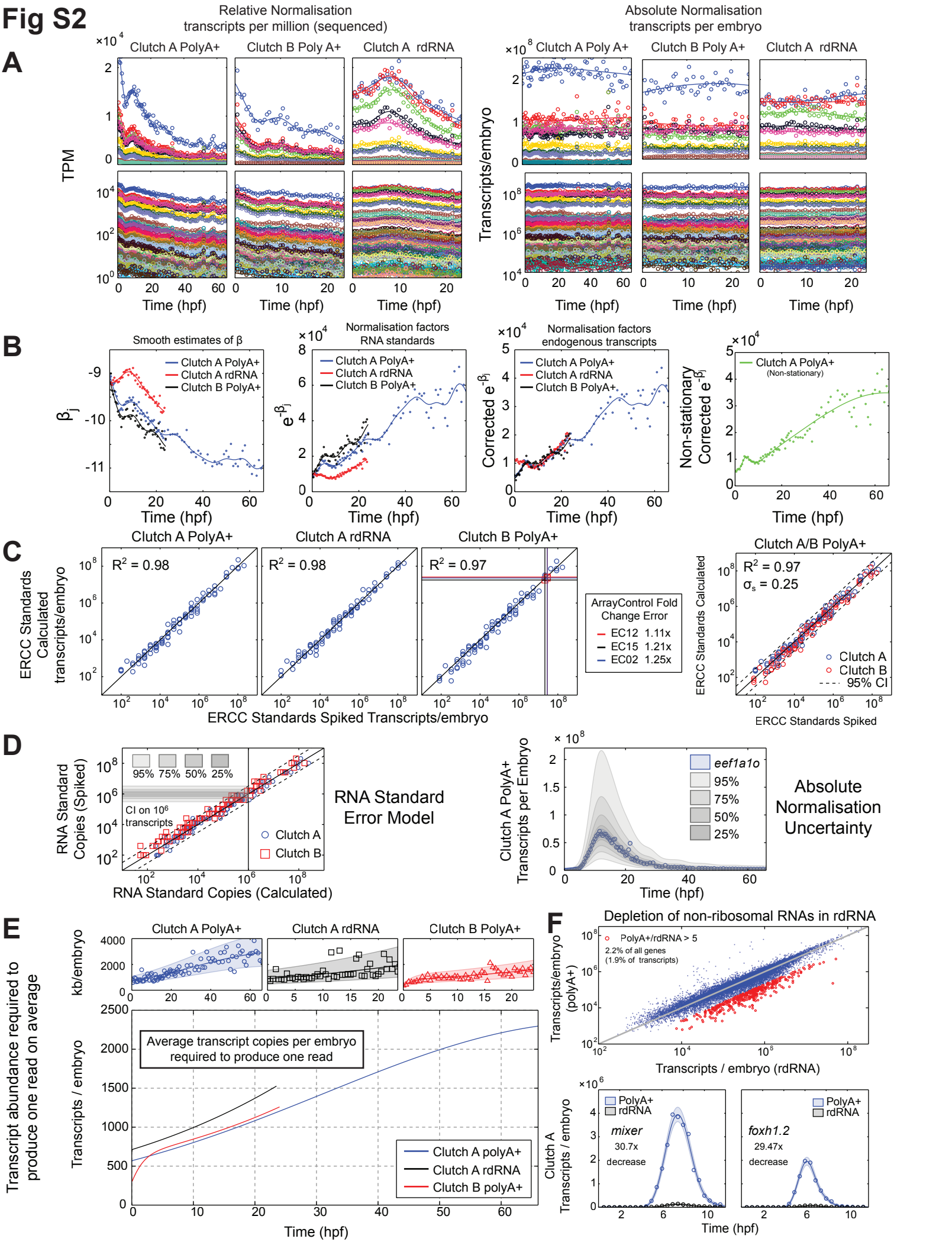
A
Relative Normalisation
transcripts per million (sequenced)
Absolute Normalisation
transcripts per embryo
Clutch A PolyA+
Clutch B Poly A+
Clutch A rdRNA
TPM
Transcripts/embryo
Time (hpf)
B
Smooth estimates of β
Normalisation factors
RNA standards
Normalisation factors
endogenous transcripts
Clutch A PolyA+
Clutch A rdRNA
Clutch B PolyA+
Corrected e^-βj
Non-stationary
Corrected e^-βj
Clutch A PolyA+
(Non-stationary)
C
Clutch A PolyA+
Clutch A rdRNA
Clutch B PolyA+
R² = 0.98
R² = 0.98
R² = 0.97
ERCC Standards
Calculated
transcripts/embryo
ERCC Standards Spiked Transcripts/embryo
ArrayControl Fold
Change Error
EC12 1.11x
EC15 1.21x
EC02 1.25x
Clutch A/B PolyA+
R² = 0.97
σs = 0.25
ERCC Standards Calculated
ERCC Standards Spiked
Clutch A
Clutch B
95% CI
D
95% 75% 50% 25%
CI on 10⁶
transcripts
RNA Standard
Copies (Spiked)
RNA Standard Copies (Calculated)
Clutch A
Clutch B
RNA Standard
Error Model
Clutch A PolyA+
Transcripts per Embryo
eef1a1o
95%
75%
50%
25%
Absolute
Normalisation
Uncertainty
E
Transcript abundance required to
produce one read on average
Clutch A PolyA+
Clutch A rdRNA
Clutch B PolyA+
kb/embryo
Transcripts / embryo
Average transcript copies per embryo
required to produce one read
Clutch A polyA+
Clutch A rdRNA
Clutch B polyA+
Time (hpf)
F
Depletion of non-ribosomal RNAs in rdRNA
PolyA+/rdRNA > 5
2.2% of all genes
(1.9% of transcripts)
Transcripts/embryo
(polyA+)
Transcripts / embryo (rdRNA)
Clutch A
Transcripts / embryo
PolyA+
rdRNA
mixer
30.7x
decrease
foxh1.2
29.47x
decrease
Time (hpf)

Fig S2 – Statistics of Absolute Normalization, related to Fig 1, Supplemental Experimental Procedures

(A) Absolute normalization of RNA standards. *Left:* relative normalization in TPM on linear and log scale for the three datasets. The trend shared by all spikes is clearly visible and differs between PolyA+ and rdRNA samples. *Right:* absolute normalization in transcripts per embryo on log and linear scale. Trend in spikes is removed, the sample noise of the spikes is retained and does not contaminate the absolute normalization of native transcripts.

(B) Absolute normalization factors. Circles mark per sample factors, lines mark Gaussian Process smoothed factors. *Left:* Estimates of β_j (Supplemental Experimental Procedures), for three datasets. *Center-left:* per dataset correction factors $\mathrm{e}^{-\beta_j}$, note discrepancy in magnitude but not in trend between datasets. *Center-right:* correction for PolyA+ bias, and averaging Clutch A and Clutch B spike levels results in excellent agreement between the correction factors. *Right:* Smoothed correction factors calculated with non-stationary Gaussian Process (Supplemental Experimental Procedures).

(C) Absolute normalization consistency and accuracy. Scatter plots of ERCC standards spiked vs calculated for Clutch A PolyA+, Clutch A rdRNA, Clutch B PolyA+ and Clutch A/Clutch B PolyA+ combined. Points are averaged over all samples in time to remove spike-in sample noise. ArrayControl spikes are given with fold change errors for Clutch B PolyA+, fold changes range from 1.11-1.25. ArrayControl spikes EC02 and EC12 have are spiked in with the same copy number and are marked with the same vertical line.Combined Clutch A/Clutch B gives 95% confidence interval for a linear model with Gaussan noise, with slope unity and with standard deviation $\sigma_s = 0.25$. Note that residuals do not depend on expression level.

(D) Absolute normalization uncertainty model. *Left:* model as described in **C** (*far-right*), with 25%, 50%, 75%, 95% confidence intervals for 10^6 transcripts. *Right:* Propagation of uncertainty model to absolute normalization of *eef1a1o* Clutch A polyA+, again 25%, 50% 75% and 95% of confidence intervals for the true number of *eef1a1o* transcripts/embryo marked.

(E) RNA-seq detection limits **–** Transcript abundance required to produce a single read with time, detection limits increase as RNA content in the embryo increases. *Top row*: Detection limits in kb, points give the per sample detection limits, lines and shaded area give Gaussian Process median and 95% confidence interval. *Bottom*: Detection limit in transcripts/embryo, calculated by averaging kb Gaussian Process distribution over all transcript lengths.

(F) The depletion of non-ribosomal RNAs in rdRNA sequencing. Comparisons between Clutch A PolyA+ and rdRNA sequencing (*top)*. Examples of extreme depletion: *mixer* and *foxh1.2*.

Fig S3

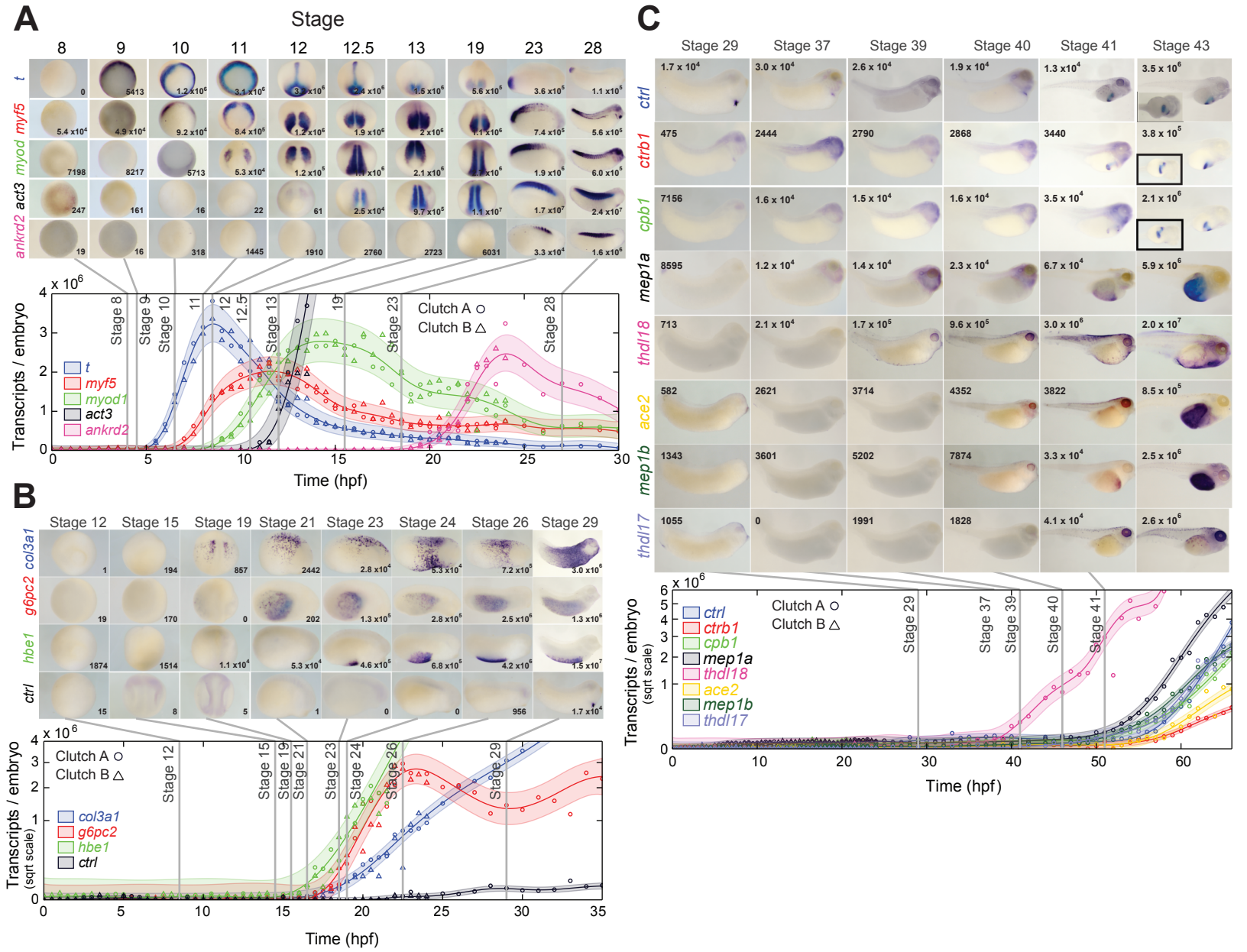

Fig S3 – Spatial expression of selected genes by in situ, related to Fig 1

All panels give Clutch A PolyA+ (circles), Clutch B PolyA+ (triangles) and Gaussian Process median line and 95% confidence intervals. Numbers on in situ images indicate the Gaussian process median number of transcripts for the given stage.

(A) Genes involved in muscle induction.

(B) Genes with onset midway through the timecourse.

(C) Genes with onset late in the timecourse.

