Single molecule imaging and modelling of mRNA decay dynamics in the *Drosophila* embryo

Lauren Forbes Beadle[1,#], Jennifer C. Love[1,#], Yuliya Shapovalova[1,3], Artem Artemev[2], Magnus Rattray[1]* and Hilary L. Ashe[1]*

Affiliations:

[1]Faculty of Biology, Medicine and Health, University of Manchester, Manchester, M13 9PT, UK.

[2]Department of Computing, Imperial College London, London, SW7 2AZ, UK.

[3]Present address: Radboud University, Postbus 9010, 6500 GL, Nijmegen, The Netherlands

[#]Joint contribution

*Correspondence: magnus.rattray@manchester.ac.uk, hilary.ashe@manchester.ac.uk

Short title: mRNA dynamics and degradation in the *Drosophila* embryo

Abstract

Regulation of mRNA degradation is critical for a diverse array of cellular processes and developmental cell fate decisions. Many methods for determining mRNA half-lives rely on transcriptional inhibition or metabolic labelling. Here we use a non-invasive method for estimating half-lives for hundreds of mRNAs in the early *Drosophila* embryo. This approach uses the intronic and exonic reads from a total RNA-seq time series and Gaussian process regression to model the dynamics of premature and mature mRNAs. We show how regulation of mRNA stability is used to establish a range of mature mRNA dynamics during embryogenesis, despite shared transcription profiles. Using single molecule imaging we provide evidence that, for the mRNAs tested, there is a correlation between short half-life and mRNA association with P-bodies. Moreover, we detect an enrichment of mRNA 3' ends in P-bodies in the early embryo, consistent with 5' to 3' degradation occurring in P-bodies for at least a subset of mRNAs. We discuss our findings in relation to recently published data suggesting that the primary function of P-bodies in other biological contexts is mRNA storage.

Introduction

Cells establish their identity by changing their gene expression patterns in response to different signals and environments. Critical to this is the ability of a cell to modulate mRNA levels. mRNA abundance depends not only on the transcription rate but also on mRNA stability. In eukaryotic cells, there are two major pathways of mRNA degradation: Xrn1 endonuclease-mediated 5'-3' decay and exosome catalysed 3'-5' decay [1,2]. Many mRNA degradation factors and mRNAs can become condensed into processing bodies (P-bodies), which are phase separated compartments in the cytoplasm implicated in mRNA storage and decay [3,4]. mRNA stability is also commonly regulated by sequences in the 3' UTR, including binding sites for RNA binding proteins or miRNAs [5]. While the two major decay pathways are responsible for general turnover of cytoplasmic mRNAs, there are also mRNA surveillance pathways that degrade aberrant mRNAs. These include mRNAs carrying a premature stop codon, lacking a stop codon, or mRNAs with paused ribosomes [6].

Regulation of mRNA degradation is essential for diverse cellular processes including proliferation, differentiation, apoptosis and immune responses [1,7–10]. Control of mRNA stability is also important for cellular decisions and behaviour during development. For example, regulation of *myc* mRNA stability fine-tunes the proliferation rate of neuroblasts in the *Drosophila* larval brain [11], an *fgf8* mRNA gradient generated by mRNA decay couples differentiation to posterior elongation of the vertebrate embryonic axis [12] and *Hes1* mRNA instability is integral to the Hes1 protein ultradian oscillations that may act as a timer for vertebrate neuronal differentiation [13]. In addition, a conserved feature of early embryogenesis is that there is bulk degradation of maternal mRNAs around the time of

zygotic genome activation [14,15]. Consistent with the key roles of mRNA stability in cell biology, mutations in many components of the degradation pathways are associated with human diseases [9,16].

While the half-lives of strictly maternal mRNAs during embryogenesis can be readily measured genome-wide [17], measuring the decay of zygotic mRNAs is more difficult due to ongoing transcription. One approach is to inhibit transcription and then follow the decline in mRNA levels over time [18–20]. Other methods involve metabolic labelling of the RNA, for example in pulse-chase or approach-to-equilibrium experiments [19–21]. Related approaches use computational models to estimate transcription and degradation rates by sequencing both the total and labelled RNA following the pulse [19]. Single molecule fluorescent in situ hybridisation (smFISH) imaging based methods for estimating mRNA half-lives have also been described. However, these methods are not high throughput and require either steady-state transcript levels [22] or a natural shut off of transcription [23].

In this study we generate a high-resolution total RNA-seq time series across early embryogenesis that we use to estimate half-lives and assign mRNAs into different stability classes. Our data suggest that some mRNAs can be degraded in P-bodies, as the unstable mRNAs we have investigated are more highly colocalised with P-bodies and we can detect 3' mRNA fragments in P-bodies. Overall, our data reveal the contribution of mRNA stability to shaping mRNA levels during early embryogenesis and provide insight into how mRNA stability is regulated.

Results

Isolation of mRNA from early embryos captures high resolution transcriptional dynamics

To investigate mRNA accumulation dynamics during early *Drosophila* embryogenesis, we first generated a total RNA-seq time series. The early *Drosophila* embryo undergoes a series of 14 nuclear cycles within a common cytoplasm (nc1-14). RNA was isolated from single early *Drosophila* embryos at 10 time points, starting at nc11, ~ 90 minutes after egg lay (AEL) and prior to the onset of bulk zygotic transcription through to the beginning of gastrulation (Figure 1A). Embryos were collected from a His2avRFP transgenic line and precisely staged at nc11, nc12, nc13 and nc14 by calculating an internuclear distance ratio (Figure S1A). Single embryos were collected in triplicate 5 min after the nc11 and nc12 divisions, both 5 and 15 min after the nc13 division, then at 15 min intervals during the long nc14 interval, with the final time point corresponding to the appearance of the cephalic furrow (Figure 1A, Table 1). As male and female embryos have differences in X chromosome transcription due to dosage compensation [24], we used PCR to determine the sex of each embryo and select female embryos for analysis. We sequenced total RNA following rRNA depletion, rather than selecting for polyadenylated RNA, allowing us to capture intronic reads and other non-coding RNA species. The intronic reads allow quantification of nascent, unspliced transcripts and also detection of early zygotic

expression by distinguishing zygotic transcripts containing introns from maternally loaded spliced mRNAs.

We detected a total of 18159 transcripts during early embryogenesis representing 9026 unique genes. Using principal component analysis (PCA) we observed that the first two principal components represented 44% and 18% of the variation respectively and the replicates at each time point clustered together (Figure 1B). This suggests the biological age of the embryos explains the majority of variation within the data rather than differences between replicates, indicating the high quality of the libraries. Transcript levels across embryogenesis were visualised as a heatmap, with the transcripts ordered based on the time point of peak expression (Figure 1C). We classified 4897 early peaking transcripts (at the 95 or 105 min time points) as maternal, and 13262 transcripts peaking after 105 min as zygotic. Of the zygotic transcripts, 23% show peak expression early in nc13 or the start of nc14 (between 115 and 160 min inclusive) and the remainder show late peak expression after 160 min. Analysis of different dynamically expressed genes showed that our dataset included well characterised maternal (*nos* and *bcd*), maternal and zygotic (*Neu3* and *da*), early zygotic (*upd1* and *dpp*) and late zygotic mRNAs (*wg* and *hnt*) (Figure 1D).

As we sequenced total RNA, we determined the number of reads that mapped to introns as well as exons and transcripts. Analysis of the distribution of intronic reads shows an even read coverage across introns over all time points (Figure S2A). Only a very small proportion of transcripts at time points 105-125 min have intronic reads (Figure 1E), suggesting there is only minor zygotic transcription of intron-containing genes during these early stages. Previous studies have shown that the earliest zygotic activation of the *Drosophila* genome is biased towards expression of short intronless genes [25,26], which we cannot distinguish from maternally deposited transcripts at the early time points in our data. In addition, the early nuclear cycles are short, limiting the time period of active transcription. Nonetheless, eight genes have detectable levels of intron signal at nc12 and nc13A, suggesting early zygotic transcription (Figure S2C-D).

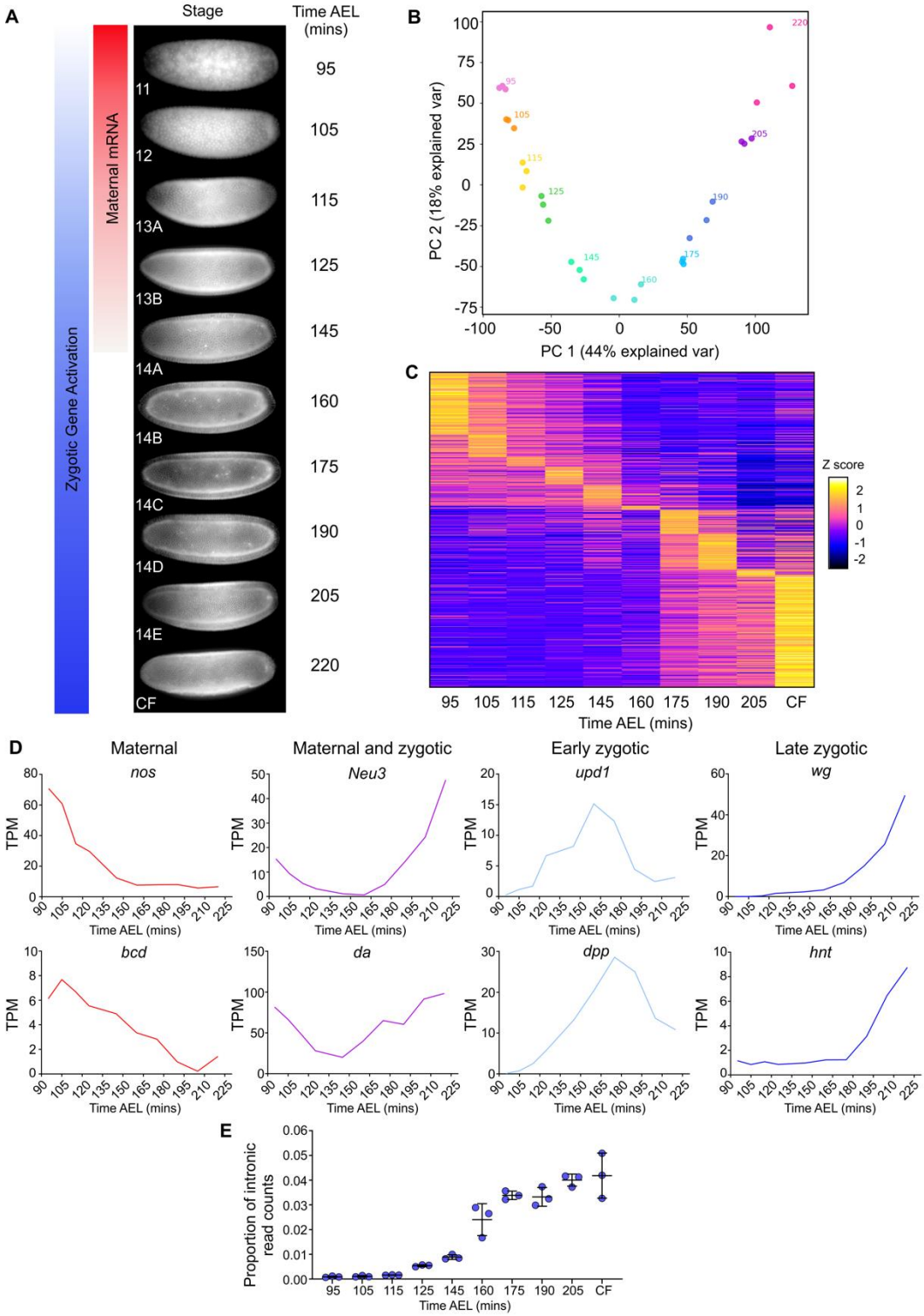


Figure 1. Total RNA-seq captures dynamic gene expression during early *Drosophila* development. (A) The time points used to collect RNA-seq samples, with approximate times after egg lay (AEL) at 20°C, are shown ranging from early nc11 through to cephalic furrow (CF) formation. The schematic highlights how the embryo switches from relying on maternally loaded RNAs (red) to activating its own zygotic transcription (blue). (B) Principal component plot of the RNA-seq samples shows tightly correlated replicates at each time point. (C) A heatmap of all expressed transcripts with

TPM >1 (18159 transcripts). Data are Z normalised and ordered along the y axis by the time point of peak expression. (D) Different examples of transcript dynamics captured by the data, as indicated by the graph titles. (E) Proportion of normalised intron reads (RPKM) throughout the time series.

The proportion of intronic reads increases significantly at 125 min (Figure S2D), then there is a further large increase around mid-nc14 (Figure 1E), when bulk activation of zygotic transcription occurs [24]. We detect 7276 zygotically expressed genes, similar to a previous estimate based on GRO-seq data [27]. The benefit of the high temporal resolution of our data can be seen in examples of transient gene expression, such as the gene *runt* (*run*) which is expressed at nc12 then peaks in early nc13 (Figure S2D). *run* has essential roles in patterning and transcriptional control of sex determination in early development, so the precise temporal regulation of its expression is likely to be important for these functions [28]. Additionally, we observe temporal changes in mRNA isoforms during development, exemplified by the genes *Meltrin* and *thickveins* (*tkv*) (Figure S3A and B). Their isoforms have altered coding sequences, which for the zygotically expressed isoform of the BMP receptor Tkv results in a shorter extracellular ligand-binding domain. We also detected expression of non-coding RNA species, such as those in the bithorax complex (Figure S3C). Overall, due to the high temporal resolution of our data and the ability to detect non-coding RNAs, we have a high quality dataset to investigate transcriptional dynamics in early *Drosophila* development.

Gaussian process regression provides estimates of transcript half-lives in early embryogenesis

As the degradation of maternal transcripts has been studied previously [17], we focused on the kinetics by which zygotic mRNAs are cleared in the early embryo. We used the intronic reads in our total RNA-seq dataset to represent pre-mRNA levels as a proxy for the transcription rate, while exonic reads reflect mature mRNA levels (Figure 2A). The intronic reads are correlated with NET-seq data from early embryos ($\rho = 0.46$, $p = 5.1 \times 10^{-11}$) (Figure S2B), consistent with the intronic signal reflecting transcription dynamics. We assume that introns are co-transcriptionally spliced and rapidly degraded. In support of this, NET-seq data suggests that >95% of splicing events are co-transcriptional in the *Drosophila* embryo [29] and modelling of metabolic labelling and sequencing data from *Drosophila* S2 cells revealed that the median half-life of introns is 2 min [30].

We used a Gaussian process (GP) regression model [31,32] of mRNA accumulation and degradation to estimate zygotic transcript half-lives from the intronic and transcript expression RNA-seq data (Figure 2B). Before fitting the GP regression, we applied a dynamic filter where we computed a log-likelihood ratio test between two GP regression models: a dynamic model with a radial basis function (RBF) kernel and a noise model to obtain genes that are differentially expressed. We then strictly filtered the dynamic data to select 593 mRNAs, which are purely zygotically transcribed and have very low reads at the first time point (<0.5 TPM). From these, we filtered further to select transcripts with a

correlation between the mRNA and pre-mRNA above 0.4. This is a mild correlation threshold that we applied to exclude transcripts for which pre-mRNA and mRNA are unrelated and it would therefore be hard to fit the model. As we filtered our data to select for zygotic genes we expect a positive correlation as gene introns and exons would increase together. The model uses a GP which specifies a prior distribution over possible underlying functions before observing the data. This non-parametric prior is governed by ordinary differential equations (ODEs), which describe the transcription regulation process. Once the data are observed, Bayesian inference is used to infer the posterior distribution. The posterior distribution allows quantifying uncertainty in the model as it reflects possible functions which can explain a given data set. Credible regions are derived from the posterior distribution to quantify the uncertainty at 95% confidence level. The ODE describing the system is shown in Figure 2B from which the splicing and degradation rates which are inferred using the GP regression. We assume that introns are spliced at the same rate for each mRNA, consistent with evidence from S2 cells that introns from the same mRNA tend to have similar splicing rates [30]. We normalised the read counts by the intron lengths so that the *S* parameter has comparable meaning for each mRNA, but allow it to differ from mRNA-to-mRNA to account for variation in splicing rates across genes [29] and also differences due to alternative splicing of different transcripts. For transcripts we used TPM units which normalises for transcript length.

The model provides half-life estimates for 263 zygotic transcripts corresponding to 186 genes (Supplementary Table 1). The distribution of these, coloured by short, medium or long half-life, can be seen in Figure 2C, with the mean half-life at 35 minutes and median at 16 minutes. Figure 2D shows examples of a gene with a short (Di) and a long (Dii) half-life, estimated using the GP model. Parameters were determined for these genes, along with associated uncertainty, using Markov chain Monte Carlo methods and the posterior distributions on the degradation rate *D* are displayed. *jim lovell* (*lov*) mRNA, a BMP target gene encoding a putative transcription factor [33,34], has a short half-life of 9 minutes whereas the *cv-2* mRNA, encoding a secreted protein that binds BMPs [35], has a longer half-life of 28 minutes. Full parameter estimates and credible regions are shown in Figure S4A.