Fig S4

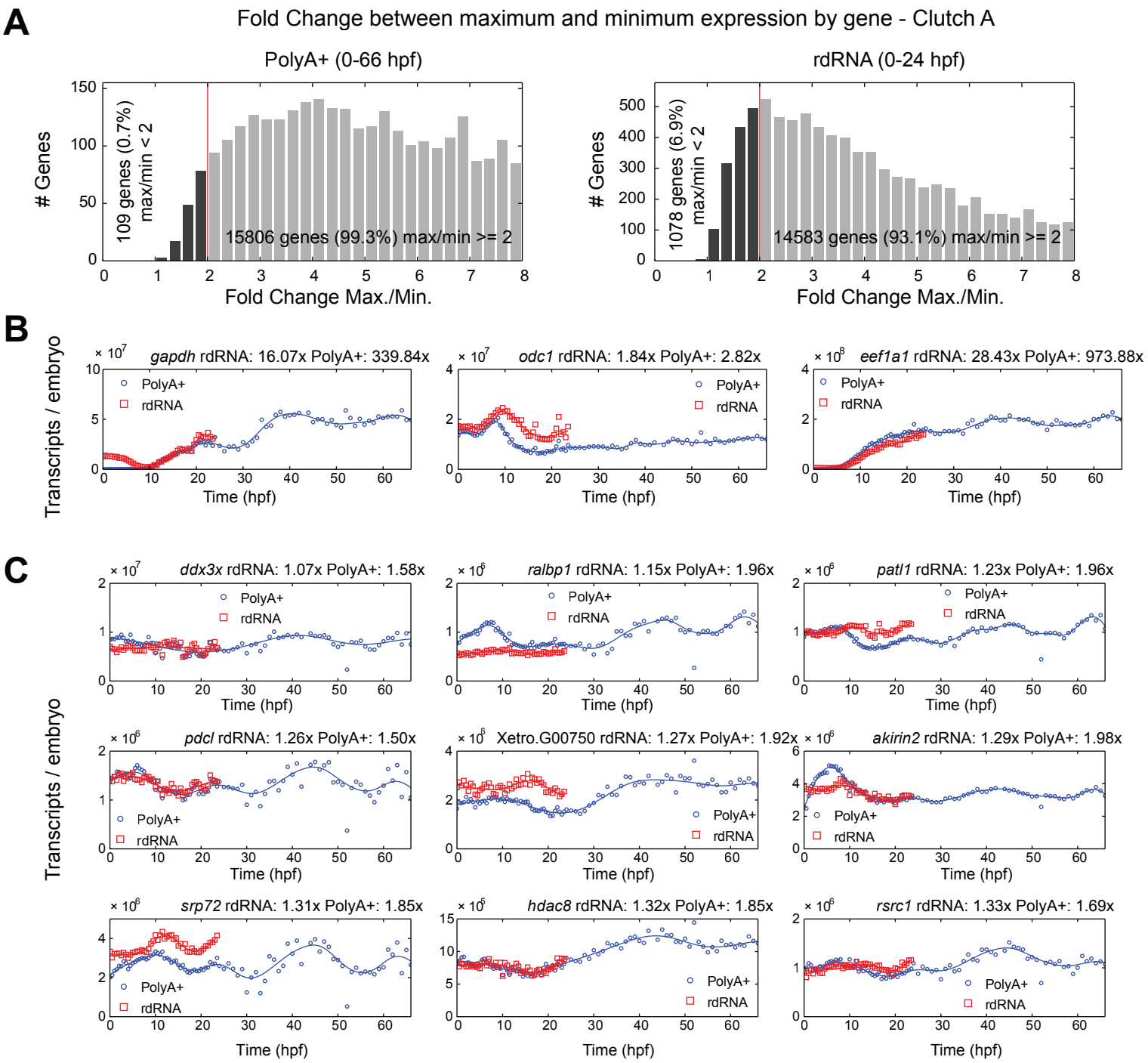

Fig S4 – Genes with constant expression – loading control candidates, related to Table S3 and Supplemental Experimental Procedures

(A) Histogram of fold change envelope between maximum and minimum Gaussian Process median abundances in transcripts per embryo for all genes for Clutch A PolyA+ and rdRNA. Candidates for loading control with a fold change < 2 are highlighted. Histogram displays up to 8-fold envelope.
(B) Traditional loading controls gapdh, odc1, eef1a1 have large fold changes between maximum and minimum expression. Fold change envelopes indicated.
(C) Candidates for loading controls with maximum/minimum fold changes < 2. Full list in Table S3.

Fig S5

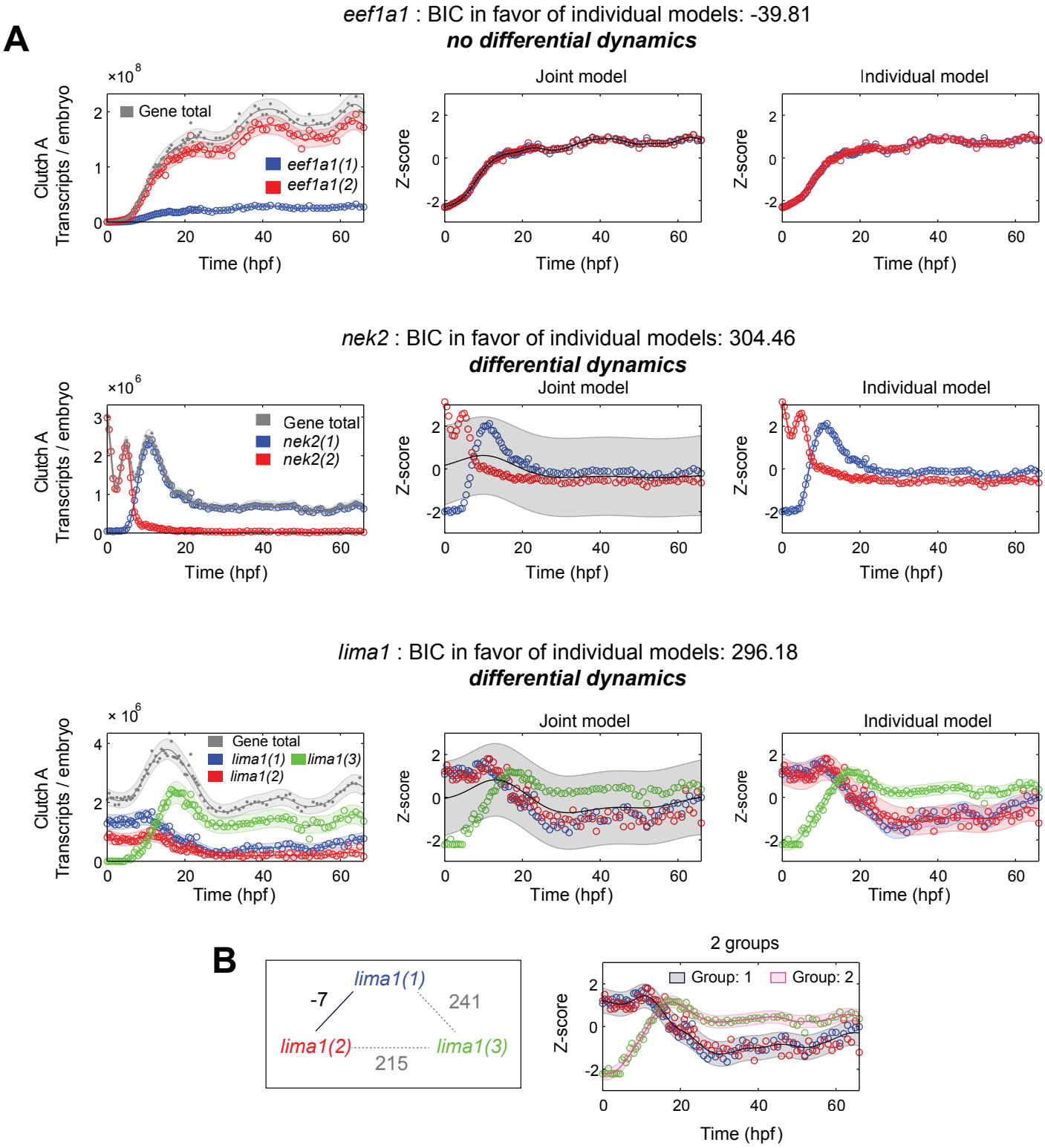

Fig S5 – Differential Isoform Dynamics in Clutch A polyA+, related to Fig 3.

(A) Explanation and detection of differential isoform dynamics. *Left*: isoform expression in transcripts/embryo. *Center*: Joint model of all isoforms assumed to have the same expression on z-normalized scale (see Supplemental Experimental Methods). *Right*: Models of all isoforms considered separately. If Bayesian Information Criterion (BIC) favors the joint model, the gene has differential abundance, whereas if the BIC favors the individual models a gene has differential dynamics. *eef1a1* has differential abundance, *nek2* and *lima1* have differential dynamics.

(B) Differential dynamics of groups of isoforms, BIC graph (left) between all *lima1* isoforms. *lima2(1)* and *lima2(2)* have differential abundance, and they have differential dynamics with *lima2(3)* defining two groups of isoforms (right).

Fig S6

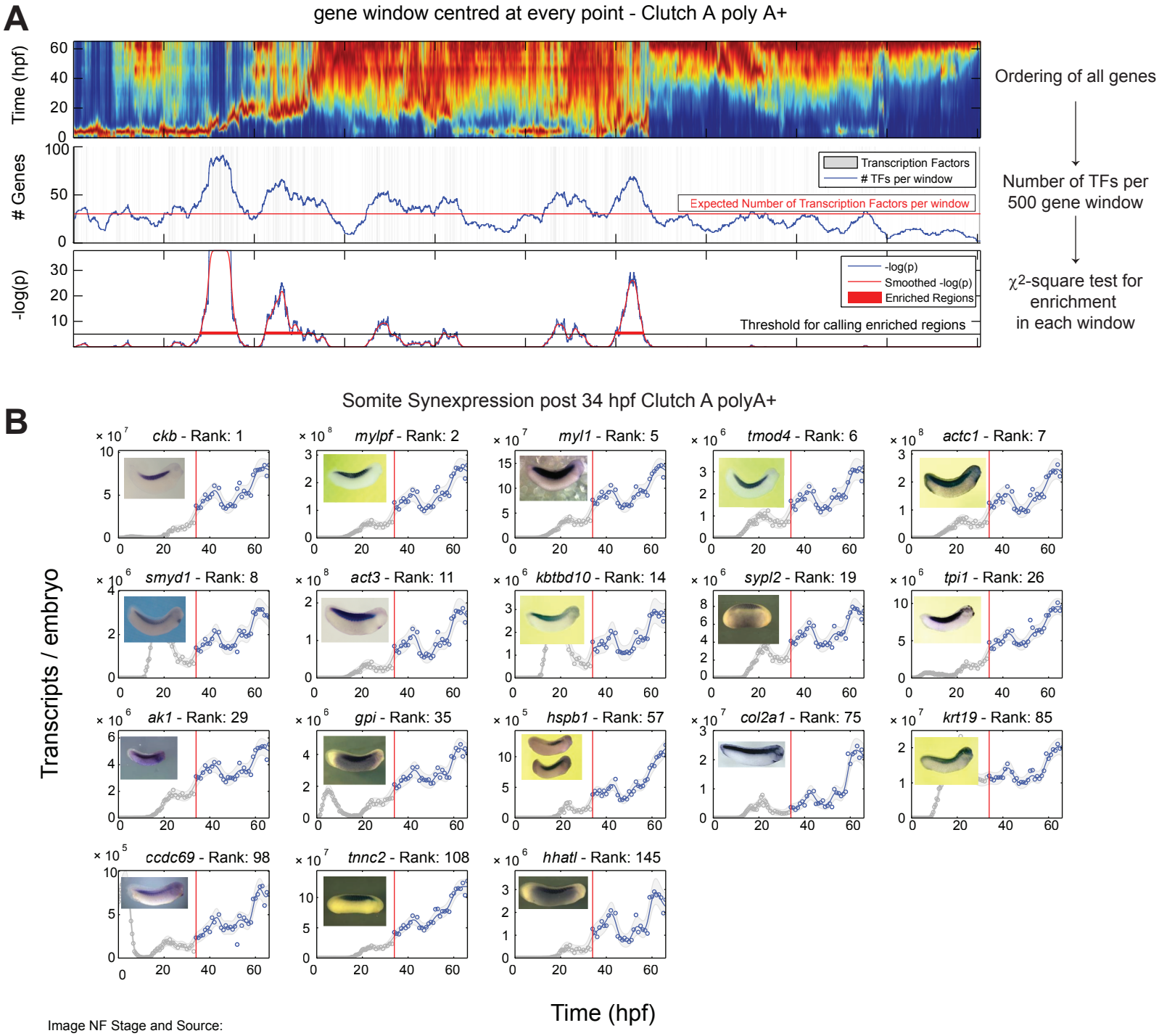

Image NF Stage and Source:
ckb - NF Stage 31-22 , this sudy
act3 - NF Stage 28, this study
mylpf, myl1, tmod4, actc1, kbtbd10, tpi1, hspb1, krt19 - NF stage 29 and 30 Copyright © CNRS UMR 8080, Nicolas Pollet Image published in: AXELDB Database
smyd1 - NF stage 29 and 30 Copyright © European Xenopus Resource Centre (EXRC)
sypl2 - NF stage 20 Copyright © XDB3, Naoto UenoImage published in: NIBB XDB3
ak1 - NF stage 28 Copyright © Papalopulu Lab, Nancy Papalopulu Image published in: Papalopulu ISH Screens
gpi - NF stage 26 Copyright © XDB3, Naoto Ueno; Image published in: NIBB XDB3 (Probe Pages)
col2a1 - NF stage 28 Copyright © Harland LabImage published in: HarlandImage source: Community submitted
tnnc2 - NF stage 27 Copyright © XDB3, Naoto Ueno; Image published in: NIBB XDB3 (Probe Pages)
hhatl - NF stage 26 Copyright © XDB3

Fig S6 – Temporal synexpression and gene ontology in Clutch A PolyA+, related to Fig 4.