As the dynamic embryonic mRNAs are not at steady-state, a previously described smFISH based method developed in human cells [22] was unsuitable for validation of half-lives. An alternative method exploited the arrest of transcription during mitosis to calculate the *snail* mRNA half-life in the *Drosophila* embryo, based on quantitation of mRNA numbers before and after mitosis [23]. However, we found the variation between transcript numbers in different embryos to be greater than any reduction that would be expected over such a short time frame (~ 4 mins) due to degradation (Figure S4B-E). As a result, any reduction due to degradation is masked by high variation between embryos, as has previously been observed for other mRNA numbers in the *Drosophila* embryo [36]. The *snail* mRNA numbers are tightly controlled by negative autoregulation [23], suggesting that *snail* may be uniquely suited to this method for calculating half-life.

In the absence of direct half-life validation, we determined whether the types of factors encoded by mRNAs with short and long half-lives have functions compatible with their inferred stabilities. Gene ontology (GO) analysis reveals that, compared to all dynamic transcripts in the RNA-seq data, those encoding transcription factors and cell adhesion proteins are enriched in the short (p-values 2.3 x 10^{-8} and 4.1 x 10^{-5}, respectively) but not long half-life mRNAs. Transcription factors have previously been reported to be encoded by unstable mRNAs [17,37,38]. Therefore, this approach has allowed the classification of transcripts into half-life categories where short half-life mRNAs are enriched for protein functions reflected by their stability.

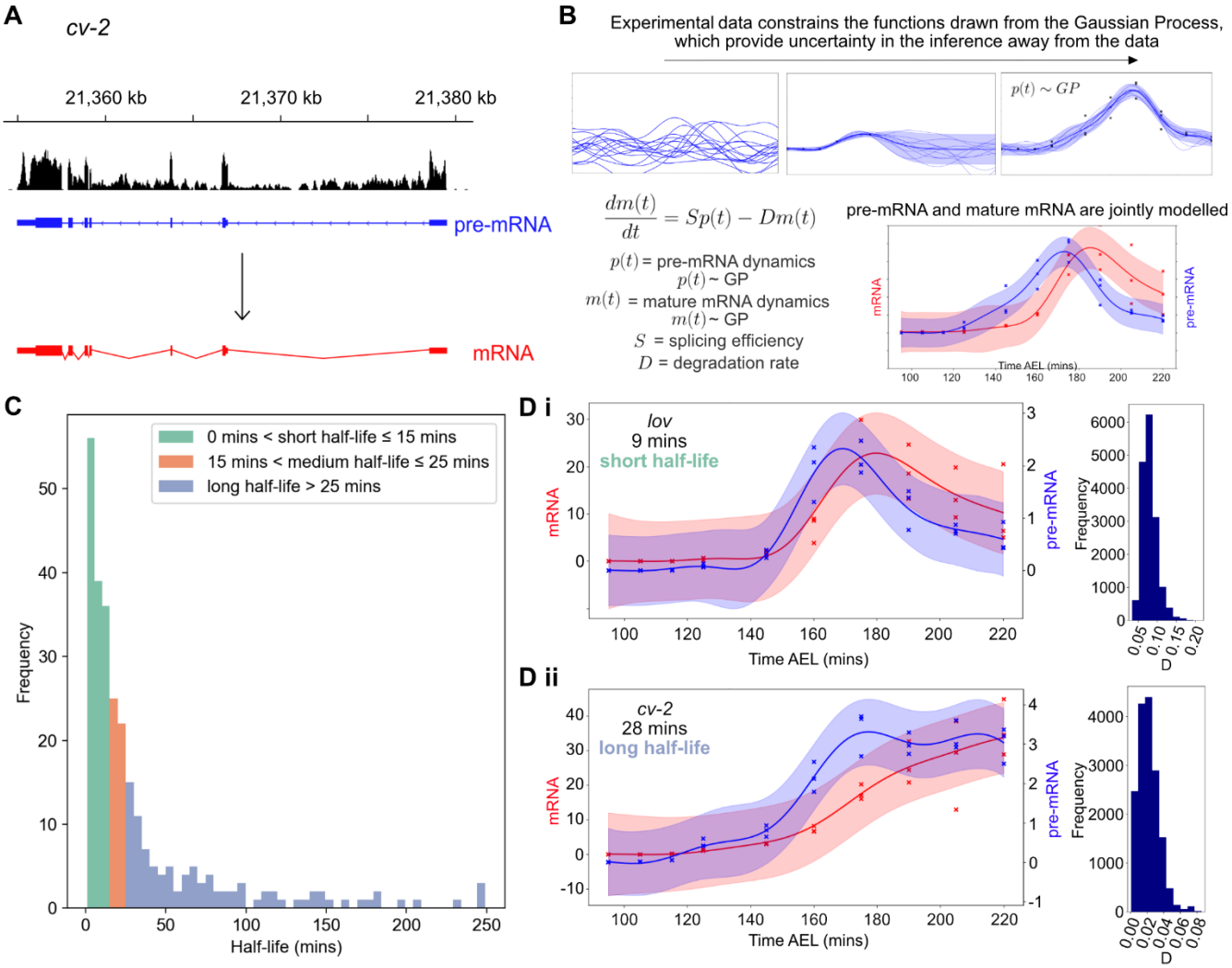


Figure 2: Gaussian process regression provides estimates of transcript half-lives in early embryogenesis. (A) Reads aligned to intronic and exonic regions of genes are used to represent pre-mRNA (blue) and mature mRNA levels (red), respectively. (B) Schematic of Gaussian Process regression and the ODE model which shows the evolution of the mature mRNA dynamics is described by the pre-mRNA data over time, sculpted by the splicing (*S*) and degradation (*D*) parameters. Pre-mRNA and mature mRNA are therefore jointly modelled using GPs related by the ODE. (C) Half-life

results for 263 transcripts estimated using the GP model. Transcripts are divided into short, medium and long half-lives and coloured accordingly. (D) Examples of data for a short (i) and a long half-life mRNA (ii) *lov* and *pxb*, fit using the GP model. Pre-mRNA is shown in blue and mature mRNA in red, shaded areas represent credible regions and crosses mark the data for each experimental replicate at each time point. Posterior distributions for the degradation parameter *D* for each gene are shown to the right.

Clustering reveals how degradation shapes mature mRNA dynamics

We next addressed how post-transcriptional regulation contributes to the range of mature mRNA dynamics seen in our data, by combining clustering analysis with our modelling of transcript half-lives. The pre-mRNA data were clustered using GPclust (Figure S5) [39]. From the intronic clusters, six highly populated intronic clusters that together exhibit a variety of interesting mRNA dynamics are shown in Figure 3Ai. The genes in each cluster share similar pre-mRNA profiles, and therefore transcription dynamics. All of the pre-mRNAs in intronic cluster 5 were then sub-clustered based on their mature-mRNA profiles (Figure 3Aii), which revealed that a range of mature mRNA dynamics arises from this single transcriptional profile. The zygotic mRNA subclusters for intronic cluster 2 also display a range of mature mRNA dynamics and are shown in Figure S6. The GP model sheds light on how these various dynamics arise, due to differences in the half-lives of transcripts in each cluster (Figure 3Aii, Figure S6). It is clear that the pattern in the shape of the time series is reflected in the different half-lives of the clusters; clusters which have a stronger peak have a shorter half-life and higher degradation rate, whereas those which continually increase across the time period have a long half-life and low degradation rate.

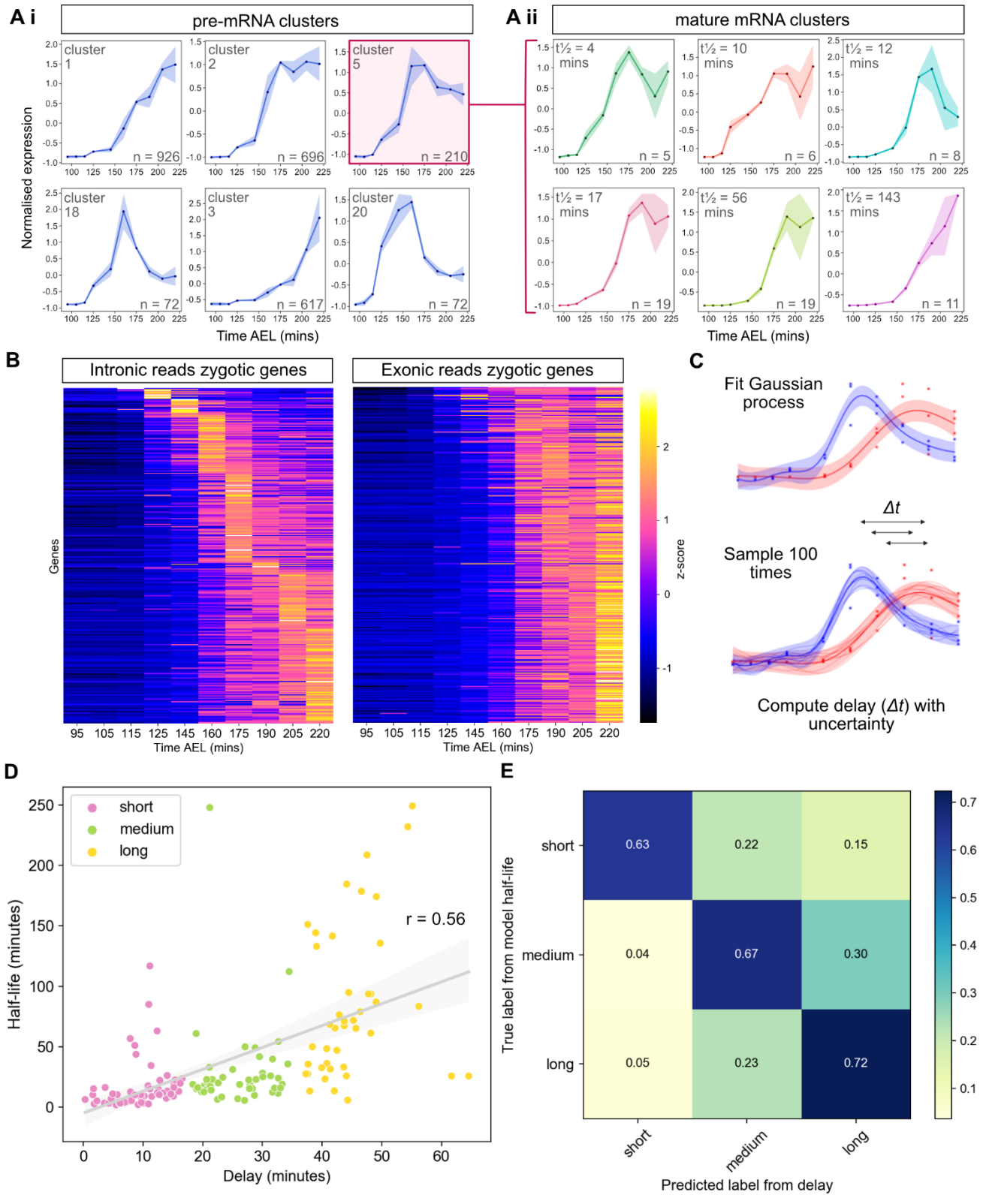


Figure 3: Degradation regulates mRNA dynamics and can be approximated using the time delay between peak pre-mRNA and mature mRNA expression. (A) Clustering of pre-mRNA (Ai) and mature (Aii) mRNA time series. Plots show the mean expression value at each time point for all mRNAs

in a cluster and the shaded area shows the standard deviation. The intronic cluster ID (Ai) or mean half-life (Aii) is shown in the top left and the number of transcripts in each cluster in the bottom right. (B) Heatmaps of pre-mRNA and mRNA expression at the gene-level for the subset of zygotic transcripts, ordered by the time point of maximum pre-mRNA expression. (C) Schematic illustrating the estimation of temporal delays between pre-mRNA and mature mRNA peak times (Δt, arrows) by fitting and sampling from a GP. (D) Correlation of the delay between the peak of pre-mRNA and mature mRNA against inferred half-life (Pearson's r = 0.56, p = 6.5 x 10^{-14} for testing non-correlation). Points representing transcripts are coloured by time delay category. (E) Confusion matrix comparing genes categorised into short, medium and long delays and their respective half-life categories. Numbers in the boxes indicate the fraction of genes with a given delay in the corresponding half-life category.

As the clustering data indicated that half-life contributes to the shape of the mature mRNA profile, we further investigated the relationship between the relative timing of the peak of the pre-mRNA and mature mRNA. Visualisation of the gene-level pre-mRNA and mature mRNA data from the zygotic subset as heatmaps, reveals that for a given pre-mRNA peak time, there are a range of mature mRNA peak times with different delays (Figure 3B). Delay is defined as the time difference at which the peak is observed for the pre- and mature mRNA. The pre-mRNA and mature mRNA data were modelled using a Gaussian process which was then sampled with n = 100, so that the time delay between the peaks could be determined and the uncertainty in the estimate quantified (Figure 3C). The relationship between delay and half-life, for each transcript that has been modelled, is shown in Figure 3D. There is a moderately positive yet significant correlation between the two variables. Figure 3E shows the data as a confusion matrix in order to assess whether delay is predictive of half-life. Enrichment along the diagonal supports this; 63% of short delay genes have short half-lives; 67% of medium delay genes have medium half-lives and 72% of long delay genes have long half-lives. Together, these results reveal how post-transcriptional regulation is able to shape mature mRNA dynamics through regulation of mRNA half-lives and that the time delay between maximum expression of the pre-mRNA and mature mRNA can be used as an indicator of mRNA stability.

The short half-life mRNAs tested are more compact

The degradation of an mRNA in the cytoplasm can be closely linked to its translation [1]. We therefore investigated how mRNA half-lives are shaped by both structural and sequence features known to influence translation. Regulatory sequences controlling mRNA degradation, translation and localisation are frequently located in the 3'UTR [40]. We found that 3'UTR length does not have any significant correlation with our inferred half-lives (Figure S7A), in agreement with previous studies of mRNA stability in late stage *Drosophila* embryos [37]. Similarly, there is no relationship between transcript length and stability in our dataset (Figure S7B).

Due to the links between mRNA decay, translation efficiency and codon optimality [41], we next investigated whether there is a relationship between half-life and both the translation efficiency and codon usage across the transcripts within our dataset. Using published ribosome footprint profiling data from 2-3 hour embryos [42] we plotted the translation efficiency and half-life for each of the transcripts within our dataset and observed no significant correlation between translation efficiency and half-life (Figure 4A). To extend this analysis, we also determined the codon stabilisation coefficient (CSC) for each codon which is a measure of the correlation between codon usage and stability of mRNAs. We plotted the CSC of each codon ordered by this value from highest to lowest (Figure 4B) and examined the identity of optimal codons previously defined in *Drosophila* embryos [43] and their occurrence within the CSC plot. The proportion of optimal codons is not significantly enriched within the positive and negative CSC groups (33% vs. 39%, p = 0.79, Figure 4B). There is also no significant difference in the proportion of optimal codons for transcripts within each of the different categories of half-life (Figure 4C) and clustering mRNAs based on codon usage showed that different clusters had similar half-lives (Figure S7C).

We next used imaging to analyse mRNA compaction in the context of stability. A more open conformation has been detected for specific mRNAs when they are being translated [44–46], raising the possibility that a particular conformation may also influence mRNA stability. We therefore selected a set of 11 zygotic mRNAs, 4 each from the medium and long half-life categories and 3 from the short half-life category (Figure S8A). A 4th short half-life mRNA, *Neu2*, is too short to separate the probe sets for compaction analysis but is included in further analysis (see later). We used dual-colour smFISH probes to visualise their 5' and 3’ ends, and quantitate the distance between them, in fixed embryos (Figure 4D). A representative smFISH image for one of the mRNAs, *lov*, is shown in Figure 4E, images for the other mRNAs tested are shown in Figure S8B.

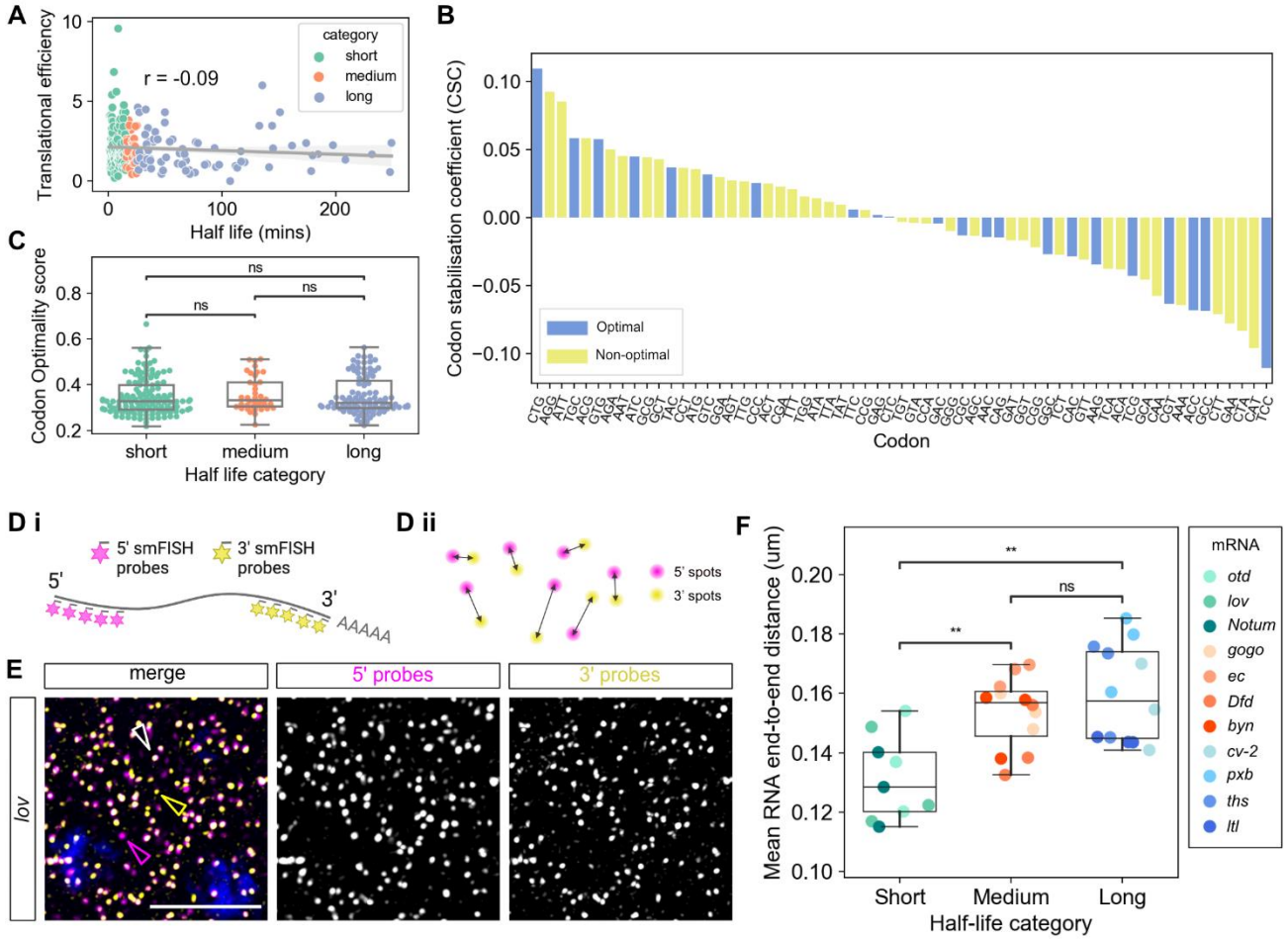


Figure 4: mRNA properties and stability. (A) Half-life (x-axis) versus the translation efficiency from 2-3 hour embryos [42]. In all panels points representing transcripts are coloured by half-life category (Pearson's r = -0.09, p = 0.18). (B) Codon stabilisation coefficients calculated from our estimated half-lives showed no difference in the proportion of optimal (blue) and non-optimal codons (yellow) (chi-squared test p = 0.79). (C) Proportion of optimal codons within transcripts from each half-life category. No significant difference was observed in the percentage of optimal codons within each category, tested by independent t-test (short vs med p = 0.5; short vs long p = 0.7; med vs long p = 0.7). (Di) Schematic showing detection of the 5' (magenta) and 3' (yellow) ends of each mRNA with different smFISH probe sets. (Dii) Spots belonging to the same mRNA are matched (see Methods). (E) Maximum projection of 6 slices from a confocal image showing smFISH detection of the 5' and 3' ends of *lov* mRNAs with lone 5' ends, 3' ends and colocalised ends labelled by magenta, yellow and white arrowheads, respectively. Scale bars: 5 µm. In the uncropped image from this embryo there are 5668, 1645 and 3620 intact, lone 5' and lone 3' signals, respectively. For absolute numbers of intact mRNAs and lone ends for all mRNAs, see Figure S9A. (F) Graph shows the end-to-end distances of mRNAs with different stabilities, n = 3 embryos per mRNA. Data are mean distances across all colocalised mRNAs in each embryo (n > 220 whole RNAs for all images). Genes are grouped by their half-life category and the hue in each category corresponds to the order of the half-lives (lighter colour refers to shorter half-life). Short half-life mRNAs are more compact than both medium ($p = 3.1 \times 10^{-3}$) and long ($p = 1.6 \times 10^{-3}$) half-life mRNAs. No significant difference in end-to-end distance was seen between medium and long half-life transcripts ($p = 8.9 \times 10^{-1}$).

For each image, the number and position of the 5' and 3' signals were collected and pairs were identified by solving a paired assignment problem (Figure 4Dii). For each pair, the distance between the 5' and 3' signals was then measured; only ends with a distance less than 300 nm were assigned as same mRNA [46]. First, we estimated our smFISH detection efficiencies using alternating fluorophores for the *otd* and *lov* mRNAs (Figure S9Bi). These data reveal mean detection efficiencies of ~70% (Figure S9Bii), which is in the 70-90% range reported from other smFISH studies [47–51]. However, we note that our detection of the 670 labelled probe sets is generally slightly poorer than that of the 570 probes due to a lower signal to noise, consistent with findings from a previous study that used 670 labelled probes [49].

The distributions of end-to-end distances for each of the mRNAs tested reveal that short half-life mRNAs are significantly more compact, based on a smaller end-to-end distance, than mRNAs in the medium and long half-life categories (Figure 4F). Considering the lower detection limit of the imaging setup we used is ~120nm (see Methods), we found that for our alternating probe sets and the 5' and 3' compaction data, the *otd* and *lov* short half-life mRNAs had an end to end distance that is very compact and close to this limit (Figure 4F). No significant difference is observed in the end-to-end distance for mRNAs in the medium and long half-life categories however we did find some mRNAs in the long category were in a more open conformation than those in the medium category (Figure 4F). We also identified unpaired mRNA ends (see later), which were further apart than the 300 nm distance threshold used. Finally, quantitation of additional control smFISH experiments for some of the test mRNAs, in which the fluorophore dyes on each set of probes were switched (Figures S9C) in order to control for detection differences between the channels mentioned above, also revealed significantly shorter end-to-end distances for short half-life mRNAs (Figure S9D).

We find no correlation between compaction and mRNA length, as *brachyenteron* (*byn*) and *echinus* (*ec*) are the shortest and longest mRNAs tested, respectively. Taken together these results suggest that within early *Drosophila* development, the decay of zygotically expressed genes is not strongly correlated with translation efficiency or codon optimality, but unstable mRNAs tend to be slightly more compact than medium and long half-life transcripts.

Embryonic P-bodies are associated with unstable mRNAs and enriched in 3' decay fragments

Cytoplasmic P-bodies have been implicated in mRNA degradation and storage in *Drosophila* [52]. Therefore, we investigated whether mRNAs with distinct stabilities are differentially localised to P-bodies. We visualised P-bodies using Me31B, a marker of P-bodies, including in *Drosophila* [53]. To detect Me31B we used a fly stock carrying a GFP-Me31B exon trap with GFP inserted into the Me31B locus [54]. We quantified both single mRNAs using smFISH and P-bodies labelled by GFP-Me31B in fixed embryos. The same set of eleven mRNAs described above was used in these experiments with

the addition of *Neu2*, a 1126 nt mRNA which was unsuitable for compaction analysis due to its short length. Many GFP-Me31B foci were detected in the cytoplasm of early nc14 embryos (Figure 5A, Figure S10A). These foci have a mean radius of 200 nm (Figure S10B), consistent with a previous observation that P-bodies in the embryo are smaller than those in the oocyte [55].

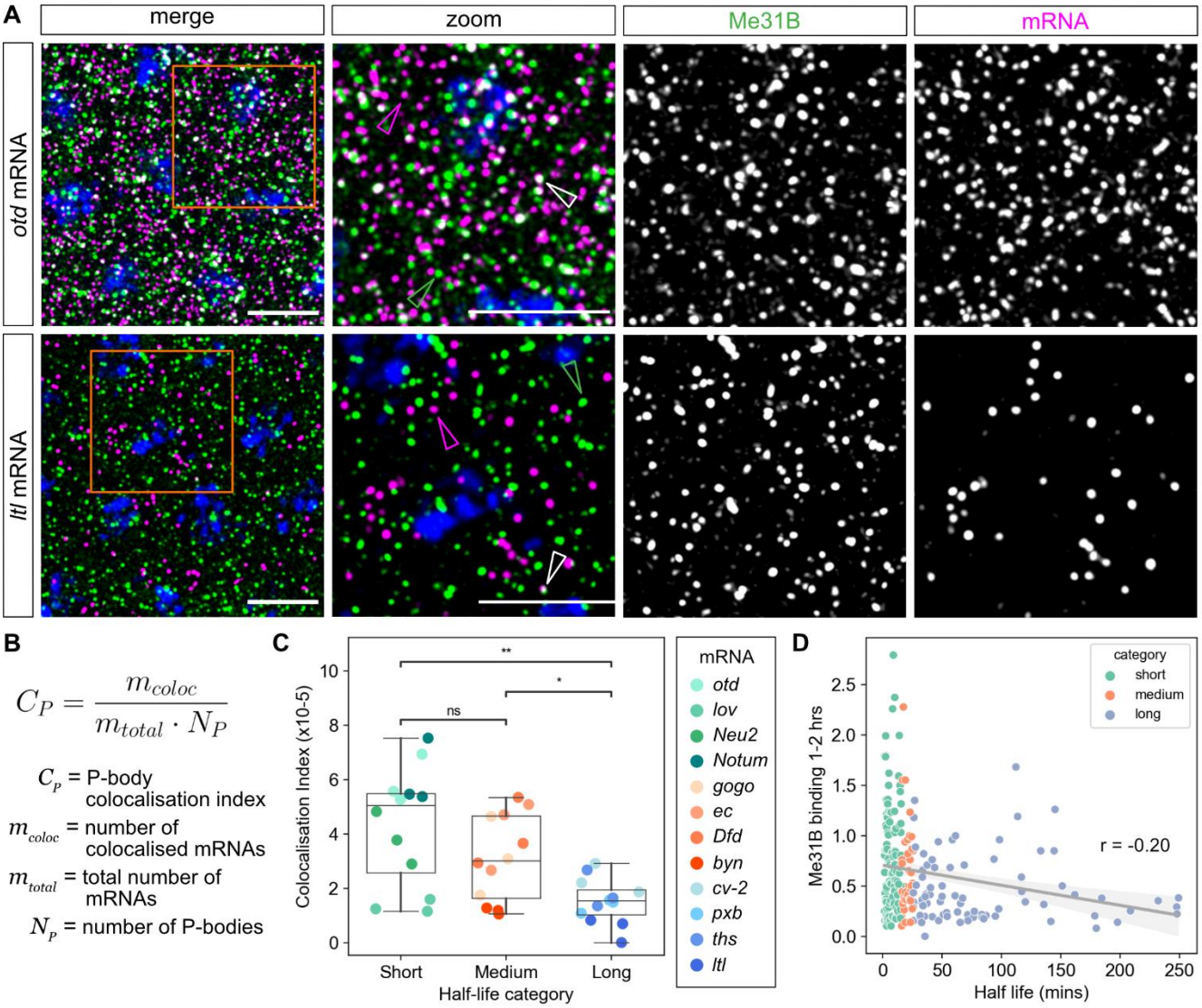


Figure 5: Short half-life mRNAs are more colocalised with P-bodies in the early embryo.
(A) Confocal images of fixed, early nc14 embryos stained with smFISH probes for the indicated mRNAs (magenta) and labelled GFP-Me31B P-bodies (green). Scale bars: 5 µm. Images are maximum projections of 7 slices, with higher magnification images of the highlighted regions (orange box) shown. Individual mRNAs (magenta arrowheads), P-bodies (green arrowheads) and colocalised mRNA and P-body signals (white arrowheads) are highlighted. (B) The P-body colocalisation index used to calculate the normalised proportion of colocalised mRNAs, facilitating comparison between different mRNAs. (C) Graph of the P-body colocalisation index for indicated mRNAs in early nc14. mRNAs are grouped by half-life category, within which they are coloured by half-life as per Figure 4F with points representing individual embryos. Both short and medium half-life mRNAs are significantly more colocalised with P-

bodies than long half-life mRNAs ($p = 1.257 \times 10^{-3}$ and $p = 1.640 \times 10^{-2}$, respectively). Short half-life mRNAs are not significantly more enriched in P-bodies than medium half-life mRNAs ($p = 4.168 \times 10^{-1}$). (D) Half-life versus Me31B binding data from 1-2 hour embryos [52] (Pearson's $r = -0.20$, $p = 1.5 \times 10^{-3}$)

For each mRNA tested, a proportion of the individual mRNA signals colocalise with P-bodies (Figure 5A, Figure S10A). As seen in Figure 5A, orthodenticle (*otd*) (also called *ocelliless*) mRNAs appear more highly colocalised with P-bodies than *ltl* mRNAs. As *otd* has a much shorter half-life than *ltl* (3 mins and 249 mins respectively), we examined whether this was a trend across the set of test mRNAs. To quantitate colocalisation, we used a colocalisation index that controls for variation in mRNA and P-body numbers between embryos (Figure 5B). This analysis reveals that the both the short and medium half-life mRNAs tested are significantly more colocalised with P-bodies than the long half-life mRNAs tested (Figure 5C). While the mean colocalisation index value for short half-life mRNAs is higher than that of the medium half-life mRNAs tested, this difference is not significant, due to higher variance in the colocalisation index of short half-life mRNAs as *lov*, a short half-life mRNA, has a particularly low colocalisation index.

Given the difference in P-body colocalisation observed for some of the test mRNAs, we extended this analysis by using published Me31B RIP-seq data from the early *Drosophila* embryo [52]. This analysis reveals a relatively weak but significant negative correlation between Me31B interaction and mRNA half-life in 1-2 hr embryos (Figure 5D) and 2-3 hr embryos (Figure S10Ci). This negative correlation between the Me31B RIP-seq data and our model half-lives is no longer significant when the RIP-seq data from 3-4 hr embryos are used (Figure S10Cii), a later stage than we have imaged. These data are consistent with a previously reported negative correlation between Me31B binding and mRNA stability in the *Drosophila* embryo, when fold change in mRNA abundance was used as a proxy for mRNA stability [52]. Together, our imaging data and the negative correlation between RIP-seq interaction and mRNA half-life suggest that in the *Drosophila* embryo P-bodies may be sites of mRNA degradation for at least a subset of mRNAs.

In our dual-colour smFISH images we observed a proportion of unpaired 5' and 3' mRNA ends suggestive of degradation intermediates (Figure 4E). We detect more lone ends when we use 5' and 3' *otd* compaction probes, compared to alternating probes (Figure S9E), providing further support that some of the lone signals are due to mRNA degradation, as detection with alternating probes is more resistant to loss of mRNA 5' and 3' sequences. In addition, due to the short length of *Drosophila* mRNAs, we are using 24-30 probes in each detection set. Therefore, it is likely that loss of binding of only a small number of probes from the 5' or 3' set is enough to take the signal below the detection threshold, facilitating our detection of partly degraded mRNAs.

In order to determine if these 5' and 3' fragments co-localised with P-bodies we assessed whether the 5' and 3' probe sets colocalised with the GFP-Me31B P-body marker. An image of an early

nc14 embryo is shown for the *Dfd* mRNA in Figure 6A, revealing that some complete mRNAs (orange arrowhead) and lone 3' ends (yellow arrowhead) are colocalised with the P-body marker Me31B. However, colocalisation of lone 5' ends with Me31B is less evident. For clarity, an equivalent region of an early nc14 embryo is shown as 3 colour images with only either the 5' or 3' end of *Dfd* mRNAs, Me31B and DAPI (Figure 6B, Figure S11). For the analysis, we identified unpaired 5' and 3' ends as described above and assessed if there is an enrichment of either end in P-bodies (Figure S12A). In general, we do not see an excess of lone 3' ends compared to 5' ends across the mRNAs we tested (Figure S9). However, quantitation of the proportion of single 5' and 3' signals that localise to P-bodies reveals a general trend of more unpaired 3' ends in P-bodies, which is significant for over half the mRNAs investigated (Figure 6C). Similar results are obtained when the fluorophores on the *otd*, *Dfd* and *cv-2* 5' and 3' probes are reversed (Figure S12Bi). Furthermore, this trend is lost when we use alternating probes for *otd* (Figure S12Bii). Taken together these results suggest that the lone 3' signals detected in P-bodies are consistent with them being 5' to 3' mRNA decay intermediates. Additionally, comparison of the proportion of lone ends versus intact mRNAs in P-bodies (relative to the total number of each) reveals that in general the proportion of intact mRNAs and lone 3' ends in P-bodies is similar but there are more intact mRNAs that lone 5' ends (Figure S12C). This detection of intact mRNAs in P-bodies may support a storage role in addition to 5' to 3' decay (see Discussion). Together, these data suggest mRNA degradation can occur within P-bodies for at least some mRNAs in the early *Drosophila* embryo.