(A) Sliding window gene ontology enrichment. Bottom: Heatmap of all genes (Fig 4) arranged by similarity. Middle: The location of all transcripts factors, blue line marks the number of transcription factors per 500 gene window. Red line marks the expected number of transcripts for any 500 gene window. Top: -log(p) chi-squared test for the association between the window position and the transcription factor density. A region that exceeds –log(p) > 5 is considered for annotation (Supplemental Experimental Procedures).
(B) Somite Synexpression. Examples from the somite temporal synexpresssion group. Ranks give similarity to ckb between 34-66 hpf (Supplemental Experimental Procedures). Image source given at bottom.

Fig S7

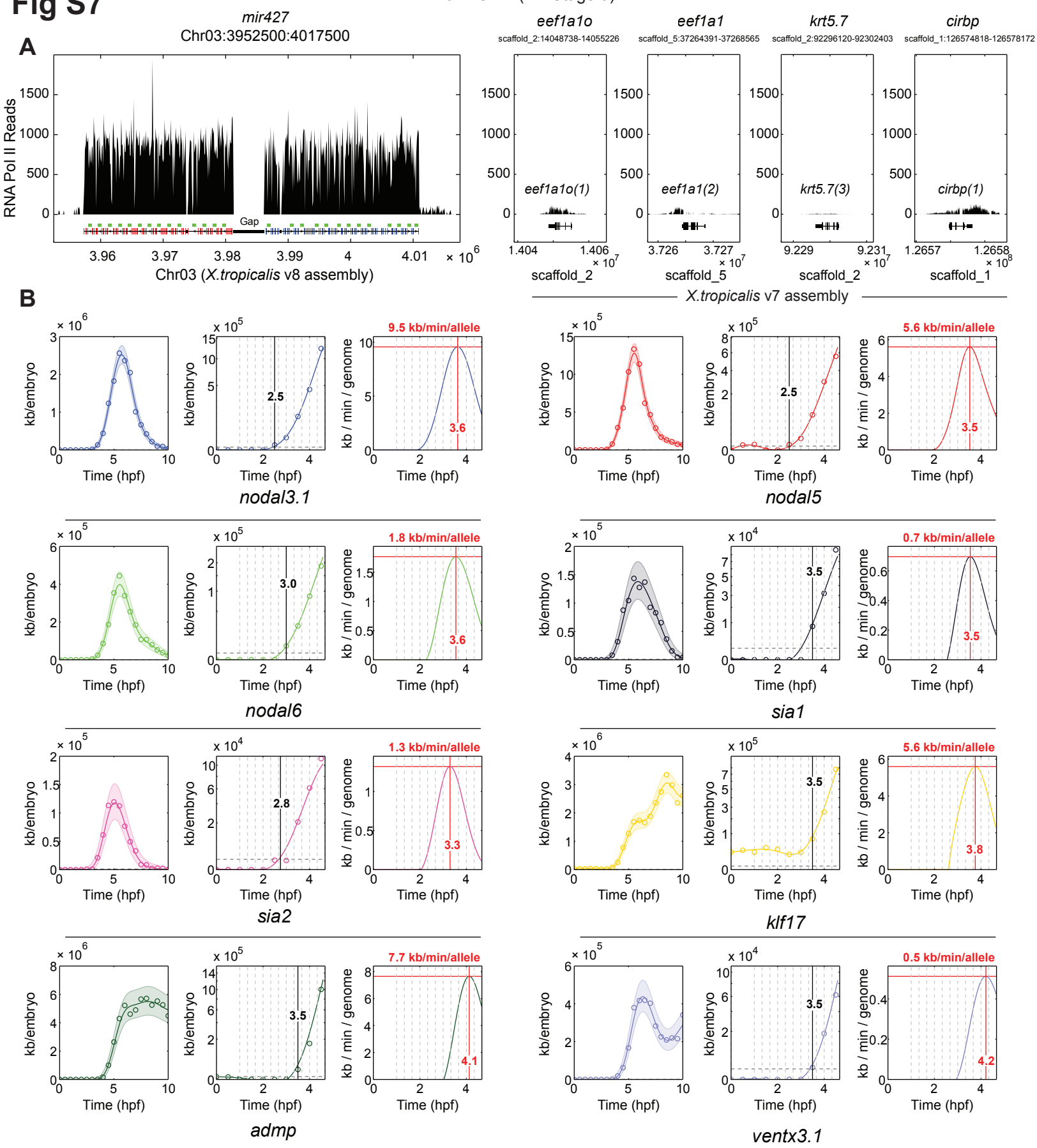

Fig S7 – Early transcription events, related to Fig 6.

(A) Stage 9 RNA pol II ChIP - Pile ups on *mir427*, *eef1a1o*, *eef1a1*, *krt5.7* and cirbp loci of RNA pol II ChIP-seq reads (van Heeringen et al., 2014). See Supplemental Experimental Procedures for calculation of *mir427* pile up.

(B) Accumulation rates in kb/min/allele averaged over all cells in the embryo – for *nodal3, nodal5, nodal6, sia1, sia2, klf17, admp* and *ventx3.1*. Left column: kb/embryo. Center column: Detection of activation of each gene during the cleavage stages. Note square root vertical scale. Vertical gray dashed lines indicate cell divisions. Horizontal dashed line gives average detection limit of 970kb for the cleavage stages. Right column: Accumulation rate in kb/min/allele, rates account for exponential increase in cell numbers and are correct if each transcript were transcribed in all cells (Supplemental Experimental Procedures).

Table S1 – Mapping of observed *Xenopus tropicalis* NF stages to time in hours post fertilisation, related to Fig 1.

NF Stage	Time (hpf)
1	0
2	1
3	1.33
4	1.67
5	2
6	2.33
7	3
8	4
9	4.5
10	6.5
11	8
12	8.5
12.5	10.5
13	12
14	13.5
15	14.5
16	15
17	15
18	15.5
19	15.5
20	16
21	16.5
22	17.5
23	18.5
24	19
25	21
26	22.5
27	24
28	27
29	29
30	29
31	30
32	31
33	33
34	33
35	36
36	37
37	38
38	39
39	41
40	46
41	51
42	56

Table S2 – Absolute Normalisation errors, related to Fig S2F, Supplemental Experimental Procedures. Genes with possible error in absolute normalization due to: 1) Non-ribosomal RNAs depleted at least 5-fold in rdRNA sequencing (Fig S2F); 2) genes with discrepant read density over UTR versus entire transcript.

Table S3 – Genes with constant expression, related to Fig S4, Supplemental Experimental Procedures. Genes with constant expression over the timecourse are candidates for loading controls. All genes with maximum/minimum fold change < 2 for PolyA+ 0-66hpf and rdRNA 0-24hpf listed (Fig S4).

Table S4 – Gene ontology enrichments, related to Fig 4, Fig 7. Gene ontology enrichments for 1) Clutch A vs Clutch B differentially expressed with large effect (Fig 2). 2) Temporal Synexpression heatmap (Fig 4), position contents and statistics of all marked blocks in main and inset heatmaps. 3) Characteristic Timescale (Fig 7). Details of min and max timescale of each block are given.

Table S5 – Somite and Vision Temporal Synexpression groups, **related to Fig 4.** Somite Synexpression: Top 150 genes with similarity to *ckb* between 34-66 hpf (Supplemental Experimental Procedures). Vision Synexpression: Identity of groups V1 and V2 given with Human orthology, OMIM and OMIA references.

Table S6 – cDNA clones used to make *in situ* probes, related to Fig 6, Fig S3.

Gene	Clone:
ace2	IMAGE:7712185
act3	Sanger: TNeu071i12
ankrd2	Sanger: TTpA004e10
col3a1	IMAGE: 7702933
cpb1	IMAGE:7003112
ctrb1	IMAGE:7840666
ctrl	IMAGE: 7007251
g6pc2	IMAGE: 7736777
hbe1	IMAGE: 7608779
mep1a	IMAGE: 7021568
mep1b	IMAGE:7623560
myf5	Sanger: TGas127b01
myod1	Sanger: Tneu017H11
pri-mir427	IMAGE:7545411
t (brachyury)	Sanger: Tgas023E12
thdl17	IMAGE: 8902999
thdl18	Sanger: TTpA005b18

Table S7 – **Primer sequences for isoform specific probes, related to Fig 3.**

	Isoform A Primers			Isoform B Primers		
Name	Forward	Reverse	Promoter	Forward	Reverse	Promoter
ckb	ACTCTCACAACACCTTGAAGC	CCCCAGTTTGAATGACATCATCC	SP6	CAATATCAGTATGGCGCAACC	GTCTGAATGACATCATCCAGGG	SP6
btbd3	AGCGGCACATAGTTCATCCC	AGATACAGCGGATCCATGCG	T7	TTGAGTCAGCGAACAGACCC	TCTTCCCTCGTGCATCTACC	T7
calhm2	TCTCAATGTGCCTACGTTGC	CCAGCCTGTAGGAGAAATGC	T7	GGGGCTGATTTCTTGCTACC	TCAGCAACAGGAAACACAGG	T7

Supplemental Data

Data S1 – Gene Models, related to Experimental Procedures. Additional gene models beyond those in *X. tropicalis* v7.2 models.

Data S2 – *mir427* locus, related to Fig 6. *mir427* locus features described in Fig 6.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Contents:

1 Sample Collection and RNA-seq Library Generation

1.1 Frog Husbandry

We employed *Xenopus tropicalis* for these studies because: 1) its high fecundity, combined with the use of *in vitro* fertilization, allows for the collection of large numbers of synchronized embryos, 2) its diploid sequenced genome facilitates genomics and 3) its evolutionary proximity to humans makes it highly applicable as a disease model and for regenerative medicine.

X. tropicalis were housed and cared for in our aquatics facility according to established protocols that were approved by Yale IACUC (Khokha et al., 2002). We obtained eggs from female frogs after injection of HCG and manually expressing eggs into a dry petri dish. IVF was performed after isolating both testes from males and applying a sperm solution to eggs after homogenizing testes with a pestle. After waiting 2 minutes, we flooded the eggs with 1/9thx MMR to activate fertilization; this was considered time zero.

1.2 RNA Isolation and Spike Ins

To obtain RNA samples for RNA-seq, we performed 2 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^{o}C, with measurements at each timepoint indicating fluctuations +/- ~1^{o}C over the entire timecourse. Unfertilized eggs and embryos were dejellied and embryos were cultured at low density, ~100 embryos per 150-mm dish containing ~120-150mls of 1/9thX MMR. For most samples, 10 embryos/timepoint were homogenized in 200 μl Trizol® and frozen at -80^{o}C, with the exception of egg, 0.5, 1, 1.5, 2, 2.5 and 3hpf timepoints, which were sampled at 25, 30, 20, 20, 20, 15 and 15 embryos respectively. Nieuwkoop and Faber staging of the embryos were noted at each time point. We labelled the progeny of the two crosses Clutch A and Clutch B. All collections from the two clutches occurred concurrently.

For the Clutch A timecourse, ERCC Spike In Mix 1 (12μl) (lot 1205006) was diluted in DEPC-treated water to a final volume of 1045μl, which was verified for accuracy by mass measurement on a Denver Instruments model 400D balance set to measure three digits past the decimal point. The dilution was performed in an Ambion Non-Stick RNAse-Free 1.5-ml microfuge tube (AM12450) to minimize loss of RNA to the tube surface. Diluted spike in mix was then pipetted directly into the Trizol homogenates such that 1μl was delivered per embryo, which in most cases required delivery of 10 μl diluted spike in mix/sample, with the exception of the first several timepoints discussed above. In all cases, a Gilson P10 pipetman was used, fitted with Denville Sharp® Precision 10μl extra long (10-XL) barrier tips (P1096-FR) to prevent loss and also to permit a visual confirmation that each 10 μl pipetting matches the 10 μl marking on the pipette tip. Spike in mix was pipeted into samples in the order of embryo sample chronology. Total RNA was purified from the embryo Trizol homogenates according to the manufacturer's recommendations. Following isopropanol precipitation, RNAs were resuspended in DEPC-treated water and any contaminating genomic DNA was removed by a subsequent overnight precipitation in 5M LiCl at 4^{o}C. RNA was pelleted and washed twice with 70% ethanol made with DEPC-treated water. All RNAs

were resuspended at 2 μl/embryo equivalent in DEPC-treated water. RNA quality was assessed by Bioanalyzer and RNA Integrity Numbers were all in excess of 9.

For the Clutch B timecourse, 8μl of ERCC spike in Mix 1 (lot 1306008) was similarly diluted to a final volume of 836μl and delivered to all Trizol® homogenates, except the pipetting was done in reverse order to the chronology of the samples. Furthermore, we added an additional spike in mix that was pipetted independently and in order of sample chronology. Ambion ArrayControl™ RNA spikes (lot 1302027) were used. Two microliters of ArrayControl™ spikes 1, 2 and 4 were diluted (using a Gilson P-2 pipetman, 10-XL tips and Ambion Nonstick tubes) to a final volume of 82 μl in DEPC-treated water to create a solution of 2.439ng for each ArrayControl spike-in RNA per microliter. Approximately 5.125μl of this dilution was then further diluted into DEPC-treated water to a final volume of 1000μl as verified by weight to create a solution of ~12.5pg/μl. One microliter per embryo (~12.5pg) was added to Trizol® homogenates by pipetting of 10μl per tube except in the cases of the 0-3hpf samples. In both timecourses these early timepoint spike-ins were pipetted by multiple rounds of 10μl (for 20 or 30μl delivery), or readjustment to 5μl (for 15μl delivery). RNAs were subsequently isolated as described above except we omitted the LiCl precipitation step. RNAs were similarly quantitated and evaluated for quality.

1.3 RNA-seq Library Construction and Sequencing

RNA quality and concentration was evaluated by running total RNA on an Agilent RNA Nano Bioanalyzer chip. All samples had an RNA integrity number (RIN) greater than 9. On average we recovered ~16μg of total RNA per sample, and 1μg of total RNA from each sample was used for library construction. Ribosomal RNA was removed via polyA-selection for mRNA, or by ribosomal RNA depletion. For polyA+ RNAs, transcripts were selected from the total RNA using two rounds of binding enrichment to oligo$(dT)_{25}$ beads (Invitrogen). For ribosome-depleted RNAs, rRNAs were eliminated from total RNA using the Ribo-Zero (Human/Mouse/Rat) Magnetic Ribosomal rRNA Removal kit (Epicenter) according to the manufacturer's recommendations. The supernatants from both approaches were each purified and concentrated using RNAclean XP beads (Beckman Coulter). Following selection, RNA was fragmented to approximately 140 base pairs and first-strand synthesis was completed with SuperScript III reverse transcriptase (Invitrogen) using random hexamers. Magnetic AMPure XP beads (Beckman Coulter) were used to purify the cDNA and remained with the sample throughout sequencing library construction. During preparation, DNA is selectively precipitated by weight and re-bound to the beads through the addition of a 20% polyethylene glycol, 2.5 M NaCl solution. We used T4 DNA polymerase and T4 polynucleotide kinase to blunt ends and phosphorylate the fragments. *E. coli* DNA polymerase (Klenow large fragment) was then used to add a single adenine residue to the 3' end of each fragment; custom adapters (IDT) are ligated using T4 DNA ligase. The remaining product is then PCR amplified using custom-made primers (IDT) containing unique 6 bp indices.

In the primary cross (Clutch A, Fig 1A), we sequenced polyA+ mRNA covering the entire time course (66 hours, 90 samples) and rdRNA for the first 24 hours (48 samples). For the secondary cross (Clutch B, Fig 1A), we sequenced polyA+ mRNA for the first 24 hours only (48 samples). We generated ~7 billion paired end reads over 145 RNA samples with an

average of ~36 million reads per library for a total of 1 trillion bp (750x the size of the *Xt* genome). Frog A polyA+ libraries yielded 34 million paired-end reads on average, Frog B polyA+ 31 million paired-end reads, and Frog A rdRNA 43 million paired-end reads.

2 High-throughput sequencing data processing

2.1 *Xenopus tropicalis* transcriptome and genome

In this study we use the *Xenopus tropicalis* version 7.1 genome (Hellsten et al., 2010; Karpinka et al., 2015) and the associated *X. tropicalis* v7.2 gene models (Karpinka et al., 2015). To this we add known off-assembly gene sequences used in a similar study (Collart et al., 2014), models for ERCC and ArrayControl spikes, and a model for the *mir427* locus (see Section 2.1.2). We corrected a number of gene models known to be erroneous in the v7.2 assembly, predominantly these are models of short locally repeated families such as the *nodal3/5*, *bix*, and *sox17* family genes. For convenience we model these short repeated families to report the aggregate signal from each copy as a total "gene" expression, and report the signal from each copy as transcript "isoforms" of the family. For example, *nodal5* is present with 4 copies, we report the aggregate of all the copies as *nodal5* as we would a gene, and we report the individual copies *nodal5.1-nodal5.4* as we would isoforms of the gene. Corrected and additional gene models beyond the *X. tropicalis* v*7.2* models are given in Data S1. The final transcriptome contains 28,274 genes with 43,568 distinct transcript isoforms. Of these isoforms, 31,928 (73%) have annotated coding and UTR sequence. Of the 28,274 genes, 15,277 (54%) have a gene symbol assigned by orthology to other species. Overall, we find that 85.9%, 79.3%, 71.6% of reads align to the genome and off-genome EST sequences for Clutch A PolyA+, Clutch A rdRNA and Clutch B respectively.

2.1.2 *mir427* locus

The sequence of the *mir427* locus is identical in *X. tropicalis* genome builds v7, v8 and v9. In v7 it is located scaffold_3b:3515270-3571270, in v8 Chr03:3956500-4012500 and v9 Chr03:133289810-133345810. We provide the coordinates for the features marked in (Fig 6), in Data S2. In this work we do not attempt to distinguish the abundance or accumulation rates of individual features of the *pri-mir427* locus, but we report the aggregate abundance and accumulation of the entire locus.

2.2 RNA-seq read alignment and estimation of relative abundance

All libraries were sequenced at 76bp paired-end on an Illumina HiSeq 2000 (GEO GSE65785). Read pairs were aligned using TopHat v2.0.10 (Trapnell et al., 2009) to the *Xenopus tropicalis* version 7.1 genome described above along with our ERCC & ArrayControl exogenous spike sequences. Spliced alignment was guided by the gene models, and TopHat was run with the following parameters "\-r 13 --mate-std-dev 60 -G v7models -g 200 --report-secondary-alignments". Overall, we find that 85.9% of reads align in Clutch A PolyA+, 79.3% of reads align in Clutch B PolyA+ and 71.6% of reads align in Clutch A rdRNA.

We estimated the relative abundance of all transcripts in *fragments per kilobase per million* (FPKM) with Cufflinks v2.1.1 (Trapnell et al., 2010) with the parameters "--compatible-hits-norm --max-bundle-frags 10000000 -p 24 -G v7models -b xt7.fa -u --max-mle-iterations 50000". As such, ambiguous reads are not discarded, and Cufflinks assigns these with an expectation-maximization procedure. For an absolute normalization to be successful is essential that no ambiguous reads are discarded. We did not allow Cuffinks to assemble

new transcripts; new transcript discovery from this data will be the subject of a subsequent manuscript.

2.2.1 Relative abundance corrections for repetitive loci

We found that Cufflinks underestimates the read mass aligning to certain locally repetitive loci.

We compared the read mass aligning to each gene reported by Cufflinks to that obtained by naïve counting in which we count the number of reads aligning to one or many transcripts of a given gene and to no transcripts of any other gene. As we discard reads that map to multiple genes, the read mass found by naïve counting should be less than or equal to that found by Cufflinks up to sequencing bias corrections applied by Cufflinks. In general, we find this to be the case with mean fold change of 0.92 (on log2 scale -0.18 +/- 0.35 [mean +/- std. dev]) between the naïve counting mass and the Cufflinks mass, and near identical results for genes with no ambiguous reads. However, we found discrepancies for a small number of genes in which the naïve counting read mass significantly exceeded that of Cufflinks. We find 12 genes for which the naïve counting read mass is two-fold larger than that of Cufflinks, and of these we find 5 known genes which exceed by over four-fold: *pri-mir427* - 190x, *nodal5* - 10x, *sst* - 9.5x, *nodal3* - 4.6x, Xetro.K01272 – 4.1x. All of these genes have locally repeated loci, *pri-mir427* is the most repetitive with over 160 copies of the mir427 hairpin, *nodal3* is present with 5 copies, *nodal5* is present with 4 copies, and *sst* is present with 2 copies. Unnamed gene model Xetro.K01272 contains a sequence repeated locally and genome wide, it is adjacent to a gap in the genome assembly on a short scaffold and its true nature is unclear. Of the remainder of the 12 with fold change greater than two-fold, most also exhibit local repetitions albeit to a lesser degree. We note that for *pri-mir427* the Cufflinks read mass is approximately equal to the number of reads that only align once to the locus. It appears that Cufflinks unduly down-weights reads that align locally many times, although other loci present in multiple local copies appear unaffected.

We correct the abundances of the most severely affected genes: *pri-mir427, nodal3, nodal5,* and *sst*. We do not correct Xetro.K01272, as we cannot be confident in its model. We return to the original read alignments, and use a simple expectation-maximization procedure that does not account for sequencing biases to assign reads between the transcripts of these genes. We discard reads that align ambiguously to one of these genes and any other gene. For *nodal3, nodal5* and *sst* we find no such reads to discard. For *pri-mir427* we find that 0.3% of reads over all samples (and 0.001% of reads between 0-10hpf in rdRNA sequencing) that align to the *mir427* locus also align elsewhere in the genome, we discard these reads. These reads map to the extreme flanks of the *mir427* locus, and importantly, they do not exhibit the same temporal profile as the remainder of the locus. We can be confident that they do not originate from *pri-mir427* transcripts and so these reads should be discarded for correct estimation of *pri-mir427* absolute abundance. We calculate the FPKM for all transcripts of these genes using effective lengths calculated using the fragment size statistics reported by Cufflinks. We update the FPKM of all transcripts in light of the additional read mass.

We convert transcript abundances in FPKM and convert to *transcripts per million* (TPM). We sum the expression of isoforms of each gene to estimate the TPM per gene.

2.3 Analysis of RNA Pol II ChIP

We demonstrate that there is significant RNA Pol II activity over the *mir427* locus. We took publically available NF stage 9 RNA Pol II ChIP-seq data, GEO accession GSE41161 (van Heeringen et al., 2014). We aligned reads to the *X. tropicalis* v7 genome with Bowtie2 with parameter "-k 80". We generated pile ups for all loci in Fig S7 from unique reads with the exception of *mir427*. Here, for a read with n best alignments within the *mir427* locus and no alignments outside the locus we assign a read mass of $1/n$ to all best alignment locations.

3 Modeling RNA-seq time series data

3.1 Relative Transcript Abundance

We introduce some concepts relevant to relative transcript abundance estimation with RNA-seq, a more complete overview is found elsewhere (Trapnell et al., 2009; Trapnell et al., 2010). Sample $j = 1, \ldots, S$ is collected at t_j hours post fertilization, and transcript $i = 1, \ldots, M$ with length l_i present with m_{ij} copies, and the total number of transcript copies at time t_j is $T_j = \sum_i m_{ij}$. From RNA-seq data we estimate two relative abundances: 1) relative transcript abundance: $\rho_{ij} = m_{ij}/T_j$; 2) relative fragment abundance (α_{ij}) and describes the expected proportion of sequenced fragments derived from transcript i: $\alpha_{ij} = m_{ij}\,\tilde{l}_i / \sum_k m_{kj}\tilde{l}_k$ here $\tilde{l}_i$ is the effective transcript length that accounts for the RNA-seq fragment size distribution (Trapnell et al., 2010). Note that the FPKM is given by $f_{ij} = 10^9 \alpha_{ij}/\tilde{l}_i$, and a FPKM abundance can be converted into the relative transcript abundance as $\rho_{ij} = f_{ij}/\sum_k f_{kj}$. We report relative transcript abundances in TPM defined as $10^6\rho_{ij}$.