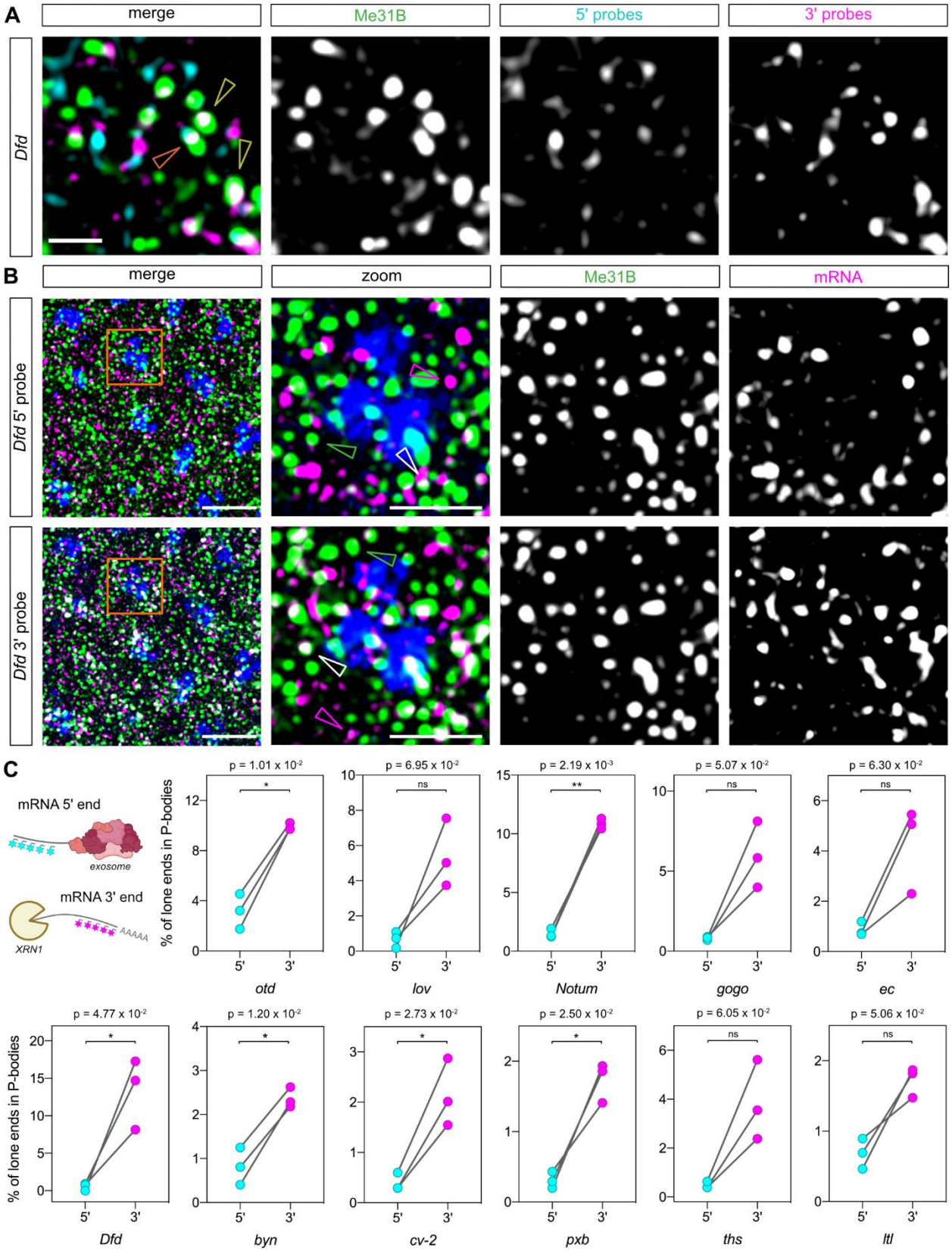


Figure 6: mRNA 3' end fragments are more associated with P-bodies than 5' ends.

(A) Maximum projection (2 slices) of a confocal image showing smFISH staining of an early nc14 embryo with probes for *Dfd* 5' (cyan) and 3' (magenta) ends, GFP-Me31B (green) and DAPI staining (blue). Examples where both the 5' and 3' ends or only the 3' end is colocalised with P-bodies are indicated by orange and yellow arrowheads, respectively. Single channels for the smFISH and GFP-Me31B are shown with the merged image. Scale bar: 1 µm. (B) As in (A) except the images (7 Z slices) show only one mRNA end (5' in the top panels, 3' in the lower panels) at a time for clarity. The mRNAs, GFP-Me31B and DAPI are shown in magenta, green and blue, respectively. A higher magnification image is shown as a merge and single channels, with individual mRNA ends (magenta arrowheads), P-bodies (green arrowheads) and colocalised mRNA end and P-body signals (white arrowheads) highlighted. Scale bars: 5 µm in merge and 2 µm in the higher magnification image. (C) Top left: Schematic showing how degradation by either pathway would result in a single fluorophore signal being detected. Quantification of the percentage of unpaired mRNA 5' and 3' ends with P-bodies relative to the total number of lone 5' or 3' ends. n = 3 embryos, paired t-test used to determine significance with $\alpha = 0.05$. For absolute numbers of intact mRNAs and lone ends, see Figure S9A and S9B.

Discussion

Here, using total RNA-seq time series data and Gaussian process regression, half-lives of ~260 mRNAs in early *Drosophila* development were derived. Our data support widespread post-transcriptional regulation of gene expression in early development, as we show that shared transcription profiles give rise to a range of mature mRNA dynamics due to differences in degradation. The RNA-seq time series that we have generated is high resolution with additional time points and over an extended period of early embryogenesis compared to published data sets [24,56]. In addition, our libraries are total RNA-seq rather than poly(A) selected, facilitating detection of non-coding RNAs and unstable RNA species, such as co-transcriptionally spliced introns. Our RNA-seq data reveal how expression of different mRNA isoforms for a given gene varies across early embryogenesis and we highlight examples where isoform changes alter the protein sequence of specific domains, potentially impacting on function.

A major advantage of our approach is that it does not require transcription inhibition, which can affect mRNA stability, or mRNA labelling, which can be difficult to achieve *in vivo* and the labelling time can influence the estimates [19,20,57]. A different method that uses RNA-seq data to estimate mRNA half-lives has been described previously, which solves ODEs describing the RNA life cycle by adopting constraints on RNA kinetic rates [58]. An advantage of our approach is that, as Gaussian process regression is non-parametric, there is greater flexibility and sensitivity in the model to more accurately represent the variety of dynamics observed in the data. Additionally, it allows us to quantify uncertainty due to biological variation and measurement noise. Using Bayesian inference, we obtain posterior distributions for degradation parameters and thus quantify uncertainty about their possible values.

Disadvantages of our approach are firstly that it is not global as ~20% of *Drosophila* genes expressed in our dataset do not contain introns. Secondly, even for mRNAs with introns, we only derived a proportion of transcript half-lives from the dataset due to strict filtering to ensure that there is signal in

both the intron and transcript expression time-series meaning genes with small introns and therefore poor signal would also be excluded. Thirdly, for genes with high degradation rates, there may be high uncertainty in the inferred degradation rate since the splicing rate and degradation rate estimates become difficult to disentangle (for simulations demonstrating the reach of the model see Supplementary Methods). Fourthly, the modelling requires some computational expertise in this area to implement on a new dataset. Potential solutions to overcoming these issues would be to generate a Pol II ChIP-seq time series for the transcription profiles that would allow intronless genes to be studied. Looser filtering could be applied to provide half-lives for more mRNAs, although this would potentially lower confidence in the estimates. Finally, the delay between the peak of the pre-mRNA and mature mRNA could be measured as a simpler approach for categorising stability, as we have shown that the stability of a transcript can be classified using this delay.

The half-lives we estimated for ~260 zygotic transcripts in the early embryo have a median of 16 minutes. Previous half-life estimates of 7-14, 13 and 60 mins described for the zygotic *fushi tarazu*, *snail* and *hunchback* mRNAs, respectively, in the early *Drosophila* embryo [23,59,60] fall within the range of half-lives we observe. Moreover, the wide range of half-lives we estimate in the embryo suggests that mRNA stability is an important checkpoint in the regulation of gene expression. The median half-life we estimate is shorter than that of 73 minutes calculated for older (stage 12-15) *Drosophila* embryos, in a study that used a 4 hour pulse-chase labelling [37]. While the pulse-labelling timing may skew some of the half-life estimates [57], the shorter median half-life in the early embryo may reflect its rapid initial development. Early embryogenesis is characterised by short mitotic cycles [61] and fast rates of transcription [62] and translation [63], with the resulting localised gene expression patterns specifying three tissues along the dorsal-ventral axis in a time period of only 90 mins [64]. Therefore, mRNA degradation rates may be faster than at other stages to limit the perdurance of transcripts encoding factors affecting cell fate.

Gene ontology analysis revealed an enrichment among the short half-life mRNAs for those encoding transcription factors and cell adhesion proteins. This is consistent with transient localised expression of key transcription factors in the early embryo and the mRNAs encoding transcription factors commonly being unstable [17,37,38]. Future studies will be able to determine how particular mRNA half-lives contribute to patterning by exploiting the extensive characterisation of gene regulatory networks in the early *Drosophila* embryo [65].

Previous studies have shown that mRNAs exist in a more open conformation during translation, while untranslated mRNAs are more compact [44–46] regardless of whether they are stress granule associated [44,45]. We found a trend that the 5’ and 3’ ends are closer for shorter half-life mRNAs. A more compact structure may facilitate degradation as 5’ to 3’ decay involves communication between deadenylation and decapping factors [66]. Alternatively, the shorter distance between 5’ and 3’ ends

could reflect a transient interaction associated with degradation, which our smFISH snapshot images capture more frequently for the less stable mRNAs.

Codon identity and translation efficiency have previously been shown to be an important determinant of mRNA stability in bacteria, yeast, *Drosophila*, zebrafish and mammalian cells [41]. Optimal codons, which are determined by codon bias in abundant mRNAs and the gene copy number of their cognate tRNA, lead to efficient translation and are enriched in stable transcripts [41]. However, our data suggest that codon optimality and translation efficiency are not major determinants of mRNA stability for early zygotic transcripts. A correlation between codon optimality and mRNA stability was observed for maternal mRNAs during the maternal to zygotic transition in the early *Drosophila* embryo, which likely contributes to clearance of maternal transcripts [67]. Optimal codons are also associated with stable mRNAs in late-stage *Drosophila* embryos, but not in neural tissues, potentially because mRNA stability regulation by RNA binding proteins dominates in the nervous system [43]. The effect of codon optimality may also be masked for early spatially regulated zygotic transcripts. This could be due to additional regulation by RNA binding proteins and miRNAs [5], a dependence on a particular distribution of non-optimal codons for instability [68] and/or tRNA abundance being a poor proxy for aminoacylated tRNA levels for a subset of tRNAs. In support of the latter, low aminoacylation of particular tRNAs has been observed in the mouse liver that may contribute to inefficient translation [69].

Our imaging data on the test set of mRNAs show that those with short half-lives tend to be more colocalised with the P-body marker Me31B than more stable mRNAs in the early embryo. Consistent with this, using published Me31B RIP-seq data from the early *Drosophila* embryo, we find a significant correlation between Me31B interaction and mRNA half-life across the set of ~260 mRNAs for which we estimated half-lives. The stronger association of short half-life mRNAs with P-bodies and our ability to detect mRNAs lacking their 5' end in P-bodies suggests that 5' to 3' mRNA decay can occur in P-bodies in the early *Drosophila* embryo. However, the majority of the lone 3' ends we detect are in the cytosol, suggesting that mRNAs can also undergo 5' to 3' decay outside of P-bodies. We also note that the localisation of mRNAs with P-bodies is variable as, within the short and medium half-life categories, the *lov* and *byn* mRNAs are less colocalised with P-bodies. This suggests that for these mRNAs in particular, degradation in P-bodies may only have a minor contribution to their turnover.

Although we generally detect similar proportions of 5' and 3' end fragments of a particular mRNA in the cytoplasm, there is weaker colocalisation of 5' end fragments with P-bodies. This observation suggests that 3' to 5' mRNA degradation by the exosome does not occur in P-bodies, consistent with components of the exosome being largely absent [4]. We also detect a similar proportion of intact mRNAs in P-bodies (relative to the total number in the cytoplasm), as we find for lone 3' ends. The presence of intact mRNAs in P-bodies may reflect an mRNA storage role. Therefore, we speculate that in the *Drosophila* embryo mRNAs enter P-bodies where they can undergo either: 1) 5' to 3' degradation

(hence the lone 3' ends detected), or 2) transient storage before exit back into the cytoplasm for translation.

A role for P-bodies in 5' to 3' decay is consistent with early studies in yeast following the discovery of P-bodies [70] and with later work in *Drosophila* suggesting that Me31B is involved in mRNA degradation in the embryo following zygotic genome activation [52] and P-bodies are sites of mRNA degradation in intestinal stem cells [71]. In addition, the Xrn1 exonuclease localises to P-bodies in yeast, *Drosophila* and mammalian cells [72]. However, P-bodies have been implicated in mRNA storage and translational repression in mature *Drosophila* oocytes [55] and Me31B represses translation of maternal mRNAs in *Drosophila* embryos prior to zygotic genome activation [52]. Moreover, many lines of evidence from other systems argue against a role for P-bodies in mRNA degradation. These include an absence of detectable mRNA decay intermediates either following purification of P-bodies [73] or based on a live imaging approach [47], mRNA degradation when P-body formation is disrupted [73] and the ability of P-body mRNAs to re-enter translation [74,75]. Although the sequencing data following P-body purification from human tissue culture cells provided evidence for mRNA storage, and do not support a role for P-bodies in bulk mRNA degradation [73], we note that two pieces of data are potentially consistent with some degradation occurring in P-bodies. Firstly, there is a weak correlation between mRNA P-body enrichment and half-life, and secondly a 3-fold difference in the median half-lives of the most strongly enriched versus depleted P-body mRNAs was observed [73].

We speculate that P-bodies are involved in both storage and degradation in an mRNA dependent manner, with features of an individual mRNA as well as the proteins present in P-bodies at a particular developmental time influencing which function dominates. In support of this, it is known that there are changes in P-bodies during *Drosophila* development, for example from being large and viscous in the oocyte to smaller, more dynamic structures in the early embryo [55]. Moreover, at the maternal-to-zygotic transition some P-body proteins are degraded, including the Cup translational repressor protein, which may increase the prevalence of mRNA decay in P-bodies [52]. Our data suggest that the degradation of mRNAs in P-bodies contributes to the post-transcriptional regulation of zygotic mRNAs in the embryo. Future studies exploiting the method developed for determining the protein and RNA contents of purified P-bodies [73], along with the power of *Drosophila* genetics and single molecule imaging, will reveal how P-bodies impact on mRNA stability or storage and cell fate decisions during development.

Materials and Methods

Biological methods

549

550 **Fly stocks**

551 All stocks were grown at 25°C and maintained at 20°C for experiments on standard fly food media (yeast
552 50g/L, glucose 78g/L, maize 72g/L, agar 8g/L, 10% nipagen in EtOH 27mL/L and proprionic acid 3mL/L).
553 The following fly lines were used in this study, y^1 w* (BDSC Stock #6599), y^1 w*; P{His2Av-mRFP1}II.2
554 (BDSC Stock #23651) and y^1 w*; P{w[+mC]=PTT-GB}me31B[CB05282] (BDSC Stock #51530).

555

556 **Staging and collection of embryos for RNA-seq**

557 Flies carrying His-RFP were allowed to lay on apple juice agar plates in small cages for 1 hour. Embryos
558 were dechorinated in 50% bleach (2.5% final concentration of sodium hypochlorite diluted in distilled
559 water) for 3 minutes and washed thoroughly in distilled water. Individual embryos were carefully
560 transferred into a droplet of halocarbon oil (Sigma-Aldrich; a mix of 700 and 27 oil at a ratio of 1:4) on a
561 rectangular coverslip (Deltalab, 24X50mm, Nr. 1) and inverted over a cavity slide (Karl Hecht). Embryos
562 were visualised and imaged with a Leica optigrid microscope at 20X magnification using a Texas red
563 filter. Embryos were timed following the observation of a nuclear division, an image was taken and the
564 embryo was immediately picked out of the oil droplet with a pipette tip and transferred to Eppendorf
565 tubes containing 50µL TRIzol Reagent (Invitrogen). Single embryos were crushed and homogenised
566 using a pipette tip and an additional 450uL Trizol added. Samples were immediately snap frozen in
567 liquid nitrogen and stored at -80°C until processing for nucleic acid extraction.

568 Ten timepoints were collected spanning early *Drosophila* embryonic development from nc11
569 through to cephalic furrow formation (Table 1). Embryos were collected 5 minutes after nuclear division
570 for nc11 and nc12, 5 and 15 minutes following the nc13 nuclear division and every 15 minutes following
571 the nc14 nuclear division as well as embryos that showed clear cephalic furrow formation. This yielded
572 samples covering every 10-15 minutes through development from nc11 to cephalic furrow formation.
573 The internuclear distance of 15-20 nuclei pairs per embryo was measured in Fiji and normalised to the
574 whole embryo length to obtain an average internuclear distance per embryo (Figure S1A and B). This
575 was compared to the internuclear distance of embryos of known stages to accurately confirm the nuclear
576 cleavage stage and age of embryos. All embryos were collected at 20°C with approximate time after
577 egg lay in minutes shown in Table 1.

578

579 **Table 1: *Drosophila* embryo samples collected for RNA-seq time course**

Sample collection time information	Sample stage ID	Approximate time after egg lay at 20°C (mins)

nc11 + 5mins	nc11	95
nc12 + 5 mins	nc12	105
nc13 + 5 mins	nc13A	115
nc13 + 15 mins	nc13B	125
nc14 + 15 mins	nc14A	145
nc14 + 30mins	nc14B	160
nc14 + 45mins	nc14C	175
nc14 + 60 mins	nc14D	190
nc14 + 75 mins	nc14E	205
CF	CF	>220

Nucleic acid extraction and embryo genotyping

Samples stored in Trizol (Invitrogen) were used for RNA and DNA extraction performed according to the manufacturer's protocol and resuspended in 10µL (RNA) or 20µL (DNA) nuclease free water. Extracted DNA was PCR amplified to sex the embryos by using Y chromosome specific primers to a region of the male fertility factor gene kl5, forward primer 5' GCTGCCGAGCGACAGAAAATAATGACT 3' and reverse primer 5' CAACGATCTGTGAGTGGCGTGATTACA 3' [24] and control primers to a region on chromosome 2R forward primer 5' TCCCAATCCAATCCAACCCA 3' and reverse primer 5' CCTACCCACAAGCAACAACC 3'. PCR reactions were performed in triplicate.

Total RNA was treated with TURBO DNA-free Kit Dnase (Invitrogen) and depleted of rRNA using the Ribo-Zero Magnetic Kit HMN/Mouse/Rat 24 Rxn (Illumina; Cat# MRZH11124) according to the manufacturer's protocol using a low input protocol with 2-4µL rRNA removal solution yielding a 20µL final sample volume added to 90µL magnetic beads. Beads were resuspended in 35µL resuspension solution and ribo-depleted total RNA was ethanol precipitated and resuspended in 18µL FPF mix prior to RNA-seq library preparation.

RNA-seq library preparation and sequencing

Three female embryos from each time point were used as replicates to make 30 individual RNA-seq libraries. Individual total RNA-seq libraries were prepared from ribo-depleted RNA using a TruSeq

599 stranded library prep kit (Illumina) according to the manufacturer's protocol. Unique dual index adaptors
600 were used for each library and they were pooled in equimolar concentration and run across 8 lanes on
601 the flow cell of the HiSeq 4000 to obtain paired end sequence reads. The average number of reads
602 obtained per library was 105 million reads.
603
604 **Embryo fixation and smFISH**
605 Flies were allowed to lay on apple juice agar plates in small cages for 2 hours at 25°C. After ageing for
606 another 2 hours, 2-4 hour old embryos were dechorinated in 50% bleach for 3 minutes and washed
607 thoroughly in distilled water. Embryos were fixed as previously described [76] and stored in methanol at
608 -20°C until required. Fixed embryos were placed in Wheaton vials (Sigma, Z188700-1PAK) for the
609 smFISH reaction as described previously [77]. mRNA targets were detected in embryos using smiFISH
610 probes designed to exonic sequence with 5' end X flap sequences [78] and using secondary detection
611 probes labelled with Quasar 570 or Quasar 670 fluorophore dyes (LGC Biosearch Technologies). Probe
612 sequences are listed in Supplementary Table 2. DAPI (500µg/ml) was added to the third of the final four
613 washes of the protocol at a concentration of 1:1000 and embryos were mounted onto slides in Prolong
614 Diamond to set overnight before imaging. To visualise the membrane to age the embryos a mouse α-
615 Spectrin antibody (DSHB, 3A9 (323 or M10-2)) with an Alexa Fluor 647 Donkey anti-Mouse IgG (H+L)
616 Highly Cross-Adsorbed Secondary Antibody (Thermo Fisher Scientific, A-31571) was used or a
617 brightfield image was taken.
618 For compaction experiments at least 24 probes were designed to each end of the mRNA (5' and 3')
619 separated by at least 1.3kb. As a control, fluorophore dyes were switched and the images from stained
620 embryos analysed and quantified. Additional controls for *otd* and *lov* used adjacent probes with
621 alternating Quasar dyes to determine the precision of detection of single mRNAs.
622
623 **Confocal microscopy of fixed embryos**
624 A Leica TCS SP8 gSTED confocal was used to acquire images of the transcription sites (TSs), single
625 mRNAs and P-bodies within cells of fixed embryos using a 100x/ 1.3 HC PI Apo Cs2 objective with 3X
626 line accumulation and 3X zoom for compaction and P-body colocalisation experiments, and 2X zoom
627 for quantifying mRNAs for the half-life validation. Confocal settings were ~0.6 airy unit pinhole, 400 Hz
628 scan speed with bidirectional line scanning and a format of 2048 x 2048 or 4096 x 4096 pixels. Laser
629 detection settings were collected as follows: PMT DAPI excitation at 405nm (collection: 417-474nm);
630 Hybrid Detectors: AlexaFluor 488 excitation at 490nm (collection: 498-548nm), Quasar 570 excitation
631 at 548nm (collection: 558-640nm) and Quasar 670 excitation at 647nm (657-779nm) with 1-6ns gating.
632 All images were collected sequentially and optical stacks were acquired at system optimised
633 spacing. Imaging of the membrane using brightfield or anti-Spectrin antibody at the mid-sagittal plane

of the embryo with 40x objective at 0.75X zoom and 1024 X 1024 format was used to measure the average length of membrane invagination from at least 5 cells. These measurements were used to select embryos of a similar age in early nuclear cycle 14 (10 μm membrane invagination). For all analysis, 3 separate embryos were imaged and quantified as independent replicates. smFISH images were deconvolved using Huygens professional deconvolution software by SVI (Scientific Volume Imaging). By deconvolving images taken on a Leica SP8 confocal we estimate that our lower detection limit is ~120nm.

Image analysis

The spot detection algorithm Airlocalize [79] was used to detect and quantify TSs, single mRNAs and P-bodies within confocal microscopy images. This software fits a 3D gaussian mask and gives the coordinates in X, Y and Z of each spot and its intensity. Z stack images were first subsetted to detect TSs within the range of Z slices around the nuclei. Images were then processed again to detect single mRNAs in the full image. The TS data was then used to remove these high intensity spots from the single mRNA data. Detection of 5' and 3' single mRNA ends and P-bodies was performed separately on each corresponding channel image as appropriate.

Half-life validation

For validation of half-lives as previously described [23], embryos were imaged at various time points during the 13th nuclear division (Figure S4) using the DAPI channel and reference movies of His-RFP [77] to carefully time the images. Single mRNAs were quantified using Airlocalize and the number per cell was calculated by dividing by the total number of pre-division cells in the images. The counts per cell were fitted with an exponential function, from which the half-life was determined. The signal to noise ratio in the data was then calculated from the change in the mean over the time course, divided by the average variance in mRNA numbers at each timepoint with sufficient data.

Computational methods

RNA-seq data processing and data filtering

The RNA-seq data were processed at the transcript level by alignment-free methods using Kallisto [80] and the *Drosophila melanogaster* Ensembl v96 reference transcriptome to infer transcript expression levels for modelling mature mRNA dynamics. Transcripts were filtered to remove any that had <1 TPM (transcripts per million) across the time course yielding 18159 remaining transcripts. The transcript-level reads were then filtered for dynamic transcripts using GPcounts [81]. This approach fits the data for each gene using a GP with a dynamic kernel, and a GP with white noise kernel which exhibits no

variation over time. The transcripts where the dynamic kernel is a better fit, measured via likelihood ratio test, are then extracted. For the transcript data this reduced the numbers of mRNAs from 18159 to 8791.

The whole-embryo total RNA-seq dataset was also processed at the gene level in order to quantify the intronic reads, by aligning data to BDGP6 (dm6) using STAR with default parameters. FeatureCounts was used to get the counts data for exons and introns, respectively. Modified RPKM (reads per kilobase of transcript per million reads mapped) normalisation was applied to exon and intron counts data, where the total mapped reads for each library were used to address the sequencing depth for exon and intron counts from the same sample yielding 11,587 genes with a detectable level of expression (RPKM > 0).

To model the pre-mRNA dynamics, any genes without introns, or with zero intronic reads across all timepoints were removed to give a set of 5035 genes and the intron sizes were then used to obtain length-normalised reads. The intronic read counts are divided by the total intron length for each gene and the exonic expression level is similarly normalised. For modelling the mature mRNA dynamics the transcript-level alignment was used. A set of strictly zygotic transcripts were extracted from the dynamic dataset (n = 8791) by filtering for transcripts with TPM < 0.5 at the first time point (t = 95) to give a set of 593 zygotic transcripts which were used in subsequent analysis. For the GP model, transcripts were subjected to a further filtering step where the correlation between pre-mRNA and mRNA was computed to extract transcripts where the correlation was above 0.4. For more details on filtering see Supplementary Methods.

Analysis of intronic read coverage

BAM files were used to calculate the read coverage for introns and exons to the Drosophila_melanogaster.BDGP6.32.107.gtf annotation using the superintronic package in R with default parameters [82].

Validation of intronic data with NET-seq data

In order to validate that the intronic data is representative of nascent zygotic gene transcription, NET-seq data from early embryos was used [29]. Both the NET-seq and intronic data were length normalised and a Spearman's Rank correlation analysis was used to analyse the relationship.

Modelling

We model dependence between pre-mRNA, $p(t)$, and mature mRNA, $m(t)$, through a Gaussian processes regression which follows dynamics of an ODE of the form

$$\frac{dm(t)}{dt} = Sp(t) - Dm(t),$$

where $p(t)$, is assumed to be a Gaussian process with RBF kernel [31,32]. This differential equation can be solved in closed form and it can be shown that $m(t)$ is also a Gaussian process with a certain kernel. For more details and specification of this kernel we refer to Supplementary Methods. As the results, $m(t)$ and $p(t)$ can be modelled jointly as a Gaussian process regression with a block kernel which depends on biologically interpretable parameters such as S and D. It is assumed that we observe $m(t)$ and $p(t)$ at discrete times with measurement noise terms which have variances σ_m^2 and σ_p^2 for mRNA and pre-mRNA respectively. Thus, we have six parameters which we estimate: two parameters of RBF kernel (l – lengthscale, σ^2 – variance, which correspondingly define smoothness and amplitude of possible functions underlying pre-mRNA dynamics), two parameters S and D, which describe the relationship between mRNA and pre-mRNA, and two measurement noise variances σ_m^2 and σ_p^2.

We assign priors to these six parameters and use the L-BFGS-B algorithm to find maximum a posteriori (MAP) point estimates. Further, we perform sampling from the posterior distribution using Metropolis-adjusted Langevin algorithm (MALA). By using gradient information this algorithm allows exploring posterior distribution more efficiently than standard MCMC methods. After inferring the posterior distribution using sampling with MALA, we can quantify uncertainty about all six parameters of the model and uncertainty about the underlying functions for pre-mRNA and mRNA dynamics. In Supplementary methods Figure S1 we demonstrate the performance of the model and inference methods on simulated data: degradation rates ranging between 0.003 and 0.05 (corresponding to half-lives of 231.0min and 13.8min, respectively) are estimated accurately by both MAP and MALA approaches. The details of the implementation, including the assumptions on prior distributions, are discussed in Supplementary Methods.

From parameter estimates of D, half-lives were obtained using the following relationship:

$$t_{1/2} = \frac{\ln(2)}{D}$$

Transcripts were then grouped into short, medium and long half-life groups, setting the boundaries at 15 and 25 minutes. Any transcripts with a half-life longer than 250 minutes were removed from the dataset.

Python implementation of the model is available from https://github.com/ManchesterBioinference/GP_Transcription_Dynamics.

734 **Clustering**

735 The intronic data (n = 5035) was z-transformed and clustered using GPclust, a package specifically
736 designed for clustering noisy time series data using Gaussian processes [39]. Intronic clusters of interest
737 exhibiting a range of expression profiles were selected (clusters 2 and 5) (Figure S5). The zygotic
738 transcripts (n = 593) corresponding to the genes in each selected intronic cluster were then normalised
739 and clustered (Figure 3; Figure S6). Summary statistics for the half-lives of the genes in each zygotic
740 cluster were then computed for clusters with >2 transcripts with estimated half-lives. A list of the
741 transcripts for intronic clusters 2 and 5 and their corresponding zygotic clusters can be found in
742 Supplementary Table 3.
743
744 **Analysis of time delays**
745 Delays between peak of pre-mRNA and mature mRNA time series from the zygotic set (n = 593) were
746 estimated by fitting a Gaussian process with RBF kernel to each time series. 100 samples from each
747 GP were taken and the delay between the peak of each sampled function of premature and mature
748 mRNA were computed to provide estimates of the delay with uncertainty. Any transcripts with delays ≤
749 0, or with mature mRNA profiles peaking at the final timepoint (t = 220), were removed. Transcripts were
750 then grouped into short, medium and long delay groups, setting the boundaries from the 33% and 66%
751 quantiles of the data (17.55 and 36.16 mins respectively). A gene was classified as short if there was
752 90% probability that the delay for that gene was in the short delay interval. All statistical analysis was
753 carried out in Python using the scipy, sklearn and statannot libraries.
754
755 **Gene ontology analysis**
756 Gene ontology analysis was conducted using Gorilla [83]. Enrichment of short and long half-life genes
757 was performed using the half-life set as the target set and the entire group of dynamic genes from the
758 RNA-seq dataset (n = 8791) as the reference set with default parameters.
759
760 **Codon usage and translation efficiency analysis**
761 The codon stabilisation coefficient (CSC) value was calculated for each codon as previously described
762 [43,84]. The CSC is equivalent to the Pearson correlation coefficient, calculated by plotting the
763 frequency of each codon per transcript within our dataset against its half-life. Classification of optimal
764 *Drosophila* codons used are as in [43]. A chi-square test of association between optimal and non-optimal
765 codons in positive and negative CSC groups was determined. The codon optimality score was
766 determined by adding the proportion of optimal codons within each transcript. Transcripts were grouped
767 by their half-life category and an independent t-test was used to determine significance in codon

optimality between groups. Translation efficiency data was obtained from previously published data of 2-3 hour embryos [42]. 3'UTR and transcript lengths were obtained from Flybase [85].

Quantification of mRNA end-to-end distance

mRNA compaction, the distance between the 5′ and 3′ ends of the transcripts, was analysed using smFISH images where the 5′ and 3′ ends are bound by probes labelled with different fluorophores. After quantifying the number and position of the mRNA ends in both channels and removing transcription sites (see Image analysis), the spot position data was analysed with a custom Python script to find optimal spot pairs by solving a paired assignment problem. The distance between *n* 5′ spots and *m* 3′ spots are computed and stored as a distance matrix. The optimal assignment of 5' and 3' pairs is then found by minimising this distance matrix to give a set of paired spots with a minimum total distance between all pairs. Spot pairs are then filtered for distances less than 300 nm where the ends are considered to be colocalised and belonging to the same RNA. This 300 nm upper threshold was selected as described in a previous study [46]. For all colocalised 5' and 3' spots the distribution of distances was then analysed and summary statistics extracted.

Analysis of mRNA colocalisation with Processing bodies

mRNA localisation within Processing bodies (P-bodies) was determined from confocal images using a custom script in Python. This script uses the position data for the mRNAs and P-bodies output from Airlocalize and calculates the distance between a given mRNA and every P-body. The minimum distance is then selected so that an mRNA is assigned to its closest P-body. If this distance is less than 200 nm (a typical radius of a P-body) then the RNA is considered to be colocalised with the P-body. The proportion of mRNAs located within and outside of P-bodies is then analysed to determine whether a given gene is enriched within P-bodies in the cytoplasm. In order to do this, we derived the P-body colocalisation index, a measure of the degree of colocalisation with P-bodies of an mRNA of interest:

$$C_P = \frac{m_{coloc}}{m_{total} \cdot N_P}$$

Where C_P is the P-body colocalisation index, m_{coloc} is the number of mRNAs colocalised with P-bodies, m_{total} is the total number of mRNAs and N_P is the number of P-bodies.

For analysis of unpaired ends, any 5' or 3' spot which was unpaired from the optimal assignment, or was more than 300 nm away from its assigned pair, was considered. The colocalisation of these with P-bodies was then analysed using a more conservative threshold of 150 nm, to ensure a sufficient

proportion of the mRNA was located inside the P-body. The enrichment of unpaired ends in P-bodies was then derived by dividing the number of unpaired ends in P-bodies by the total number of unpaired ends for each channel.

Acknowledgements

We thank Jing Yang for processing the RNA-seq data, Lijing Lin for help with the intronic read coverage, Nuha BinTayyash for filtering the data for dynamics, Mark Ashe for helpful discussions, the University of Manchester Genomics Technology Core Facility, Fly Facility and Bioimaging Facility for support, and Mark Ashe, Sophie Frampton and Catherine Sutcliffe for comments on the manuscript. Schematics in figures were created using BioRender.com. This research was funded by a Wellcome Trust Investigator award to H.L.A. and M.R. (204832/Z/16/Z) and a Wellcome Trust PhD studentship to J.C.L. (222814/Z/21/Z). For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Author Contributions

Conceptualisation, L.F.B, J.C.L, Y.S, M.R and H.L.A; Investigation, L.F.B, J.C.L, Y.S; Software, Y.S., A.A; Writing; L.F.B, J.C.L, H.L.A, Y.S; Funding Acquisition, M.R. and H.L.A.

Declaration of interests

The authors declare no competing interests.

Availability of data and materials

RNA-seq data are deposited in ArrayExpress under accession number: E-MTAB-11580. Python implementation of the model is available from:
https://github.com/ManchesterBioinference/GP_Transcription_Dynamics.