3.2 Biological and Technical Noise

The technical noise of re-sequencing the same library has been shown to be well approximated by a Poisson distribution (Marioni et al., 2008). If sample j is sequenced to a depth of d_j and $v_{ij}$is the number of fragments sequenced from transcript i then $\mathbb{E}[v_{ij}] = d_j\mathbb{E}[\alpha_{ij}]$ and by the law of total variance $\mathbb{V}[v_{ij}] = d_j\mathbb{E}[\alpha_{ij}] + d_j^2\mathbb{V}[\alpha_{ij}]$. Consequently v_{ij} is overdispersed or heteroskedastic; its variance is greater than its mean by the variance of α_{ij} (the biological noise) scaled by the sequencing depth. We can apply a transform to approximately stabilize the variance of v_{ij}. If the variance of α_{ij} depends linearly on $\mathbb{E}[\alpha_{ij}]$ and some biological noise parameter σ_{ij} such that $\mathbb{V}[\alpha_{ij}] = \beta_1\mathbb{E}[\alpha_{ij}] + \beta_2\sigma_{ij}$ and if $y_{ij} = \sqrt{v_{ij}}$ then $\mathbb{V}[y_{ij}] \approx a + b\sigma_{ij}/\mathbb{E}[\alpha_{ij}]$ for positive constants a and b. The square root transformation is not a perfect variance stabilizing transformation, but significantly reduces the dependence of the variance on the expression level.

3.3 Gaussian Process Models of Gene Expression

We model our gene expression time series with Gaussian Processes (Rasmussen and Williams, 2006), examples of their application to biological data are increasing, including examples of approaches similar to ours (Aijo et al., 2014; Heinonen et al., 2014; Honkela et al., 2010; Stegle et al., 2010; wa Maina et al., 2014). We have estimates of a gene's abundance $\mathbf{g} = (g_1, g_2, \dots, g_n)^\top$ measured at times $\mathbf{t} = (t_1, t_2, \dots, t_n)^\top$. Here, $\mathbf{g}$ is non-negative and has variance as outlined above. Rather than modelling the likelihood of this data explicitly, we opt to transform the data using the Anscombe transform: $\mathbf{y} = h(\mathbf{y}) = \sqrt{\mathbf{g} + 3/8}$, and then assume that the transformed data are Gaussian distributed, we address the validity of this assumption below.

We assume that the transformed variables are noisy observations from some underlying process, $f(t)$, that describes gene expression:

$$y_t = f(t) + \epsilon, \quad \text{with } \epsilon \sim \mathcal{N}(0, \sigma_n^2),$$

ϵ is a zero mean Gaussian random variable with variance σ_n^2. We apply a Gaussian process prior $p(\mathbf{f}|\mathbf{t}) = \mathcal{N}(\mathbf{0}, \mathbf{K})$ with covariance matrix entries defined by the covariance function $\mathbf{K}_{ij} = K(t_i, t_j)$. In the majority of cases we use the stationary Matérn class of covariance functions:

$$K_M(t_i, t_j) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}r}{\tau}\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}r}{\tau}\right), \quad \text{with } r = |t_i - t_j|;$$

$\nu > 0$ controls the roughness of the process; $\tau > 0$ describes the characteristic timescale; and $K_\nu$is a modified Bessel function. We select the Matérn covariance function over the infinitely differentiable squared exponential, $\exp\left(-\frac{r}{2\tau^2}\right)$, as we find it better describes the features of gene expression, and it has been pointed out that squared exponential is unrealistically smooth for modelling a physical process (Rasmussen and Williams, 2006; Stein, 1991). Rather than treating ν as a hyperparameter we find that setting $\nu = 5/2$ robustly captures the features of gene expression in our data. In this case the covariance function reduces to $K_{M,\nu=5/2}(t_i, t_j) = \left(1 + \sqrt{5}\, r/\tau + 5r^2/(3\tau^2)\right)\exp(-\sqrt{5}r/\tau)$. We scale our covariance by a parameter σ_f so $\mathbf{K}_{ij} = \sigma_f K_M(t_i, tj)$. We treat σ_f, τ, σ_n as hyperparameters, and select their values by optimizing the log marginal likelihood (see below).

The timescale τ is particularly important here. In Gaussian Process literature it is more commonly known as the *length-scale* (Rasmussen and Williams, 2006), due to our temporal setting, we refer to it as the characteristic timescale of gene expression (Fig 7). The timescale can be understood by the number of upcrossings of the process above a particular threshold over the unit interval (Rasmussen and Williams, 2006). Roughly, for a gene whose expression is characterized by a process with a timescale τ and the Matern covariance function with $\nu = 5/2$, the expected number of times a gene is able to cross above its mean level of expression in one hour will be $\sqrt{5/3}(2\pi\tau)^{-1}$. Or, that a gene's expression will cross above its mean level of expression once every $2\pi\tau\sqrt{3/5}$ hours on average.

In certain instances we make use of the non-stationary Gibbs covariance function to allow a time-varying timescale:

$$K_G(t_i, t_j) = \left(\frac{2\tau(t_i)\tau(t_j)}{\tau(t_i)^2 + \tau(t_j)^2}\right)^{1/2} \exp\left[-\frac{(t_i - t_j)^2}{\tau(t_i)^2 + \tau(t_j)^2}\right],$$

we set $\tau(t) = \tau_e - (\tau_e - \tau_s)\mathrm{e}^{-t/\tau_r}$ so that the timescale is τ_s at $t = 0$ and increases or decreases toward τ_e at rate τ_r. We treat τ_s, τ_e, τ_r as hyperparameters and constrain them to be positive. Similarly, we make use of a covariance function to describe time varying noise:

$$K_N(t_i, t_j) = \begin{cases} \sigma_n(t_i) & \text{when } i = j \\ 0 & \text{otherwise} \end{cases},$$

with $\sigma_{n(t)} = \sigma_e - (\sigma_e - \sigma_s)\mathrm{e}^{-t/\sigma_r}$ and $\sigma_s, \sigma_e, \sigma_r$ as hyperparameters and similarity constrained to be positive.

Conditioning on observed data yields a Gaussian predictive distribution for y at time $t_\star$, $y_{t_\star} \sim \mathcal{N}(\mu_\star, \sigma_\star^2)$:

$$\mu_\star = \mathbf{K}_{t_\star,\mathbf{t}}\left[\mathbf{K}_{\mathbf{t},\mathbf{t}} + \sigma_n^2\mathbf{I}\right]^{-1}\mathbf{y},$$
$$\sigma_\star = \mathbf{K}_{t_\star,t_\star} - \mathbf{K}_{t_\star,\mathbf{t}}\left[\mathbf{K}_{\mathbf{t},\mathbf{t}} + \sigma_n^2\mathbf{I}\right]^{-1}\mathbf{K}_{\mathbf{t},t_\star} + \sigma_n^2 .$$

To report predictions we invert our original data transform and report quantiles of the transformed distribution. Commonly, we report the median $h^{-1}(\mu_\star)$ and the 95% confidence interval: $h^{-1}(\mu_\star \pm 1.96\sigma_\star)$.

We treat our scale, timescale and noise parameters as hyperparameters $\boldsymbol{\theta} = (\sigma_f, \tau, \sigma_n)$ in case of the Matérn covariance function, and $\boldsymbol{\theta} = (\sigma_f, \tau_s, \tau_e, \tau_r, \sigma_s, \sigma_e, \sigma_r)$ in the nonstationary case. We select our hyperparameters by maximizing the *marginal likelihood* (or the *evidence*):

$$p(\mathbf{y}|\mathbf{t}, \boldsymbol{\theta}) = \int p(\mathbf{y}|\mathbf{f}, \mathbf{t}, \boldsymbol{\theta})p(\mathbf{f}|\mathbf{t}, \boldsymbol{\theta})d\mathbf{f}$$

The log marginal likelihood with the Gaussian prior and noise model used here is given by $\log p(\mathbf{y}|\mathbf{t}, \boldsymbol{\theta}) \equiv \mathcal{L}(\mathbf{y}|\mathbf{t}, \boldsymbol{\theta})$:

$$\mathcal{L}(\mathbf{y}, \mathbf{t}, \boldsymbol{\theta}) = -\frac{1}{2}\mathbf{y}^\top\left[\mathbf{K}_{\mathbf{t},\mathbf{t}} + \sigma_n^2\mathbf{I}\right]^{-1}\mathbf{y} - \frac{1}{2}\log\left|\mathbf{K}_{\mathbf{t},\mathbf{t}} + \sigma_n^2\mathbf{I}\right| - \frac{n}{2}\log 2\pi.$$

To perform the optimization, we used gradient descent to minimize the negative log marginal likelihood using `fminunc` in (MATLAB, 2014). To help ensure that a global minimum had been reached, we used a further round of minimization with the Nelder-Mead Simplex method using `fminsearch` (MATLAB, 2014). As all hyperparameters are positive we perform the optimization in the log-space of all parameters. Occasionally, this optimization procedure selects biologically unrealistic values for hyperparameters. In almost all of these cases, the optimization selects a timescale that is excessively short with a noise level that is unrealistically small. To ensure an optimum is selected within a biologically reasonable region of parameter space, we apply a prior to hyperparameters. To select a suitable prior we empirically evaluate the distribution of the maximum likelihood estimates of $\log\widehat{\boldsymbol{\theta}}$ found by maximizing the likelihood without a prior. Treating each parameter independently we find that a log-normal distribution to be a good fit for the main mass of each $\hat{\theta}_i$. We place a

Gaussian prior on each $\log \theta_i$ centered on the empirical mode of each $\hat{\theta}_i$, with variance greater than the empirical variance and then peak and minimize $-\mathcal{L}(\mathbf{y}, \mathbf{t}, \boldsymbol{\theta}) - \log p(\boldsymbol{\theta})$.

Our sampling intervals $\Delta t = 0.5, 1$ define a detection limit for genes with fast timescales. If we consider the power spectrum (Rasmussen and Williams, 2006) of the covariance function, then we will not be able to reliably detect those frequency components that are faster than the Nyquist frequency associated with our sampling rate. The power spectrum for the Matérn covariance function in one dimension with $\nu = 5/2$ is:

$$S_M(s, \tau) = 400\sqrt{5}\tau(5 + 4\pi^2 s^2 \tau^2)^{-3} .$$

We calculate a theoretical timescale detection limit by finding the smallest value of τ such that $\int_{-1/2\Delta t}^{1/2\Delta t} S_M(s, \tau) > \alpha$. Setting $\alpha = 0.99$ we find we can reliably detect timescales as small as 0.6 hours and 1.1 hours for 0.5 hour and 1 hour sampling respectively. As our measurements are not noise free our actual detection limits are larger than this. We are nevertheless confident that we capture all relevant behavior as the mass distribution of calculated timescales lies well away from these limits. Further, extreme transcriptional events with the fastest observed timescales, such as *pri-mir427* (Fig 6, Fig 7), are well characterized by our 30 minute sampling interval.

4 Data Quality

Alignment rates were high for the *X. tropicalis* genome. In Clutch A polyA+ 84.9% of all pairs mapped; in Clutch A rdRNA 70.5% of pairs mapped, with only 3.5% of reads aligning to ribosomal sequence demonstrating excellent ribosomal RNA depletion; in Clutch B polyA+ 78.9% of reads mapped.

To assess the quality of our data, we compare the Spearman correlation shared between all pairs of samples within each series (Fig S1A). The samples exhibit excellent correlation, neighboring samples share a Spearman correlation coefficient close to 1; they are as similar as biological replicates. Roughly, the samples fall into two groups: those prior to 10 *hours post fertilization* (hpf) and those after 10 hpf. Poorly performing samples, such as 52 hpf, in Clutch A PolyA+ are clearly visible, we address these samples below. Note that all samples prior to 4.5 hpf (the period before maternal clearance and widespread zygotic transcription) in the rdRNA share very high correlation.

Fig S1B gives the Spearman correlation between the three series as a whole. The reproducibility is greater between biological replicates Clutch A and Clutch B, then between different sequencing protocols of the same RNA polyA+ and rdRNA. This is due to large changes in polyadenylation state of many maternal genes post-fertilization.

We assessed the reproducibility of our spiking of RNA standards (Fig 1C). We observed excellent reproducibility between each series and within each series. The largest discrepancy can be seen in comparisons between polyA+ and rdRNA sequencing. Nearly all standards have a greater relative abundance in rdRNA samples and specific spike species can be seen to shift consistently away from the identity line. We attribute this to different polyA+ capture efficiencies of the spike in species (Qing et al., 2013; SEQC MAQC-III Consortium, 2014) and that polyA+ selection is less efficient for our RNA standards than for endogenous genes. ERCC-00116 performs especially poorly and is excluded from absolute normalization calculations. We note that the relative abundances of the RNA standards in Clutch A samples are approximately 1.6 fold greater than those in Clutch B samples, reflecting differences in the quantity of RNA standard in each sample due to batch and preparation variations.