References

1. Mugridge JS, Coller J, Gross JD. Structural and molecular mechanisms for the control of eukaryotic 5′–3′ mRNA decay. Nat Struct Mol Biol. 2018;25: 1077–1085. doi:10.1038/s41594-018-0164-z

2. Weick E-M, Lima CD. RNA helicases are hubs that orchestrate exosome-dependent 3′–5′ decay. Current Opinion in Structural Biology. 2021;67: 86–94. doi:10.1016/j.sbi.2020.09.010

3. Ivanov P, Kedersha N, Anderson P. Stress Granules and Processing Bodies in Translational Control. Cold Spring Harb Perspect Biol. 2019;11: a032813. doi:10.1101/cshperspect.a032813

4. Standart N, Weil D. P-Bodies: Cytosolic Droplets for Coordinated mRNA Storage. Trends in Genetics. 2018;34: 612–626. doi:10.1016/j.tig.2018.05.005

5. Mayya VK, Duchaine TF. Ciphers and Executioners: How 3′-Untranslated Regions Determine the Fate of Messenger RNAs. Front Genet. 2019;10: 6. doi:10.3389/fgene.2019.00006

6. Morris C, Cluet D, Ricci EP. Ribosome dynamics and mRNA turnover, a complex relationship under constant cellular scrutiny. WIREs RNA. 2021;12. doi:10.1002/wrna.1658

7. Akira S, Maeda K. Control of RNA Stability in Immunity. Annu Rev Immunol. 2021;39: 481–509. doi:10.1146/annurev-immunol-101819-075147

8. Akiyama T, Suzuki T, Yamamoto T. RNA decay machinery safeguards immune cell development and immunological responses. Trends in Immunology. 2021;42: 447–460. doi:10.1016/j.it.2021.03.008

9. Fraga de Andrade I, Mehta C, Bresnick EH. Post-transcriptional control of cellular differentiation by the RNA exosome complex. Nucleic Acids Research. 2020;48: 11913–11928. doi:10.1093/nar/gkaa883

10. Luan S, Luo J, Liu H, Li Z. Regulation of RNA decay and cellular function by 3′-5′ exoribonuclease DIS3L2. RNA Biology. 2019;16: 160–165. doi:10.1080/15476286.2018.1564466

11. Samuels TJ, Järvelin AI, Ish-Horowicz D, Davis I. Imp/IGF2BP levels modulate individual neural stem cell growth and division through myc mRNA stability. eLife. 2020;9: e51529. doi:10.7554/eLife.51529

12. Dubrulle J, Pourquié O. fgf8 mRNA decay establishes a gradient that couples axial elongation to patterning in the vertebrate embryo. Nature. 2004;427: 419–422. doi:10.1038/nature02216

13. Bonev B, Stanley P, Papalopulu N. MicroRNA-9 Modulates Hes1 Ultradian Oscillations by Forming a Double-Negative Feedback Loop. Cell Reports. 2012;2: 10–18. doi:10.1016/j.celrep.2012.05.017

14. Vastenhouw NL, Cao WX, Lipshitz HD. The maternal-to-zygotic transition revisited. Development. 2019;146: dev161471. doi:10.1242/dev.161471

15. Yartseva V, Giraldez AJ. The Maternal-to-Zygotic Transition During Vertebrate Development. Current Topics in Developmental Biology. Elsevier; 2015. pp. 191–232. doi:10.1016/bs.ctdb.2015.07.020

16. Pashler AL, Towler BP, Jones CI, Newbury SF. The roles of the exoribonucleases DIS3L2 and XRN1 in human disease. Biochemical Society Transactions. 2016;44: 1377–1384. doi:10.1042/BST20160107

17. Thomsen S, Anders S, Janga SC, Huber W, Alonso CR. Genome-wide analysis of mRNA decay patterns during early Drosophila development. Genome Biol. 2010;11: R93. doi:10.1186/gb-2010-11-9-r93

18. Brown AJP, Sagliocco FA. mRNA Abundance and Half-Life Measurements. Yeast Protocols. New Jersey: Humana Press; 1996. pp. 277–296. doi:10.1385/0-89603-319-8:277

19. Furlan M, de Pretis S, Pelizzola M. Dynamics of transcriptional and post-transcriptional regulation. Briefings in Bioinformatics. 2021;22: bbaa389. doi:10.1093/bib/bbaa389

20. Tani H, Akimitsu N. Genome-wide technology for determining RNA stability in mammalian cells: Historical perspective and recent advantages based on modified nucleotide labeling. RNA Biology. 2012;9: 1233–1238. doi:10.4161/rna.22036

21. Lugowski A, Nicholson B, Rissland OS. Determining mRNA half-lives on a transcriptome-wide scale. Methods. 2018;137: 90–98. doi:10.1016/j.ymeth.2017.12.006

22. Bahar Halpern K, Itzkovitz S. Single molecule approaches for quantifying transcription and degradation rates in intact mammalian tissues. Methods. 2016;98: 134–142. doi:10.1016/j.ymeth.2015.11.015

23. Boettiger AN, Levine M. Rapid Transcription Fosters Coordinate snail Expression in the Drosophila Embryo. Cell Reports. 2013;3: 8–15. doi:10.1016/j.celrep.2012.12.015

24. Lott SE, Villalta JE, Schroth GP, Luo S, Tonkin LA, Eisen MB. Noncanonical Compensation of Zygotic X Transcription in Early Drosophila melanogaster Development Revealed through Single-Embryo RNA-Seq. Hawley RS, editor. PLoS Biol. 2011;9: e1000590. doi:10.1371/journal.pbio.1000590

25. De Renzis S, Elemento O, Tavazoie S, Wieschaus EF. Unmasking Activation of the Zygotic Genome Using Chromosomal Deletions in the Drosophila Embryo. Kornberg T, editor. PLoS Biol. 2007;5: e117. doi:10.1371/journal.pbio.0050117

26. Heyn P, Kircher M, Dahl A, Kelso J, Tomancak P, Kalinka AT, et al. The Earliest Transcribed Zygotic Genes Are Short, Newly Evolved, and Different across Species. Cell Reports. 2014;6: 285–292. doi:10.1016/j.celrep.2013.12.030

27. Saunders A, Core LJ, Sutcliffe C, Lis JT, Ashe HL. Extensive polymerase pausing during Drosophila axis patterning enables high-level and pliable transcription. Genes & Development. 2013;27: 1146–1158. doi:10.1101/gad.215459.113

28. Wheeler JC, Shigesada K, Peter Gergen J, Ito Y. Mechanisms of transcriptional regulation by Runt domain proteins. Seminars in Cell & Developmental Biology. 2000;11: 369–375. doi:10.1006/scdb.2000.0184

29. Prudêncio P, Savisaar R, Rebelo K, Martinho RG, Carmo-Fonseca M. Transcription and splicing dynamics during early *Drosophila* development. RNA. 2022;28: 139–161. doi:10.1261/rna.078933.121

30. Pai AA, Henriques T, McCue K, Burkholder A, Adelman K, Burge CB. The kinetics of pre-mRNA splicing in the Drosophila genome and the influence of gene architecture. eLife. 2017;6: e32537. doi:10.7554/eLife.32537

31. Honkela A, Peltonen J, Topa H, Charapitsa I, Matarese F, Grote K, et al. Genome-wide modeling of transcription kinetics reveals patterns of RNA production delays. Proc Natl Acad Sci U S A. 2015;112: 13115–13120. doi:10.1073/pnas.1420404112

32. Lawrence ND, Sanguinetti G, Rattray M. Modelling Transcriptional Regulation Using Gaussian Processes. Proceedings of the 19th International Conference on Neural Information Processing Systems. Cambridge, MA, USA: MIT Press; 2006. pp. 785–792.

33. Deignan L, Pinheiro MT, Sutcliffe C, Saunders A, Wilcockson SG, Zeef LAH, et al. Regulation of the BMP Signaling-Responsive Transcriptional Network in the Drosophila Embryo. Perrimon N, editor. PLoS Genet. 2016;12: e1006164. doi:10.1371/journal.pgen.1006164

34. Bjorum SM, Simonette RA, Alanis R, Wang JE, Lewis BM, Trejo MH, et al. The Drosophila BTB Domain Protein Jim Lovell Has Roles in Multiple Larval and Adult Behaviors. Callaerts P, editor. PLoS ONE. 2013;8: e61270. doi:10.1371/journal.pone.0061270

35. Serpe M, Umulis D, Ralston A, Chen J, Olson DJ, Avanesov A, et al. The BMP-Binding Protein Crossveinless 2 Is a Short-Range, Concentration-Dependent, Biphasic Modulator of BMP Signaling in Drosophila. Developmental Cell. 2008;14: 940–953. doi:10.1016/j.devcel.2008.03.023

36. Calvo L, Ronshaugen M, Pettini T. smiFISH and embryo segmentation for single-cell multi-gene RNA quantification in arthropods. Commun Biol. 2021;4: 352. doi:10.1038/s42003-021-01803-0

37. Burow DA, Umeh-Garcia MC, True MB, Bakhaj CD, Ardell DH, Cleary MD. Dynamic regulation of mRNA decay during neural development. Neural Dev. 2015;10: 11. doi:10.1186/s13064-015-0038-6

38. Edgar BA, Odell GM, Schubiger G. A genetic switch, based on negative regulation, sharpens stripes inDrosophila embryos. Dev Genet. 1989;10: 124–142. doi:10.1002/dvg.1020100303

39. Hensman J, Rattray M, Lawrence ND. Fast Nonparametric Clustering of Structured Time-Series. IEEE Trans Pattern Anal Mach Intell. 2015;37: 383–393. doi:10.1109/TPAMI.2014.2318711

40. Mayr C. Regulation by 3′-Untranslated Regions. Annu Rev Genet. 2017;51: 171–194. doi:10.1146/annurev-genet-120116-024704

41. Hanson G, Coller J. Codon optimality, bias and usage in translation and mRNA decay. Nat Rev Mol Cell Biol. 2018;19: 20–30. doi:10.1038/nrm.2017.91

42. Eichhorn SW, Subtelny AO, Kronja I, Kwasnieski JC, Orr-Weaver TL, Bartel DP. mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in Drosophila oocytes and early embryos. eLife. 2016;5: e16955. doi:10.7554/eLife.16955

43. Burow DA, Martin S, Quail JF, Alhusaini N, Coller J, Cleary MD. Attenuated Codon Optimality Contributes to Neural-Specific mRNA Decay in Drosophila. Cell Reports. 2018;24: 1704–1712. doi:10.1016/j.celrep.2018.07.039

942 44. Adivarahan S, Livingston N, Nicholson B, Rahman S, Wu B, Rissland OS, et al. Spatial
943 Organization of Single mRNPs at Different Stages of the Gene Expression Pathway. Molecular
944 Cell. 2018;72: 727-738.e5. doi:10.1016/j.molcel.2018.10.010

945 45. Khong A, Parker R. mRNP architecture in translating and stress conditions reveals an ordered
946 pathway of mRNP compaction. Journal of Cell Biology. 2018;217: 4124–4140.
947 doi:10.1083/jcb.201806183

948 46. Vinter DJ, Hoppe C, Minchington TG, Sutcliffe C, Ashe HL. Dynamics of *hunchback* translation in
949 real-time and at single-mRNA resolution in the *Drosophila* embryo. Development. 2021;148:
950 dev196121. doi:10.1242/dev.196121

951 47. Horvathova I, Voigt F, Kotrys AV, Zhan Y, Artus-Revel CG, Eglinger J, et al. The Dynamics of
952 mRNA Turnover Revealed by Single-Molecule Imaging in Single Cells. Molecular Cell. 2017;68:
953 615-625.e9. doi:10.1016/j.molcel.2017.09.030

954 48. Stapel LC, Lombardot B, Broaddus C, Kainmueller D, Jug F, Myers EW, et al. Automated detection
955 and quantification of single RNAs at cellular resolution in zebrafish embryos. Development. 2015;
956 dev.128918. doi:10.1242/dev.128918

957 49. Oka Y, Sato TN. Whole-mount single molecule FISH method for zebrafish embryo. Sci Rep.
958 2015;5: 8571. doi:10.1038/srep08571

959 50. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA
960 molecules using multiple singly labeled probes. Nat Methods. 2008;5: 877–879.
961 doi:10.1038/nmeth.1253

962 51. Batish M, van den Bogaard P, Kramer FR, Tyagi S. Neuronal mRNAs travel singly into dendrites.
963 Proc Natl Acad Sci USA. 2012;109: 4645–4650. doi:10.1073/pnas.1111226109

964 52. Wang M, Ly M, Lugowski A, Laver JD, Lipshitz HD, Smibert CA, et al. ME31B globally represses
965 maternal mRNAs by two distinct mechanisms during the Drosophila maternal-to-zygotic transition.
966 eLife. 2017;6: e27891. doi:10.7554/eLife.27891

967 53. Patel PH, Barbee SA, Blankenship JT. GW-Bodies and P-Bodies Constitute Two Separate Pools
968 of Sequestered Non-Translating RNAs. Artero R, editor. PLoS ONE. 2016;11: e0150291.
969 doi:10.1371/journal.pone.0150291

970 54. Buszczak M, Paterno S, Lighthouse D, Bachman J, Planck J, Owen S, et al. The Carnegie Protein
971 Trap Library: A Versatile Tool for Drosophila Developmental Studies. Genetics. 2007;175: 1505–
972 1531. doi:10.1534/genetics.106.065961

973 55. Sankaranarayanan M, Emenecker RJ, Wilby EL, Jahnel M, Trussina IREA, Wayland M, et al.
974 Adaptable P body physical states differentially regulate bicoid mRNA storage during early
975 Drosophila development. Developmental Cell. 2021;56: 2886-2901.e6.
976 doi:10.1016/j.devcel.2021.09.021

977 56. Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, et al. The developmental
978 transcriptome of Drosophila melanogaster. Nature. 2011;471: 473–479. doi:10.1038/nature09715

57. Uvarovskii A, Naarmann-de Vries IS, Dieterich C. On the optimal design of metabolic RNA labeling experiments. Erhard F, editor. PLoS Comput Biol. 2019;15: e1007252. doi:10.1371/journal.pcbi.1007252

58. Furlan M, Galeota E, Gaudio ND, Dassi E, Caselle M, de Pretis S, et al. Genome-wide dynamics of RNA synthesis, processing, and degradation without RNA metabolic labeling. Genome Res. 2020;30: 1492–1507. doi:10.1101/gr.260984.120

59. Edgar BA, Weir MP, Schubiger G, Kornberg T. Repression and turnover pattern fushi tarazu RNA in the early Drosophila embryo. Cell. 1986;47: 747–754. doi:10.1016/0092-8674(86)90517-9

60. Little SC, Tikhonov M, Gregor T. Precise Developmental Gene Expression Arises from Globally Stochastic Transcriptional Activity. Cell. 2013;154: 789–800. doi:10.1016/j.cell.2013.07.025

61. Ferree PL, Deneke VE, Di Talia S. Measuring time during early embryonic development. Seminars in Cell & Developmental Biology. 2016;55: 80–88. doi:10.1016/j.semcdb.2016.03.013

62. Fukaya T, Lim B, Levine M. Rapid Rates of Pol II Elongation in the Drosophila Embryo. Current Biology. 2017;27: 1387–1391. doi:10.1016/j.cub.2017.03.069

63. Dufourt J, Bellec M, Trullo A, Dejean M, De Rossi S, Favard C, et al. Imaging translation dynamics in live embryos reveals spatial heterogeneities. Science. 2021;372: 840–844. doi:10.1126/science.abc3483

64. Levine M, Davidson EH. Gene regulatory networks for development. Proceedings of the National Academy of Sciences. 2005;102: 4936–4942. doi:10.1073/pnas.0408031102

65. Stathopoulos A, Levine M. Genomic Regulatory Networks and Animal Development. Developmental Cell. 2005;9: 449–462. doi:10.1016/j.devcel.2005.09.005

66. Ermolenko DN, Mathews DH. Making ends meet: New functions of mRNA secondary structure. WIREs RNA. 2021;12. doi:10.1002/wrna.1611

67. Bazzini AA, Viso F, Moreno-Mateos MA, Johnstone TG, Vejnar CE, Qin Y, et al. Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. EMBO J. 2016;35: 2087–2103. doi:10.15252/embj.201694699

68. He F, Celik A, Wu C, Jacobson A. General decapping activators target different subsets of inefficiently translated mRNAs. eLife. 2018;7: e34409. doi:10.7554/eLife.34409

69. Gobet C, Weger BD, Marquis J, Martin E, Neelagandan N, Gachon F, et al. Robust landscapes of ribosome dwell times and aminoacyl-tRNAs in response to nutrient stress in liver. Proc Natl Acad Sci USA. 2020;117: 9630–9641. doi:10.1073/pnas.1918145117

70. Sheth U, Parker R. Decapping and Decay of Messenger RNA Occur in Cytoplasmic Processing Bodies. Science. 2003;300: 805–808. doi:10.1126/science.1082320

71. Buddika K, Huang Y-T, Ariyapala IS, Butrum-Griffith A, Norrell SA, O'Connor AM, et al. Coordinated repression of pro-differentiation genes via P-bodies and transcription maintains Drosophila intestinal stem cell identity. Current Biology. 2021; S0960982221015918. doi:10.1016/j.cub.2021.11.032

72. Jones CI, Zabolotskaya MV, Newbury SF. The 5′ → 3′ exoribonuclease XRN1/Pacman and its functions in cellular processes and development: The 5′ → 3′ exoribonuclease XRN1/Pacman and its functions. WIREs RNA. 2012;3: 455–468. doi:10.1002/wrna.1109

73. Hubstenberger A, Courel M, Bénard M, Souquere S, Ernoult-Lange M, Chouaib R, et al. P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. Molecular Cell. 2017;68: 144-157.e5. doi:10.1016/j.molcel.2017.09.003

74. Bhattacharyya SN, Habermacher R, Martine U, Closs EI, Filipowicz W. Relief of microRNA-Mediated Translational Repression in Human Cells Subjected to Stress. Cell. 2006;125: 1111–1124. doi:10.1016/j.cell.2006.04.031

75. Brengues M, Teixeira D, Parker R. Movement of Eukaryotic mRNAs Between Polysomes and Cytoplasmic Processing Bodies. Science. 2005;310: 486–489. doi:10.1126/science.1115791

76. Kosman D, Mizutani CM, Lemons D, Cox WG, McGinnis W, Bier E. Multiplex Detection of RNA Expression in *Drosophila* Embryos. Science. 2004;305: 846–846. doi:10.1126/science.1099247

77. Hoppe C, Bowles JR, Minchington TG, Sutcliffe C, Upadhyai P, Rattray M, et al. Modulation of the Promoter Activation Rate Dictates the Transcriptional Response to Graded BMP Signaling Levels in the Drosophila Embryo. Developmental Cell. 2020;54: 727-741.e7. doi:10.1016/j.devcel.2020.07.007

78. Tsanov N, Samacoits A, Chouaib R, Traboulsi A-M, Gostan T, Weber C, et al. smiFISH and FISH-quant – a flexible single RNA detection approach with super-resolution capability. Nucleic Acids Res. 2016;44: e165–e165. doi:10.1093/nar/gkw784

79. Trcek T, Lionnet T, Shroff H, Lehmann R. mRNA quantification using single-molecule FISH in Drosophila embryos. Nat Protoc. 2017;12: 1326–1348. doi:10.1038/nprot.2017.030

80. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34: 525–527. doi:10.1038/nbt.3519

81. BinTayyash N, Georgaka S, John ST, Ahmed S, Boukouvalas A, Hensman J, et al. Non-parametric modelling of temporal and spatial counts data from RNA-seq experiments. Mathelier A, editor. Bioinformatics. 2021;37: 3788–3795. doi:10.1093/bioinformatics/btab486

82. Lee S, Zhang AY, Su S, Ng AP, Holik AZ, Asselin-Labat M-L, et al. Covering all your bases: incorporating intron signal from RNA-seq data. NAR Genomics and Bioinformatics. 2020;2: lqaa073. doi:10.1093/nargab/lqaa073

83. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. 2009;10: 48. doi:10.1186/1471-2105-10-48

84. Presnyak V, Alhusaini N, Chen Y-H, Martin S, Morris N, Kline N, et al. Codon Optimality Is a Major Determinant of mRNA Stability. Cell. 2015;160: 1111–1124. doi:10.1016/j.cell.2015.02.029

85. Larkin A, Marygold SJ, Antonazzo G, Attrill H, dos Santos G, Garapati PV, et al. FlyBase: updates to the *Drosophila melanogaster* knowledge base. Nucleic Acids Research. 2021;49: D899–D907. doi:10.1093/nar/gkaa1026

Supplementary Data

Supplementary Figures S1-S12

Supplementary Table 1: Table of half-lives

Supplementary Table 2: Table of smFISH probes used in this study

Supplementary Table 3: Intronic and zygotic clusters and IDs

Supplementary Methods: Filtering of the genes, model formulation and Bayesian inference with MCMC, simulation studies

Supplementary Figures

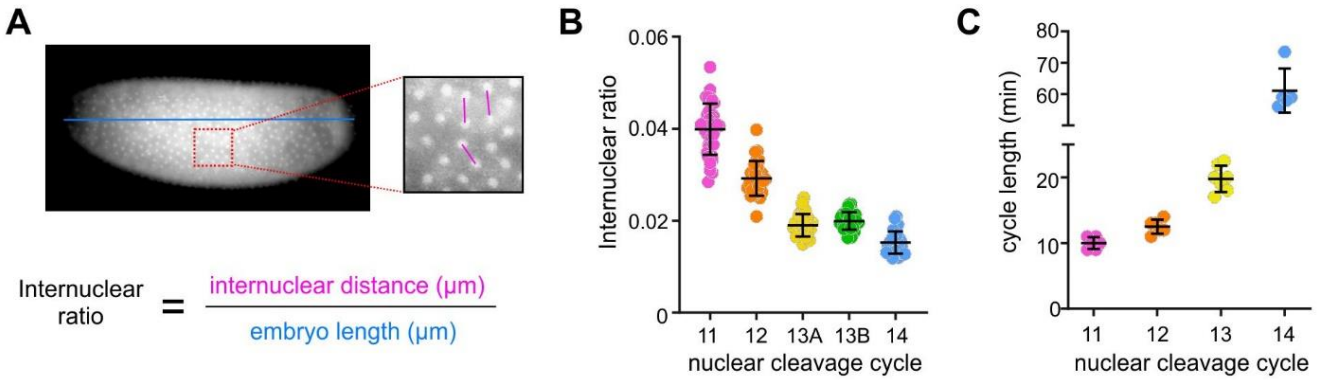


Figure S1. Staging of early embryos for RNA-seq. (A) Images of embryos were captured immediately prior to collection and the internuclear distance and embryo length were measured for each to give an internuclear ratio. (B) The internuclear ratio at nuclear cleavage cycles (nc) can be used to accurately stage embryos. (C) Timing of each nc at 20°C used in experiments.

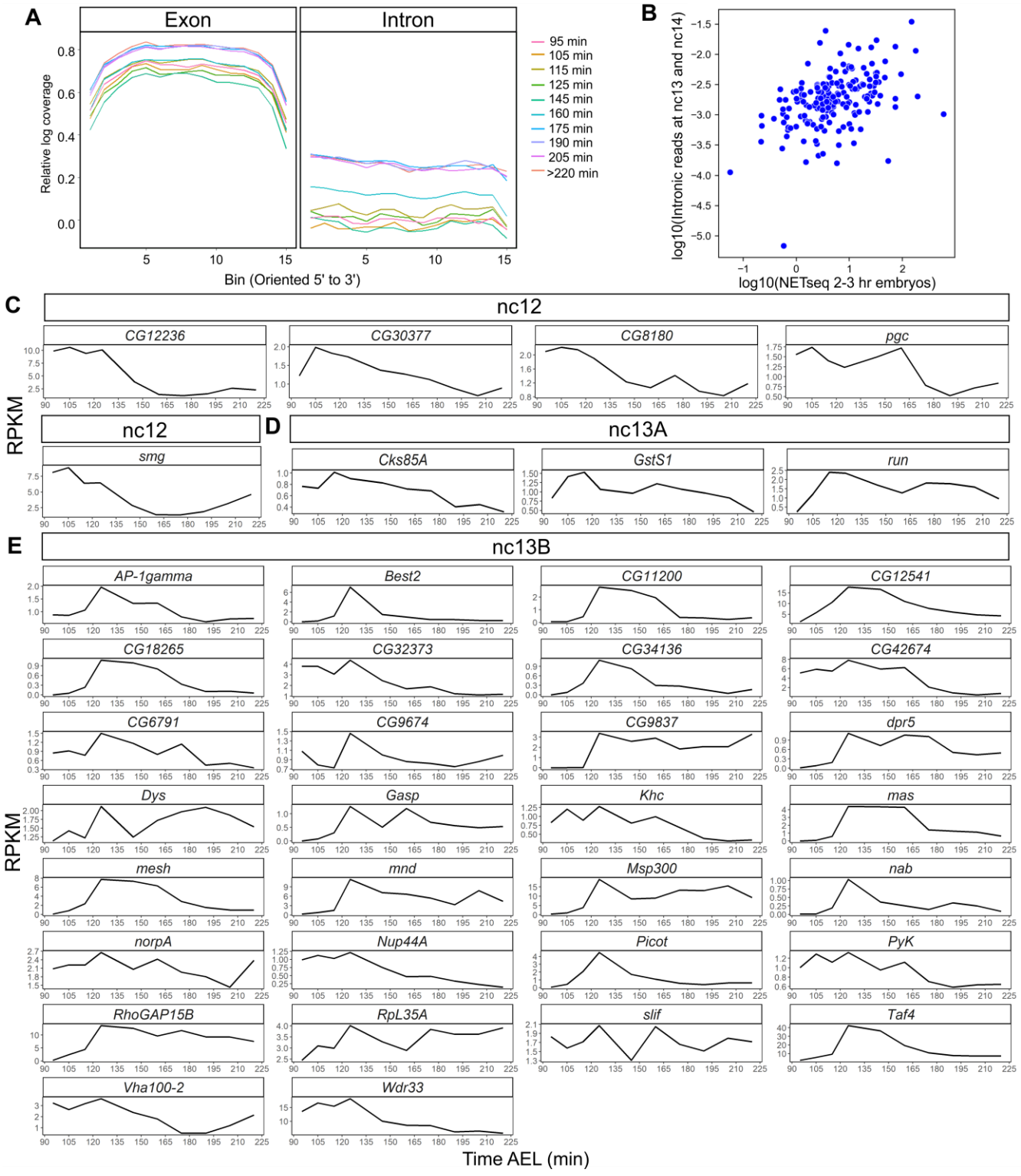


Figure S2. Read coverage and pre-mRNA profiles of early zygotic genes. (A) Binned read coverage across exons and introns for one replicate for each of the time points within the time series. (B) Scatterplot of our RNAseq reads versus NET-seq read counts showing a relationship (Spearman's Rank correlation $\rho = 0.46$, $p = 5.1 \times 10^{-11}$). Intronic read counts (RPKM) of genes that show early transcription

1088 in our dataset, grouped depending on peak expression at (C) nc12 (105 min), (D) nc13A (115 min) or
1089 (E) nc13B (125 min), based on timings at 20°C.

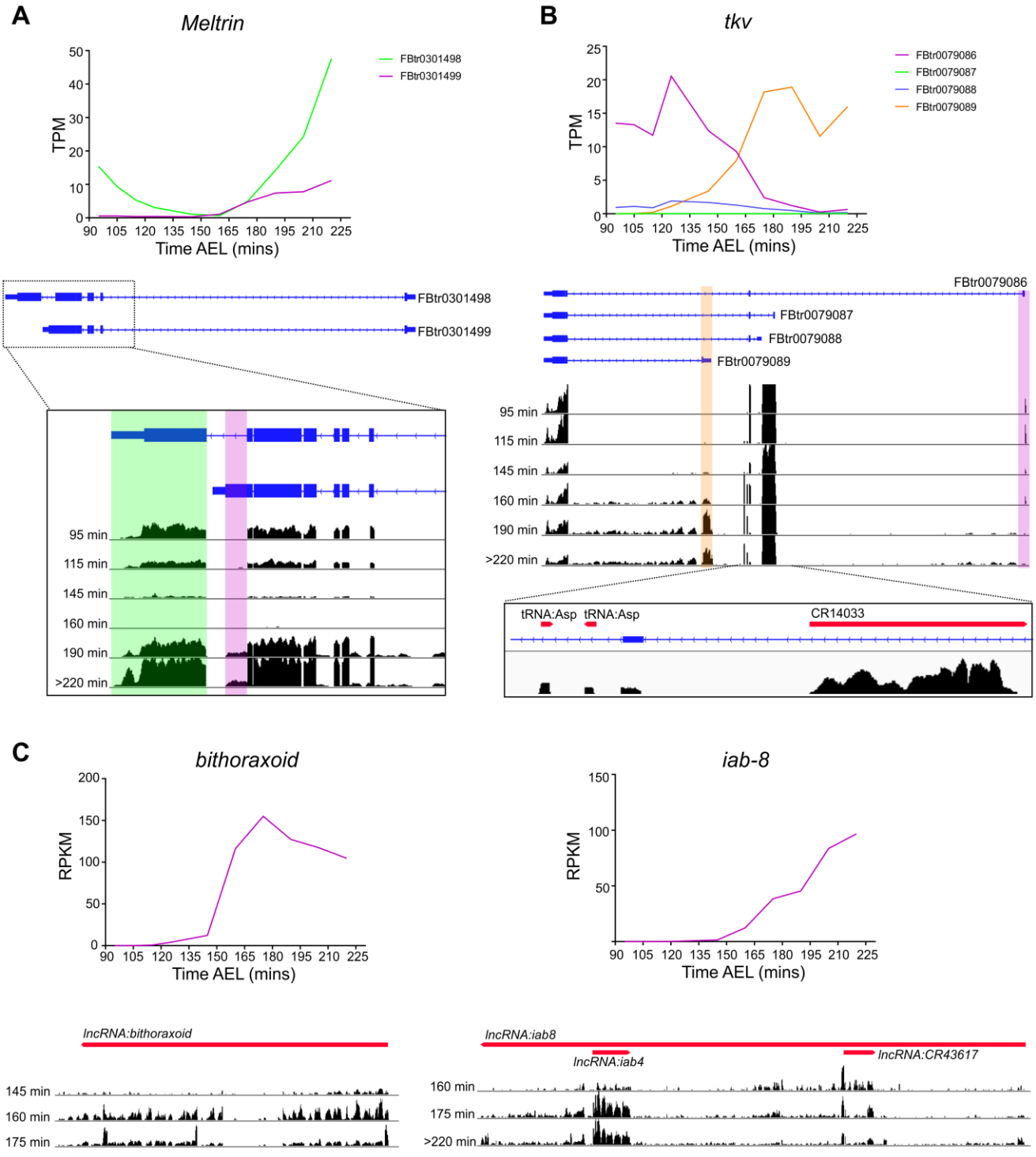


Figure S3. Examples of alternative isoform and non coding RNA expression during development in *Drosophila* embryos. (A) Quantification of the two different transcripts (TPM) of the *Meltrin* gene. The FBtr0301499 isoform (purple) is not detected in embryos <160 min AEL but is detected at later time points in addition to the FBtr0301498 (green) transcript. Mapped reads are shown below with a region highlighted in purple depicting the increase in the alternative transcript at time point 190 and 220 min. (B) As in (A) but for the gene *tkv.* Transcription of the FBtr0079086 (purple) and FBtr0079089 (orange)

isoforms switches during the time course of development, as highlighted on the mapped reads below. Expression of non coding RNAs, including two copies of the tRNA-Asp and a pseudogene (CR14033), overlap the *tkv* locus as shown in the expanded region below from time point 220 min. (C) Expression of the non-coding RNAs *bithoraxoid* and *iab-8.* Gene level counts (RPKM) show dynamic expression across the time course of these two non-coding genes. Gene regions for the ncRNAs (red) are shown with the genome browser tracks below.

A

Gene	Transcript ID	*D*	MAP half-life (mins)	Lower credible interval (mins)	Upper credible interval (mins)
lov	FBtr0072458	0.0752	9	6	15
cv-2	FBtr0071610	0.0246	28	17	937

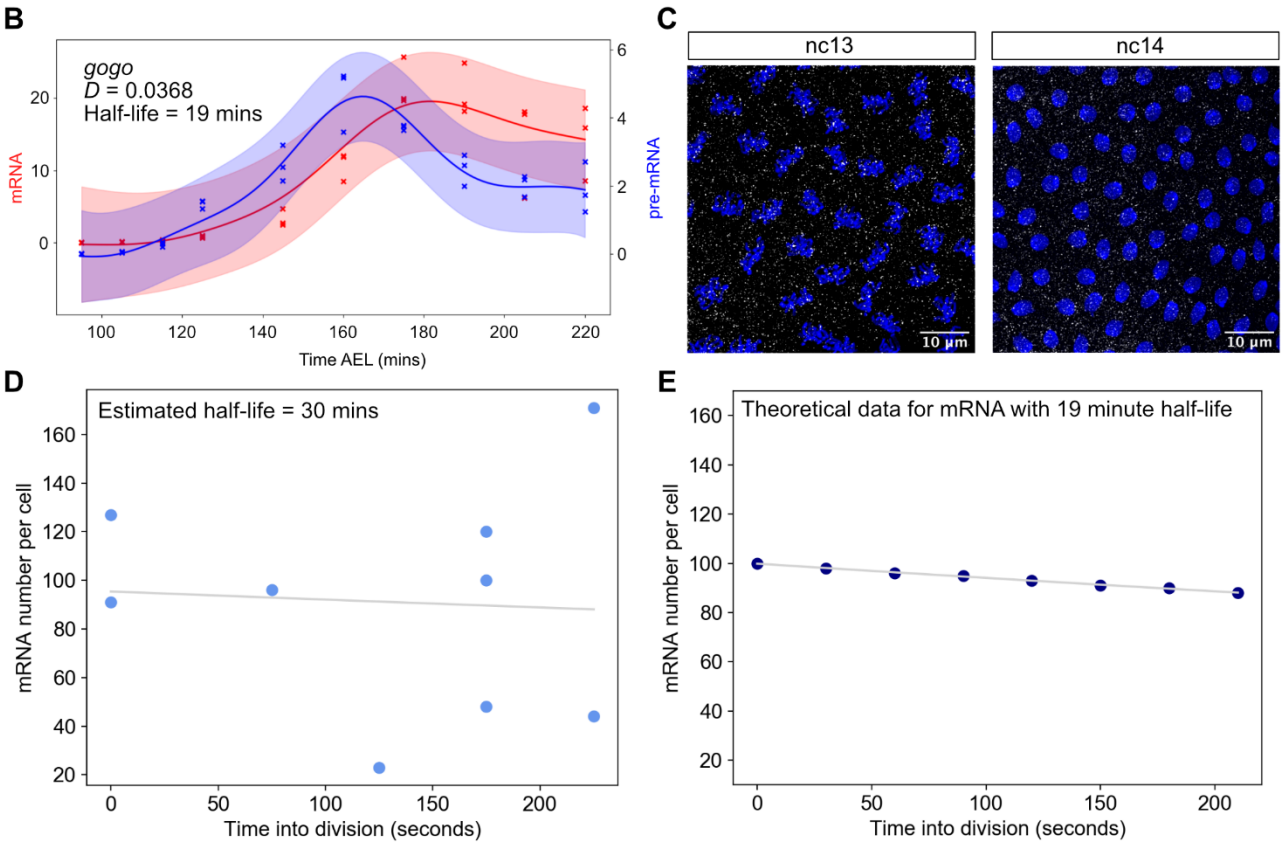


Figure S4: High embryo to embryo variation in transcript numbers masks variation due to degradation over cell division. (A) Parameter estimates for degradation rates and credible regions of short and long half-life genes shown in Figure 2D. (B) RNA-seq data for *gogo* fitted with the Gaussian process model. A half-life of 19 minutes is inferred for this gene. (C) Confocal images of pre- and post-division during the 13th and 14th nuclear cycles showing nuclei stained with DAPI (blue) and single mRNAs (white). (D) Data for the number of mRNAs per cell for embryos at various time points during the cell division, fitted with an exponential function from which a half-life of 30 min was calculated. The data has a low signal to noise ratio of 0.0013, meaning that the reduction in transcript numbers over time due to degradation (signal) is much smaller than the natural embryo to embryo variation in transcript numbers (noise). (E) Theoretical data demonstrating the reduction in transcript numbers that would be expected for a mRNA with a 19 min half-life over a 210 second time frame, which is a reduction of ~10%.