We found TPM abundances of certain samples to be consistently diverged from the trend of many genes. This effect is partly visible in the PCA (Clutch A polyA+ 52 hpf sample, Clutch B polyA+ sample 22.5 hpf). To investigate this further, we evaluated *leave-one-out* (LOO) (Rasmussen and Williams, 2006) residuals, in this scheme we leave out each sample in turn and calculate its predictive distribution given the remainder of the data: $p(y_i|\mathbf{y_{-i}}, \mathbf{t}, \boldsymbol{\theta})$. We maximize the log marginal likelihood on the full data, and then the LOO mean and variance: μ_{ij} and σ_{ij}^2 for gene i in sample j can be efficiently calculated as $\mu_{ij} = y_{ij} - [\mathbf{K}^{-1}\mathbf{y}]_j / [\mathbf{K}^{-1}]_{jj}$ and $\sigma_{ij}^2 = 1/[\mathbf{K}^{-1}]_{jj}$. We sum the squared normalized residuals over all m genes for each sample $q_j = \sum_{i=1}^{m} \left(y_{ij} - \mu_{ij}\right)^2 / \sigma_{ij}^2$, and so q_j will by χ^2-distributed with m degrees of freedom. We calculate a LOO score: $z_j = (q_j - m)/\sqrt{2m}$ will approximately have a standard normal distribution. In reality, q_j will have less than m degrees of freedom as some genes will have zero expression, but the construction serves well to identify outlying samples (Fig S1D). Note, those samples adjacent to poor performing samples can also receive a large positive value of z_j, we confirm there are no adjacent poorly performing samples by recalculating a

similar score to z_j but with adjacent large scoring samples removed (data not shown). We select the six samples with largest z_j in Clutch A PolyA+, and returned to the original RNA purifications and re-generated new RNA libraries and re-sequenced. We observe a remarkable technical reproducibility (Fig S1E), including highly reproducible aberrant behavior. The cause is not clear, but AT-rich transcripts are most affected. We excluded the 3 worst samples: 21.5 hpf, 33hpf and 52 hpf from all further Clutch A analysis, and we exclude sample 22.5 hpf from all further Clutch B analysis. Despite the exclusion of these samples from our analysis, we continue to plot these data throughout this work.

To further demonstrate the strong temporal relationship between our samples we performed a principle component analysis (PCA) on each of the three series independently (Fig S1F). The samples plot a clear and dense trajectory through the projections onto the first two principle components. Moreover, the major transcriptomic events, particularly those related to the maternal to zygotic transition (Fig 1, Fig 5), are clearly observable. The influence of changes in polyadenylation is present in polyA+ sequencing, but not rdRNA sequencing.

5 Direct Absolute Normalization of RNA-seq data

To convert relative read transcript abundances into absolute transcript numbers, we added a mix of 92 spike-in RNAs from the External RNA Controls Consortium (ERCC; Ambion) along with an independently generated mix of 3 spike-in RNAs derived from the ArrayControl set as RNA quantitation standards to each time point sample. We avoided adding spike-in RNAs per quantity of recovered, purified RNA since this approach is highly error-prone. Instead, we added a known amount of these spike-in RNAs directly to each embryo Trizol homogenate on a per embryo basis prior to RNA extraction, subjecting the RNA standards to the same variations in recovery as the endogenous embryonic transcripts during RNA purification, library preparation, and sequencing.

5.1 Absolute Normalization Methodology

We apply a two-step normalization. First, we perform a relative normalization for sequencing depth and calculate the relative abundance in TPM of all RNA standards and native transcripts. Here, it is key to note that the relative abundance of the RNA standards decrease with a smooth trend with time (Fig 1B, S2A). This reflects the increasing RNA content of the embryo with time.

We combine the known transcript copy numbers of each RNA standard with their relative transcript abundances to estimate the total number of transcripts per sample and discover conversion factors from relative to absolute abundances.

Formally, we add $q = 1, \ldots, Q$ RNA standard species each present with a copy number s_q to each sample, the total number of copies of standards is $S = \sum_q s_q$. The relative abundance of an endogenous transcript is $\rho_{ij} = m_{ij}/(T_j + S)$ and the relative abundance of a RNA standard is $r_{qj} = s_q/(T_j + S)$. The variation in r_{qj} for a given RNA standard species over all samples will arise from the following sources:

1. Batch variation, both in manufacturing and in RNA standard sample preparation. This results in the TPM abundances of spikes in Clutch A PolyA+ being approximately 1.6 fold higher than those in Clutch B PolyA+ (Fig S1C).
2. Changes in the total number of transcripts T_j. We assume that T_j varies smoothly with time, and samples close in time will have a more highly correlated total number of endogenous transcripts than those more distant in time. This can clearly be seen in Fig 1B, S2A, where the trend of relative normalized spikes vary smoothly with time.
3. Sample noise, including pipetting error. This is also observable in Fig 1B, Fig S2A as the deviations from the smooth spike trend lines.

We opt for a practical strategy for dealing with the above sources of variation: we estimate a parameter related to T_j for each library, we then smooth these estimates to obtain the trend in T_j that is not contaminated by RNA standard sample noise. We combine a Generalized Linear Model with our Gaussian Process framework. We use a generalized linear model with a dispersed Poisson likelihood to each sample with the following link function:

$$\log \hat{r}_{qj} = \beta_j + \log s_q,$$

where $\hat{r}_{qj}$ is the abundance of RNA standard q in sample j in TPM, and as a consequence $\mathrm{e}^{-\beta_j} \approx (T_j - S)/10^6$. We then calculate smooth $\bar{\beta}_j$ using Gaussian Processes with Matérn

covariance function, which describe the trend without sample noise (Fig S2B). We then can convert a relative transcript abundance to TPM, $\hat{\rho}_{ij}$, to an absolute abundance by $\hat{m}_{ij} = \hat{\rho}_{ij}\mathrm{e}^{-\overline{\beta}_j}$, and similarly for the RNA standards $\hat{s}_q = \hat{r}_{qj}\mathrm{e}^{-\overline{\beta}_j}$.

5.2 Corrections for PolyA+ Sequencing Bias and Batch Variation

The performance of ERCC spikes in PolyA+ sequencing has been reported to be poor (Qing et al., 2013; SEQC MAQC-III Consortium, 2014), and our data is in agreement as the spike transcripts are captured less well than their endogenous counterparts in a polyA+ selection (Fig S1C). Indeed, we find the calculated absolute normalization factors to be higher for Clutch A polyA+ samples than rdRNA samples (Fig S2B). We address this by assuming that on average only a proportion $0 < \phi < 1$ of all PolyA+ spikes present in the embryo homogenate are captured. The relative abundance in PolyA+ samples in TPM will now become:

$$\tilde{\rho}_{ij} = \frac{10^6 m_{ij}}{T_j + \phi S} \text{ for endogenous transcripts and } \tilde{r}_{qj} = \frac{10^6 \phi s_q}{T_j + \phi S} \text{ for RNA standards.}$$

Applying the absolute normalization as above will still yield the correct result for RNA standards as $\hat{s}_q = \tilde{r}_{qj}\mathrm{e}^{-\overline{\beta}_j}$, but will be inaccurate for endogenous transcripts as $\hat{m}_{ij}\phi^{-1} = \tilde{\rho}_{ij}\mathrm{e}^{-\overline{\beta}_j}$. To address this we obtain an estimate of ϕ by assuming that the majority of measured transcripts are polyadenylated over a particular interval in development. This amounts to assuming the T_j are approximately equal in the polyA+ and rdRNA preparations for certain samples, so $\hat{\phi}_t = \left(10^6\mathrm{e}^{\overline{\beta}_{j_r}} - S\right)/\left(10^6\mathrm{e}^{\overline{\beta}_{j_p}} - S\right)$ where j_r and j_p denote rdRNA and polyA+ samples at time t respectively. To correct our absolute normalization we use the median value of $\hat{\phi}_t$ for t in the range [6, 23.5] when the polyadenylation of maternal transcripts has completed and the clearance of non-polyadenylated transcripts has initiated and it is reasonable to assume that all mRNAs that the vast majority measure are polyadenylated; this gives $\hat{\phi} = 0.56$.

In Clutch B polyA+ the ERCC spikes have lower relative abundance estimates than in the Clutch A polyA+. This reflects variation in spike batch preparations and the result of these differences inflates Clutch B polyA+ absolute normalization correction factors (Fig S2B). We assume that both sets of normalization factors are equally likely and average their differences by shifting the normalization factors to the median geometric mean over all time points.

After applying PolyA+ and Clutch A/B corrections we observe excellent agreement between the otherwise independently calculated $\mathrm{e}^{-\beta_j}$ for the three datasets (Fig S2B). We use corrected factors to absolutely normalize endogenous genes and use uncorrected factors to normalize RNA standards.

5.3 Absolute Normalization between 40-66 hpf

There appears to be a greater level of variability in the RNA standards between 40-66 hpf (Fig S2B); the variability results in a wave in the absolute normalization correction factors. Contrasting the loading control examples (Fig S4), we find the wave is absent from the absolute normalized expression profiles of some but not all genes. This suggests that for some, but not all genes, this wave in the correction factors is appropriate. In the absence of

replicates covering this region we cannot ascertain whether this is a biological or a technical phenomenon. As the effect is small enough not to alter any of our conclusions we do not attempt to alter the correction factors for any individual gene traces. However, when presenting certain genome wide results, particularly, the ng /embryo estimates (Fig 1C) and the transcriptome synexpression map (Fig 4), we opt to use a stronger smoothing of absolute normalization factors. We use the non-stationary Gibbs covariance function in combination with the time varying noise covariance function to produce a covariance function with increasing timescale and noise level that increase with time (Fig S2B, far-right). In bias-variance trade-off terms, this equates to a higher bias and lower variance model of the data.

5.4 Performance and Uncertainty of Absolute Normalization

We applied the absolute normalization correction factors to the RNA standards. As desired the standards lost the trend present in relative normalization but retained their sample noise (Fig 1B, Fig S2A).

ERCC spike-ins are manufactured to be accurate up to $\pm 5\%$ on a $\log_2$ scale (Ambion technical service, personal communication), all measures of accuracy are therefore performed on a $\log$ scale.

We evaluate the consistency and accuracy of our absolute normalization by comparison of calculated to actual RNA standard transcript copy number abundances. We time average the abundance of all RNA standards, then we compare the three series individually and a normalization involving Clutch A/B PolyA+ combined (Fig S2C). The normalization shows excellent performance with an $R^2 = 0.97 - 0.98$ error independent of expression level. Moreover, calibration of ArrayControl standards by ERCC standards (each standard pipetted independently into each embryo Trizol homogenate) alone reveals of 1.11-1.25 fold inaccuracy.

There is a spike-in species specific bias which is reproducible over replicates and partially independent of sequencing protocol. This can most readily be seen in the residuals of absolute normalization model validations (Fig S2C), these residuals are highly correlated in Clutch A and Clutch B polyA+ sequencing (Pearson correlation coefficient 0.86), and are still correlated to a lesser degree between Clutch A polyA+ and rdRNA (Pearson correlation coefficient 0.59). We conclude that the spike-in polyA+ bias varies between spike-in species and may be the result of a variable polyA tail length (indeed, this appears to be the origin of the poor performance of ERCC-00116 (Fig S1C)). Further, there is either a sequencing bias that affects both polyA+ and rdRNA sequencing, or there is a consistent deviation in manufacture of ERCC spike-ins from their stated datasheet abundances.

We use the variation between Clutch A and Clutch B with the spike-in species specific bias to quantify the uncertainty in our absolute normalization. Here, a generative model of the error based on distribution of the $\bar{\beta}_j$ would be most ideal, however, we do not have the necessary replicates of the rdRNA in both clutches to derive confidence intervals for rdRNA data that accounts for between clutch batch effects. We opt to estimate empirical uncertainty for true transcript numbers based on the residuals of the spikes in ClutchA/B polyA+ sequencing. We apply this uncertainty to all polyA+ and rdRNA datasets.

If s_q is the true number of transcripts for a given RNA standard and $\hat{s}_q$ is our calculated absolute abundance, we find a simple log model summarizes the residuals in our absolute normalization well. That is, $\log s_q = \log \hat{s}_q + \epsilon$, where ϵ is zero mean Gaussian random variable with standard deviation $\sigma_s = 0.59$ (or 0.25 on a $\log_{10}$ scale Fig S2C). Given absolute normalized estimates for a transcript's abundance $\hat{\mathbf{m}}$ measured at times $\mathbf{t}$ and a set of hyperparameters $\hat{\boldsymbol{\theta}}$ chosen by maximizing the marginal likelihood, the distribution of the true transcript number $m_\star$ at time $t_\star$ is:

$$p\left(m_\star \middle| \sigma_s, t_\star, \hat{\mathbf{m}}, \mathbf{t}, \hat{\boldsymbol{\theta}}\right) = \int p(m_\star | \log h^{-1}(y_\star), \sigma_s) p\left(y_\star \middle| t_\star, h(\hat{\mathbf{m}}), \mathbf{t}, \hat{\boldsymbol{\theta}}\right) dy_\star.$$

Where $p\left(y_\star \middle| t_\star, h(\hat{\mathbf{m}}), \mathbf{t}, \hat{\boldsymbol{\theta}}\right)$ is the Gaussian Process predictive distribution on data transformed by h and $p(m_\star | \log h^{-1}(y_\star), \sigma_s)$ is a log-normal distribution with mean $\log h^{-1}(y_\star)$ and standard deviation σ_s. We evaluate this distribution by Monte Carlo methods. This uncertainty is largely a gene independent, time independent phenomenon, and so we do not report it as standard as it obscures gene dynamics and gene-gene comparisons. As example, we report the uncertainty distributions quantiles for *eef1a1o* (Fig S2D) and we use the distribution to average over the uncertainty of our absolute normalization when calculating the concentration of mRNA in ng / embryo (Fig 1C).

We now explore three factors that confound the absolute normalization of RNA-seq data: 1) library amplification artefacts, 2) gene model correctness and 3) ribosomal-depletion sequencing bias.

5.4.1 Library Amplification Artefacts

We explore the influence of library amplification artefacts on our data and subsequent absolute normalization. We evaluate the occurrence of RNA-seq fragments containing the same start and end coordinates, for all genes and spikes. These fragments will be both genuine and artefactual, that is produced both by different mRNA molecules and PCR duplication. We cannot distinguish genuine from artefactual duplicates, but we find an equivalent amount of duplication in spikes as endogenous genes:

We examine transcript length against duplicated read rate (duplicated reads mapping to a transcript / total reads mapping to a transcript), to ensure that we have a clear picture of duplication we restrict our concern to the most abundant transcript of each gene with at least 5,000 reads. We find a linear relationship between log transcript length and log duplicated read rate. For example, for Clutch A polyA+ at 6.0hpf (a representative but more deeply sequenced sample with 46 million reads), we find a Pearson correlation -0.73 between transcript length and duplicated read rate. We fit a linear model: $\log_{10} r = -0.75 \log_{10} l + 1.73$, where r is the duplication rate and l is the transcript length. Examining the residuals of this model a two sample Kolmogorov-Smirnov test fails to find any differences between the residuals of the spikes and genes (p = 0.45). Extending this analysis to all samples in Clutch A polyA+, we do not find any differences between the length corrected duplication rate of the spikes compared to genes. Thus, in terms of read duplication rates we find both spikes and genes have a similar relationship between molecule numbers and read counts. Therefore, any library amplification artefacts present will be accounted for in normalization, and so will have a limited undue influence on relative or absolute transcript number predictions.

5.4.2 Gene Model Correctness

In this work we do not assemble or correct the *X.tropicalis* v7.2 models based on our RNA-seq data. Incorrect gene models can lead to incorrect estimates of absolute abundance. In instances in which an exon is missing from either the 5' or the 3' end of a transcript model we will estimate the correct transcript copy number, but will underestimate the absolute abundance in kb/embryo. Conversely, instances in which the 5' or 3' end of a transcript model are unduly extended we will incorrectly estimate the transcript copy number, but will correctly estimate the abundance in kb/embryo.

We quantify this latter issue. For the most abundant isoform of each gene we evaluate the density of unique reads aligning to the entire transcript versus the UTR. In Table S2, we report the 725 transcripts for which these densities differ by at least 2-fold. We calculate a predicted fold change error of the absolute normalization due to this discrepancy in densities. The is error is calculated by estimating the degree to which the UTR needs to be shortened to produce an equal read density over the UTR and coding regions. If a transcript model has problems beyond or different to an extended UTR, the transcript will often be reported within the list, but the fold change prediction will be incorrect. As such the absolute copy numbers of any transcript in this list should be treated with caution. If we combine the errors for all 725 transcripts discussed above, this produces a 4.7% error in the total number of transcripts present in the embryo (or 1.047 fold error), and this effect is considerably less than the uncertainty of our absolute normalization.

5.4.3 rdRNA Depletion of non-ribosomal-RNA transcripts

Certain transcripts are under-represented in ribosomal depleted sequencing over polyA+ sequencing, for example *mixer* and *foxh1.2* (Fig S2F lower panels). We find a severe reduction (a greater than 5-fold decrease in rdRNA sequencing compared in polyA+) in 2.2% of all genes, this corresponds to a loss of 1.9% of all transcripts. Further, we find that only 12% of all transcripts experience a decrease greater than 1.1 fold. This is within the error of our absolute normalization and so we do not correct further for rdRNA dropout. We provide a list of all those genes affected (Table S2). We hypothesize that this effect may be due to RiboZero probes matching regions within these mRNAs and note that a sequence based regression model would be a fruitful way to account for and measure this bias.

6 Gaussian Process Model Validity

In our standard model of gene expression we opted for the pragmatic assumptions that square root transformed transcripts per embryo abundances are noisy observations from a stationary Gaussian Process with Gaussian likelihood. Visual inspection suggests the assumptions and Gaussian Process construction offer a good representation of the data.

To validate our assumptions, we calculate the sum of normalized residuals of the transformed expression of each gene over all samples, $w_i = \sum_j^n \left(y_{ij} - \mu_{ij}\right)^2 / \sigma_{ij}^2$, with $\mu_{ij}$and $\sigma_{ij}$are the Gaussian Process mean and variance for gene i in sample j. Note, these are no longer the leave-one-out mean and variance as before, but the mean and variance calculated on the full data. If our assumptions are correct, then w_i will follow a χ^2 distribution with n degrees of freedom. Taking this as the null hypothesis, accounting for samples in which a gene has zero expression we reject the null hypothesis with p < 0.01 for 8.8% of genes in Clutch A polyA+, 0.6% of genes in Clutch A rdRNA and 1.2% of genes in Clutch B PolyA+. For our shorter 0-24 hpf data we reject the null approximately the expected number of times. Clearly, failing to reject the null hypothesis does not imply that the data follow our assumed distribution, but we use it as an indicator that our assumptions are reasonable.

For Clutch A PolyA+ we note that 98% of the cases in which we reject the null hypothesis have a characteristic timescale of less than 15 hours. Inspection reveals that it is the stationary assumption at fault, and a non-stationary covariance function with a time varying timescale or sample noise would be more suitable to describe the data. In the majority of cases gene expression appears more tightly controlled in early development. The covariance is best characterized by a short timescale and low sample noise during very early development (0-24 hpf). Then it is better characterized by a lengthening timescale and increasing sample noise throughout the remainder of development. In these cases our maximization of the log marginal likelihood most often to select hyperparameters that match the early phase where we have most dense sampling and the largest amount of information on a gene's expression.

Clearly, the origin of this non-stationarity could be explained by a gene executing different programs during development. However, we rarely observe the converse in which a gene's expression becomes more tightly controlled with developmental time. This suggests two confounding factors. First, the embryos in the clutch become more de-synchronized with time. Second and related, as development progresses the cells within the embryo lose the synchronicity they display during the cleavage stages as they execute increasingly complex programs of development. Our observed data becomes the aggregate of different processes, and our ability to detect fast timescale processes with clarity diminishes and this contributes to non-stationary behavior in a gene's expression.

As our stationary and data transformation assumptions perform well in the shorter time series and those genes with non-stationary behavior are well described early in development in the longer series, we opt to not employ the non-stationary Gibbs covariance function for these genes.

7 Detection Limits and mRNA content of the Embryo

7.1 Spike in Detection Limit

We evaluate our detection limit by calculating the average number of transcripts required to produce a single read. We do so by first calculating the detection limit in transcript abundance in kb for each sample from the three series (Fig S2E). We find the detection limit increases, as expected, with developmental time as the RNA levels in the embryo increase (see below. We average over a Gaussian Process model of these per sample detection limits, along with averaging over all transcript sizes to calculate the mean detection limit in transcripts/embryo with time (Fig S2D). If we further average over all samples and data sets we find a mean detection limit of ≈ 1300 transcripts/embryo for the entire data.

7.2 Total mRNA per Embryo

We calculate the total measured mRNA (PolyA+ and rdRNA) in the embryo in nanograms (Fig 1C). To evaluate the per-sample variation of this normalization, the data points in these figures are estimates of total mRNA calculated from non-smooth correction factors $e^{-\beta_j}$ (rather than Gaussian Process smoothed $e^{-\bar{\beta}_j}$). We average over the uncertainty in our absolute normalization, convert abundances to kilobases and then to nanograms, and sum up the abundance for all measured transcripts excluding RNA standards. For visualization, we smooth the resulting ng/embryo estimates with a Gaussian Process using a non-stationary Gibbs covariance function and time-varying noise covariance function.

We compare and validate our *X. tropicalis* polyA+ mRNA ng/embryo abundances to experimentally measured *X. laevis* polyA+ mRNA yields (Sagata et al., 1980). We first map the *X. laevis* abundances that are given by developmental stage onto *X. tropicalis* stages based on our table of stages and timings (Table S1). The *X. laevis* polyA+ mRNA mass is greater than that of the *X. tropicalis*. A reduction of the Sagata et al. *X. laevis* mRNA recoveries by a factor of 3.31 +/- 0.76 [mean +/- std. dev.] maximizes similarity to *X. tropicalis* (Fig 1C). This is in remarkable agreement with the ratio of volumes of *X. laevis* and *X. tropicalis* egg sizes. In (Crowder et al., 2015) they find *X. laevis* diameters to be 1.19 +/- 0.07 mm [mean +/- std. dev., n = 17], and *X. tropicalis* egg diameters to be 0.80 +/- 0.05 mm [mean +/- std. dev., n = 13], and so the volume ratio is 3.31 +/- 0.76. We observe good agreement between *X. laevis* and *X. tropicalis* in the wave of polyadenylated RNA that increases to N&F stage 8/9 (*X.t.* 4.5 hpf), then decreases to stage 12.5 (*X.t.* 10 hpf), then increases again through early development.

8 Differential Expression in time series

8.1 Clutch A vs Clutch B Differential Expression

To determine which genes exhibit temporal differential expression, we use our Gaussian Process framework. We test the hypothesis that a gene in Clutch A and Clutch B has equal expression in time, versus the hypothesis that the gene has different expression in Clutch A and Clutch B in time. The latter hypothesis describes differential expression. We ask whether a single Gaussian process model best describes the data, or whether two models (one for Clutch A and one for Clutch B) best describe the data. We restrict our consideration to those timepoints at which Clutch A samples and Clutch B samples overlap, and only those genes with a mean abundance greater than 3,000 transcripts per embryo over all Clutch A and Clutch B samples.

We have Clutch A abundance measurements $\widehat{\mathbf{m}}_A = h^{-1}(\mathbf{y}_A)$ measured at times $\mathbf{t}_A$ and similarly $\widehat{\mathbf{m}}_B = h^{-1}(\mathbf{y}_B)$ at $\mathbf{t}_B$ for Clutch B, and the combined data with length m is $\mathbf{y} = [\mathbf{y}_A\ \mathbf{y}_B]$ and $\mathbf{t} = [\mathbf{t}_A\ \mathbf{t}_B]$. We look for differential expression by comparing the log marginal likelihoods of the two hypotheses. To account for the difference in the number of parameters we use the Bayesian Information Criterion (BIC):

$$\mathrm{BIC}(\mathbf{y}_A, \mathbf{y}_B) = \mathcal{L}(\mathbf{y}_A, \mathbf{t}_A, \boldsymbol{\theta}_A) + \mathcal{L}(\mathbf{y}_B, \mathbf{t}_B, \boldsymbol{\theta}_B) - \mathcal{L}(\mathbf{y}, \mathbf{t}, \boldsymbol{\theta}) - 3\log[m]/2.$$

A positive value of $\mathrm{BIC}(\mathbf{y}_A, \mathbf{y}_B)$ would indicate a preference for the hypothesis that Clutch A and Clutch B have differential expression. We report the distribution of $\mathrm{BIC}(\mathbf{y}_A, \mathbf{y}_B)$ for all genes, along with examples of genes with $\mathrm{BIC}(\mathbf{y}_A, \mathbf{y}_B) > 0$ (Fig 2B,C). We wish to focus on those genes that are differentially expressed with the largest statistical effect. We calculate the time dependent overlap between the models:

$$\alpha_t = \int_{-\infty}^{\infty} \min[p(y_t|\mathbf{y}_A, \mathbf{t}_A, \boldsymbol{\theta}_A), p(y_t|\mathbf{y}_B, \mathbf{t}_B, \boldsymbol{\theta}_B)]\, dy_t.$$

If a gene has greater expression in Clutch A over Clutch B, and an overlap of $\bar{\alpha} = 0.025$, then the upper 95% confidence interval of B would be equal to the lower 95% confidence of A on average. We plot $\mathrm{BIC}(\mathbf{y}_A, \mathbf{y}_B)$ against mean overlap over time for all genes (Fig 2A). We find that approximately all genes with $\mathrm{BIC}(\mathbf{y}_A, \mathbf{y}_B) > 60$ have $\bar{\alpha} < 0.1$, and all genes with $\bar{\alpha} < 0.025$ have $\mathrm{BIC}(\mathbf{y}_A, \mathbf{y}_B) > 60$. The mean overlap dramatically decreases for all genes with $\mathrm{BIC}(\mathbf{y}_A, \mathbf{y}_B) > 60$. We divide genes into three categories:

1. No differential expression: $\mathrm{BIC}(\mathbf{y}_A, \mathbf{y}_B) \leq 0$, 12,062 genes.
2. Differential Expression Small effect: $0 < \mathrm{BIC}(\mathbf{y}_A, \mathbf{y}_B) \leq 60$, 4,561 genes. These genes are statistically distinguishable, but have minor differences.
3. Differential expression Large effect: $\mathrm{BIC}(\mathbf{y}_A, \mathbf{y}_B) > 60$, 291 genes. These genes are strongly differentially expressed.

We calculate gene ontology enrichment for strongly differentially expressed 291 genes in the large effect category. All gene ontology terms over represented with $p < 0.05$ by χ^2 test are reported in Table S4 along with Benjamini Hochberg FDR values and Fisher exact p-values. Interestingly, there is a strong enrichment for GO terms related to cell division.

8.2 Isoform Differential Temporal Dynamics

We now consider differential isoform expression; our primary interest is finding the most dramatic examples of differential isoform expression. We use a procedure very similar to that used in Clutch A vs Clutch B differential expression; however, we now distinguish between *differential abundance* and *differential temporal dynamics*. We say that two isoforms whose expression ratio is constant with time have differential abundance. These isoforms have the same temporal pattern of expression but they are expressed at different levels. Conversely, two isoforms whose expression ratio varies with time have differential temporal dynamics. Examples of differential abundance and differential dynamics are given in Fig S5.

To detect differential dynamics, we consider differential expression on scale normalized data. If a gene has $k = 1, \ldots, P$ isoforms and if isoform k has abundance estimates $\widehat{\mathbf{m}}_k = h^{-1}(\mathbf{y}_k)$, then we look for differential expression between normalized $\mathbf{z}_k = (\mathbf{y}_k - \mu_{\mathbf{y}_k})/\sigma_{\mathbf{y}_k}$ where $\mu_{\mathbf{y}_k}$ and $\sigma_{\mathbf{y}_k}$ are the mean and standard deviation of vector $\mathbf{y}_k$ respectively. We use multiple conditions to detect differential isoform dynamics. First we evaluate the Bayesian Information Criterion to decide if we prefer a single model for all isoforms or a different model for each individual isoform:

$$\mathrm{BIC}(\mathbf{z}) = \left(\sum_{k}^{P} \mathcal{L}(\mathbf{z}_k, \mathbf{t}, \boldsymbol{\theta}_k)\right) - \mathcal{L}(\mathbf{z}, \mathbf{t}, \boldsymbol{\theta}) - 3(P-1)\log[mP]/2$$

With this condition any two isoforms whose expression is linearly related will have $\mathrm{BIC}(\mathbf{z}) \leq 0$ and so will have differential or identical abundance. Similar to Clutch A vs Clutch B differential expression, we focus our attention on those with $\mathrm{BIC}(\mathbf{z}) > 60$. If this first condition is satisfied, we next consider whether any of the pairs of isoforms of a gene have differential abundance rather than differential dynamics. Consider the graph with nodes for each isoform and edges between those isoforms which have $\mathrm{BIC}(\mathbf{z}_k, \mathbf{z}_l) < 60$. Then, consider the transitive closure of this graph. We say that any connected pair of nodes has equal or differential abundance, and any disconnected pair of nodes has differential dynamics. This procedure may connect two isoforms $\mathbf{z}_k$ and $\mathbf{z}_l$ which have $\mathrm{BIC}(\mathbf{z}_k, \mathbf{z}_l) > 60$, and it may result in a fully connected graph. In these cases the differential dynamics exhibited is not strong enough for consideration here. This condition will aid to remove those genes with a large number of annotated isoforms and few disambiguating reads.

Disconnected sub-graphs constitute groups of isoforms with equal or differential abundance, see Fig S5 for example. For our final condition, we return to the original reads and require unique reads disambiguate each group of isoforms.

9 Temporal Synexpression

9.1 Temporal Map of the Transcriptome

We generated a temporal map of the transcriptome (Fig 4A), we included all genes with a mean expression of at least 3,000 transcripts over the 66 hours of Clutch A PolyA+. To visualize the broad behaviors of the transcriptome, we took the transformed Gaussian Process medians from absolute normalization using the non-stationary Gibbs and noise covariance functions (Fig S2B, far right) and then normalized each trace by its maximum. Our intention is to project all variations in gene expression onto a single dimension, and to do so we performed a hierarchical clustering and found and ordered the resulting dendrogram so that the leaves were most similar (Bar-Joseph et al., 2001). The resultant ordering of all genes is more important than the clustering itself.

While exploring temporal synexpression, we noticed a number of genes with a similar expression pattern to *ckb*, this lead us to identify temporal synexpression groups **S1** and **S2** (Fig 4A,B). To expand on these we take the z-normalized expression between 34 and 66 hpf and take the 150 genes most similar to *ckb* by Euclidean distance (Table S5). Similarly, we identified **V1** and **V2**, by ranking by similarity to *rho* over 40-66 hpf (Table S5).

9.2 Gene Ontology Annotation and Sliding Window Overrepresentation

We use a previously published *X. tropicalis* gene ontology (GO) annotation generated by blast2go (Collart et al., 2014). To augment this GO annotation we include a previously published list of transcription factors (Collart et al., 2014), we treat this list in the same manner as any GO category when calculating overrepresentations.

In temporal synexpression (Fig 4) and characteristic timescale analysis (Fig 7) we look for locally overrepresented categories in ordered lists of genes. To calculate this overrepresentation we slide a window of length 500 (Fig 4) and 600 (Fig 7) along a list of genes (Fig S6), at each window position we perform a χ^2 tested for an association between the genes in the window and each GO category. We calculate $-\log(p)$ of the χ^2 test at each position in the list, and smooth this to obtain the trend (Fig S6). We look for regions in which $-\log(p) > 5$ or equivalently $p < 0.0067$. From these regions we retain and report those that contain at least 5 genes of the given category, have their maximum $-\log(p) > 10$ and the χ^2 test for the region taken as a whole has $p < 0.05$.

All GO overrepresentations are reported in Table S4.

10 Transcript Kinetics Per Allele

We calculate the rate of transcript accumulation per allele during the cleavage stages. We do this for *pri-mir427* (Fig 6), and *nodal3, nodal5, nodal6, sia1, sia2, klf17, admp, ventx3.1* (Fig S7).

We use an idealized model of cell numbers in which the first cell division occurs at 1 hpf and divisions occur every 20 minutes thereafter until the 12th division. We use the following continuous approximation for cell numbers:

$$c(t) = \begin{cases} 1 & \text{for } t \leq 1 \\ 2^{3(t-1)+1} & \text{for } 1 < t \leq 4\,{}^{2}/_{3}, \end{cases}$$

t gives the time in hpf. During the growth phase of cells the rate of change of cells is $\dot{c}(t) = \nu c(t)$, with growth rate $\nu = 3\log 2$. We consider a transcript whose Gaussian Process abundance in kb at time t is given by $g(t)$. We are interested in the transcript abundance per cell and the rate of production by each cell. The cell transcript abundance is given by $\eta(t) = g(t)/c(t)$, changes in $\eta(t)$ will be caused by production of transcript by each cell, degradation and dilution amongst the cells, and so:

$$\dot{\eta}(t) = p(t) - \nu\eta(t),$$

where $p(t)$ is the combination of production and degradation rates. By definition of $\eta(t)$:

$$\dot{\eta}(t) = \frac{\dot{g}(t)}{c(t)} - \frac{g(t)\dot{c}(t)}{c(t)^2} = \frac{\dot{g}(t)}{c(t)} - \nu\eta(t) \qquad \text{for } 1 < t < 4\,{}^{2}/_{3},$$

and as a consequence $p(t) = \dot{g}(t)/c(t)$ and the combined production and degradation rate per allele is $p(t)/2$. During the cleavage stages, we assume degradation is negligible, and $p(t)/2$ will be approximately equal to the production rate of transcript per allele.

In Fig 6 we report the rate of accumulation for *pri-mir427*, the 95% confidence intervals reported are those that arise from differentiating the *pri-mir427* Gaussian process, that is, they do not include uncertainty due to the absolute normalization. We measure the maximum accumulation rate of all *pri-mir427* transcript at 2560 (95% CI: 2280 to 2840) kb/min/allele. There is a gap in the genome assembly present at the center of the locus (Fig 6D). This gap will not unduly influence this maximum accumulation rate as long as the gap contains sequence exhibiting the same repetitive structure (i.e. the mir427 hairpin and inter-hairpin repeat) as the remainder of the locus, as we will still detect all reads generated by the *mir427* locus.

We provide context to the maximum accumulation rate of *pri-mir427* by calculating the theoretical accumulation rate assuming maximal RNA pol II transcription. RNA pol II elongation has been measured at 4.3 kb/min (Ardehali and Lis, 2009), the maximum loading RNA pol II loading density observed in amphibian oocyte lampbrush chromosome is ~10 pol II/kb (Miller and Hamkalo, 1972). Assuming RNA pol II transcription is uniform over the *mir427* locus the theoretical maximal rate of accumulation of *pri-mir427* is,

$$4.3\ \frac{\text{kb}}{\text{min pol II}} \times 55\text{kb} \times 10\ \frac{\text{pol II}}{\text{kb}} = 2365\ \frac{\text{kb}}{\text{min}}.$$

This theoretical rate compares well with the 2560 (95% CI: 2280 to 2840) kb/min/allele we measure, and implies that *pri-mir427* is transcribed maximally during the early rapid cell

divisions of the frog embryo. The above theoretical rate calculation does not include transcription from the gap in the genome assembly in the *mir427* locus, and therefore a RNA pol II elongation rate of 4.3 kb/min and a loading density of 10 pol II/kb constitute an upper bound to the RNA pol II activity required.

The above assumes that the transcript is ubiquitously transcribed. If a transcript is transcribed in a fraction of cells that undergo exponential increase during the cleavage stages, then the above will underestimate the rate proportional to the fraction of transcribing cells. We can be confident that *pri-mir427* is ubiquitously transcribed by *in situ* (Fig 6) and that the pol II elongation rates and loading densities required to explain the observed accumulation rates are maximal. However, we cannot make any such assumption for genes whose rates are given in Fig S7. For these genes, these rates are underestimates of the true accumulation per allele. Nevertheless, these rates can be used to derive bounds and estimates of relevant biological information. For example, the peak accumulation rate of *nodal6* is 1.8 (95% CI: 1.3 to 2.2) kb/min/allele averaged over all allelic copies in the embryo (Fig S7), and is transcribed from a 1.6kb locus. We can derive a lower bound for the number of cells transcribing *nodal6* by assuming a maximal RNA pol II output similar to *pri-mir427* (Fig 6). If RNA pol II loads at ~10 pol II/kb and elongates at ~4kb/min/pol II, then a cell is capable of producing *nodal6* transcript at ~64kb/min. This implies that at least 2.0% - 3.4% of the embryo must be transcribing *nodal6*. Conversely, we can run the argument in the other direction. Literature suggests that approximately 10% of the cells of the Stage 8-9 embryo transcribe *nodal6* (Skirkanich et al., 2011). This suggests that the true peak accumulation rate of *nodal6* is 13-22 kb/min/allele, which could be achieved by 3-5 RNA pol IIs elongating at 4 kb/min at each active *nodal6* locus.

11 *in situ* Hybridization

RNA *in-situ* hybridization was done as previously described (Khokha et al., 2002). We generated probes from cDNA clones available in either the Sanger full-length (Gilchrist et al., 2004) clone collection or the IMAGE clone repositories. We generated digoxigenin-labelled antisense probes using *in vitro* transcription according to the manufacturer's instructions (T7 High Yield Kit, New England Biolabs). In the case of *pri-mir427* the clone was cut with XhoI-SacI, blunted and re-ligated in order to remove a long polyA tail. All cDNA clones used to make probes given in Table S6.

We generated isoform specific probes from cloned PCR products using a TOPO® TA Cloning® Kit, Dual Promoter (Life Technologies). We used either SP6 or T7 RNA polymerases to generate RNA probes depending on insert orientation. For isoform specific *in situs*, we also used sense strand probes side-by-side and only report antisense signals that are clearly expressed above the sense control. All primer sequences for the isoform specific probes are given in Table S7.

SUPPLEMENTAL REFERENCES

Bar-Joseph, Z., Gifford, D.K., and Jaakkola, T.S. (2001). Fast optimal leaf ordering for hierarchical clustering. Bioinformatics *17 Suppl 1*, S22-29.

Gilchrist, M.J., Zorn, A.M., Voigt, J., Smith, J.C., Papalopulu, N., and Amaya, E. (2004). Defining a large set of full-length clones from a Xenopus tropicalis EST project. Developmental biology *271*, 498-516.

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.*, et al.* (2002). Techniques and probes for the study of Xenopus tropicalis development. Developmental dynamics : an official publication of the American Association of Anatomists *225*, 499-510.

MATLAB (2014) (Natick, Massachusetts: The MathWorks Inc.).

Miller, O.L., Jr., and Hamkalo, B.A. (1972). Visualization of RNA synthesis on chromosomes. International review of cytology *33*, 1-25.

Stein, M.L. (1991). Interpolation of Spatial Data (New York: Springer-Verlad).