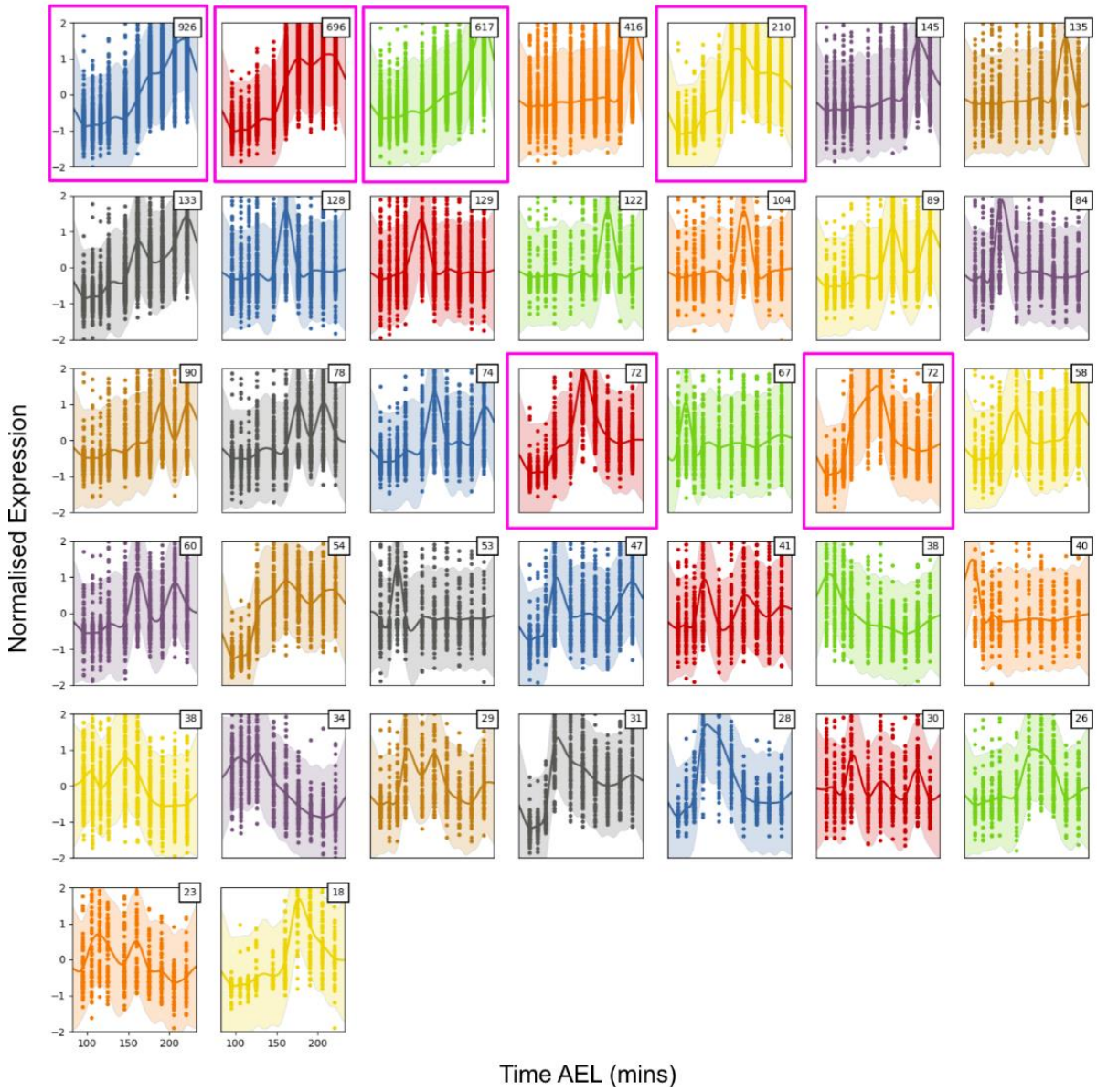


Figure S5: Output from clustering of 5035 intronic transcripts using GPclust. Data for all pre-mRNAs in the cluster are shown with shaded credible regions and inferred function as a solid line. The number of pre-mRNAs in each cluster is shown in the top right corner of each plot and graphs are arbitrarily coloured. Clusters that show interesting dynamics and contain high numbers of pre-mRNAs (1,2,3,5,18,20) are highlighted and also displayed in Figure 3Ai.

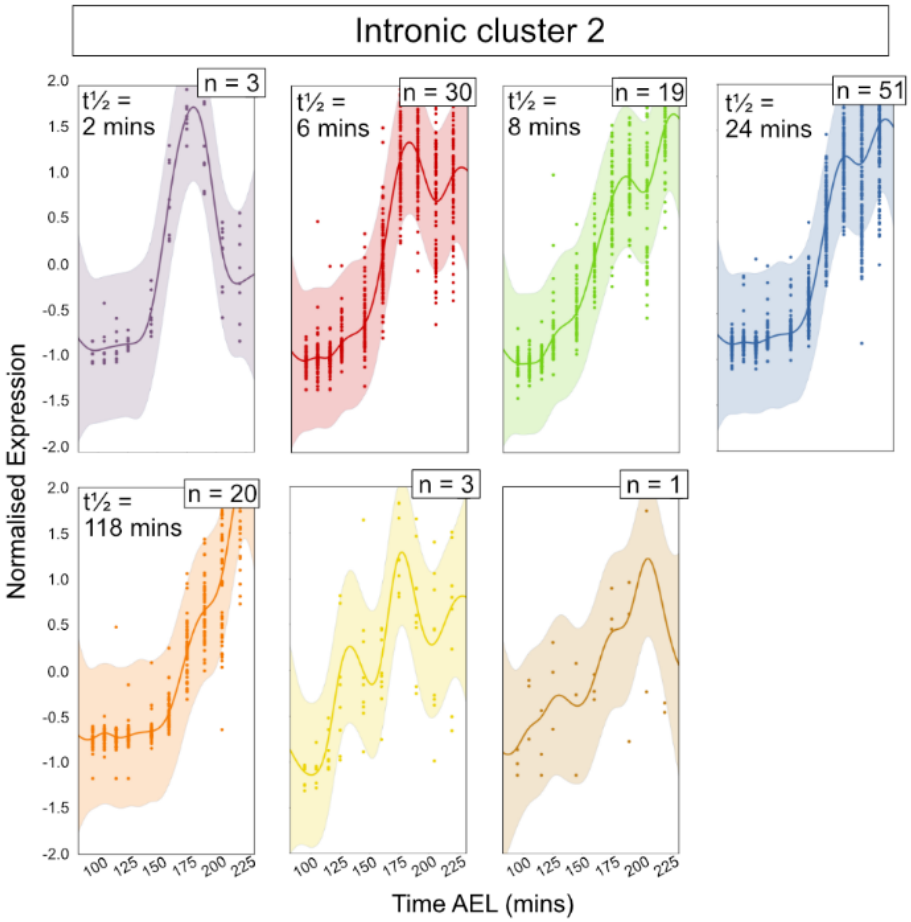


Figure S6: Sub-clustering of zygotic transcripts with intronic data in intronic cluster 2 using GPclust. Data for all mRNAs in the cluster are shown with shaded credible regions and inferred function as a solid line for intronic cluster 2. The number of mRNAs in each cluster is shown in the top right corner of each plot and graphs are arbitrarily coloured. The mean half-lives of the transcripts in the cluster are shown where the cluster has >2 transcripts with estimated half-lives.

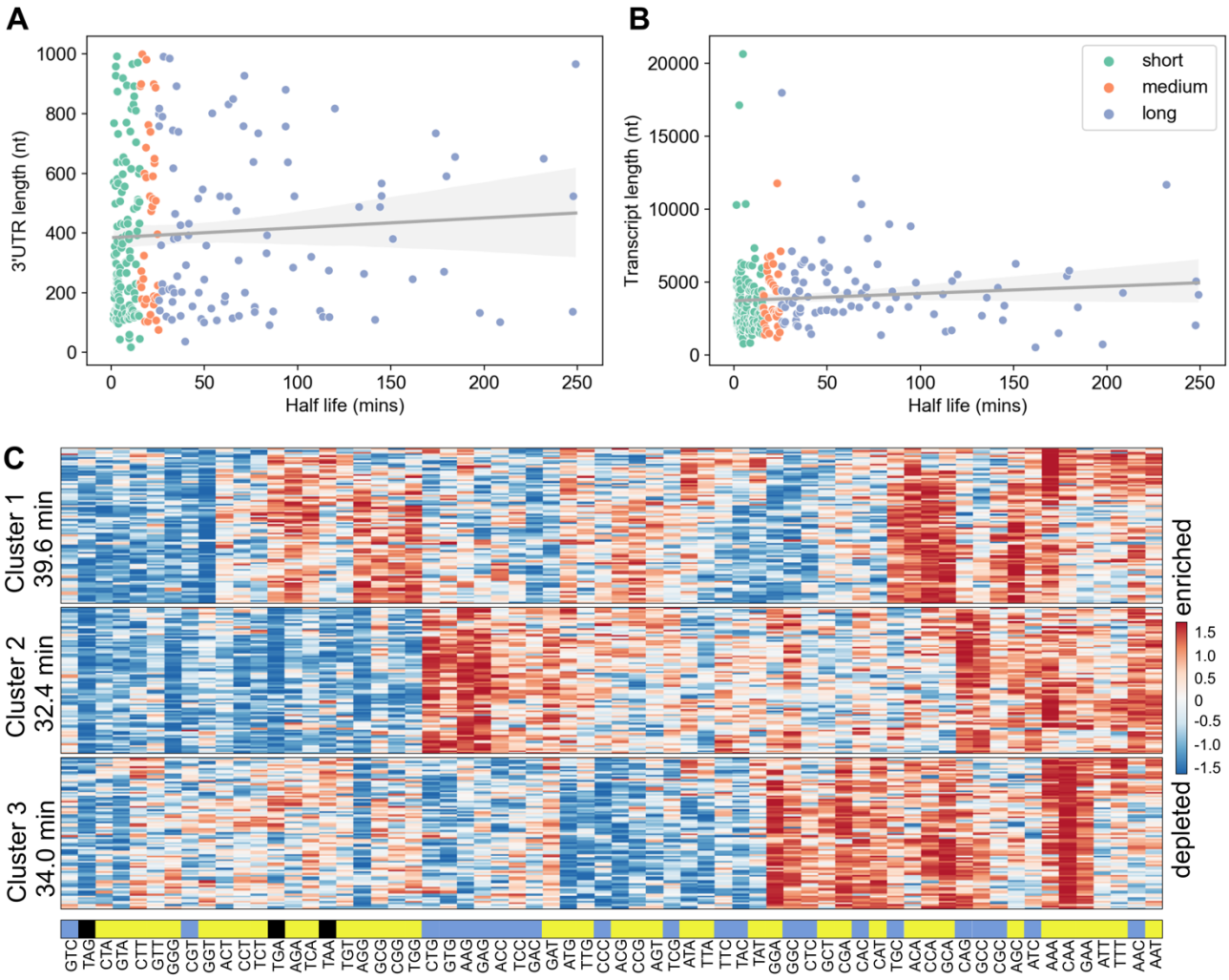


Figure S7: mRNA properties and stability. (A) Graph showing the correlation between mRNA half-life and 3' UTR length. Data fit with linear regression model, Pearson's r = 0.06, p = 0.34. (B) Correlation between mRNA total length and half-life. Transcripts are coloured by half-life category short (green), medium (orange) or long (blue). Data fit with linear regression model, Pearson's r = 0.09, p = 0.13. (C) Transcripts were clustered based on similar codon usage using K-means and the level of enrichment of each codon within the transcript is shown in the heatmap. The average half-life of each cluster is shown to the left. Codons are coloured by whether they are optimal (blue) vs non-optimal (yellow) with stop codons shown in black.

A

Transcript half-lives used in smFISH experiments

Gene	Transcript ID	Half-life	Category
otd	FBtr0331305	3	short
Neu2	FBtr0332718	5	short
lov	FBtr0072458/FBtr0330231	9/9	short
Notum	FBtr0075435	12	short
gogo	FBtr0074848	19	medium
ec	FBtr0112802/FBtr0070606	18/20	medium
Dfd	FBtr0081621	22	medium
byn	FBtr0332071/FBtr0332070	22/23	medium
cv-2	FBtr0071610	28	long
pxb	FBtr0083162	47	long
ths	FBtr0088030	54	long
ltl	FBtr0332144	249	long

B

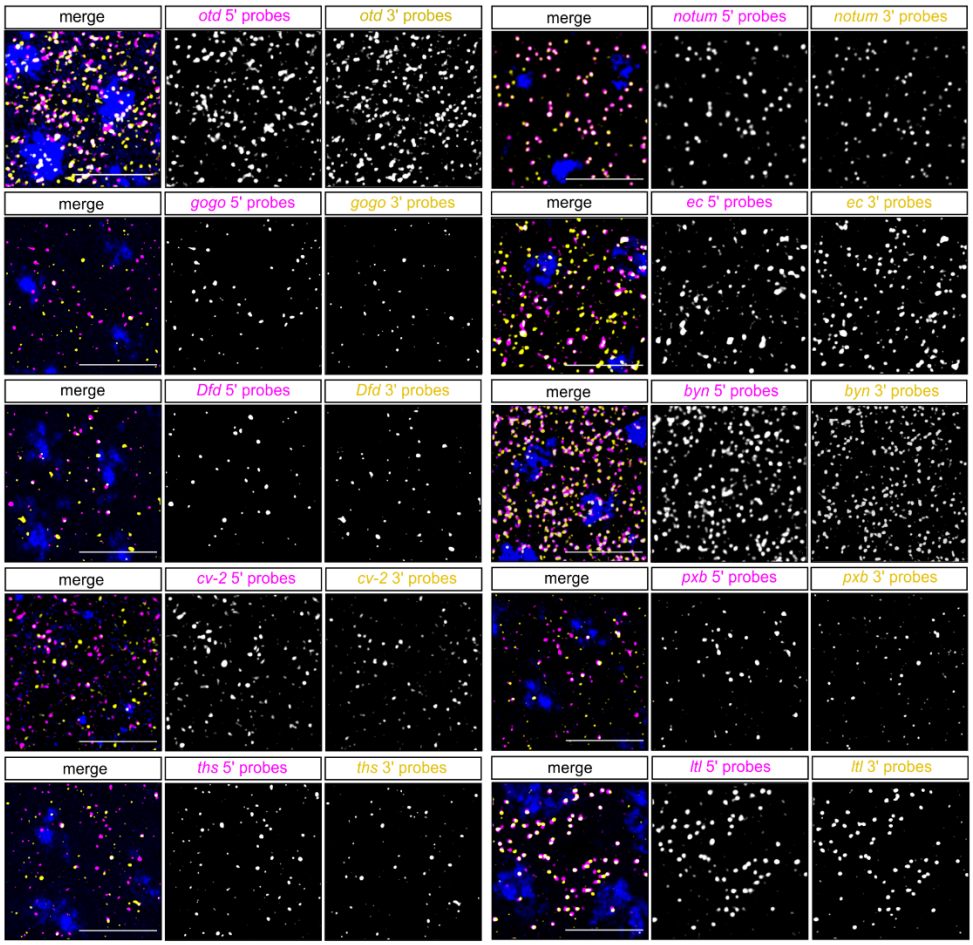
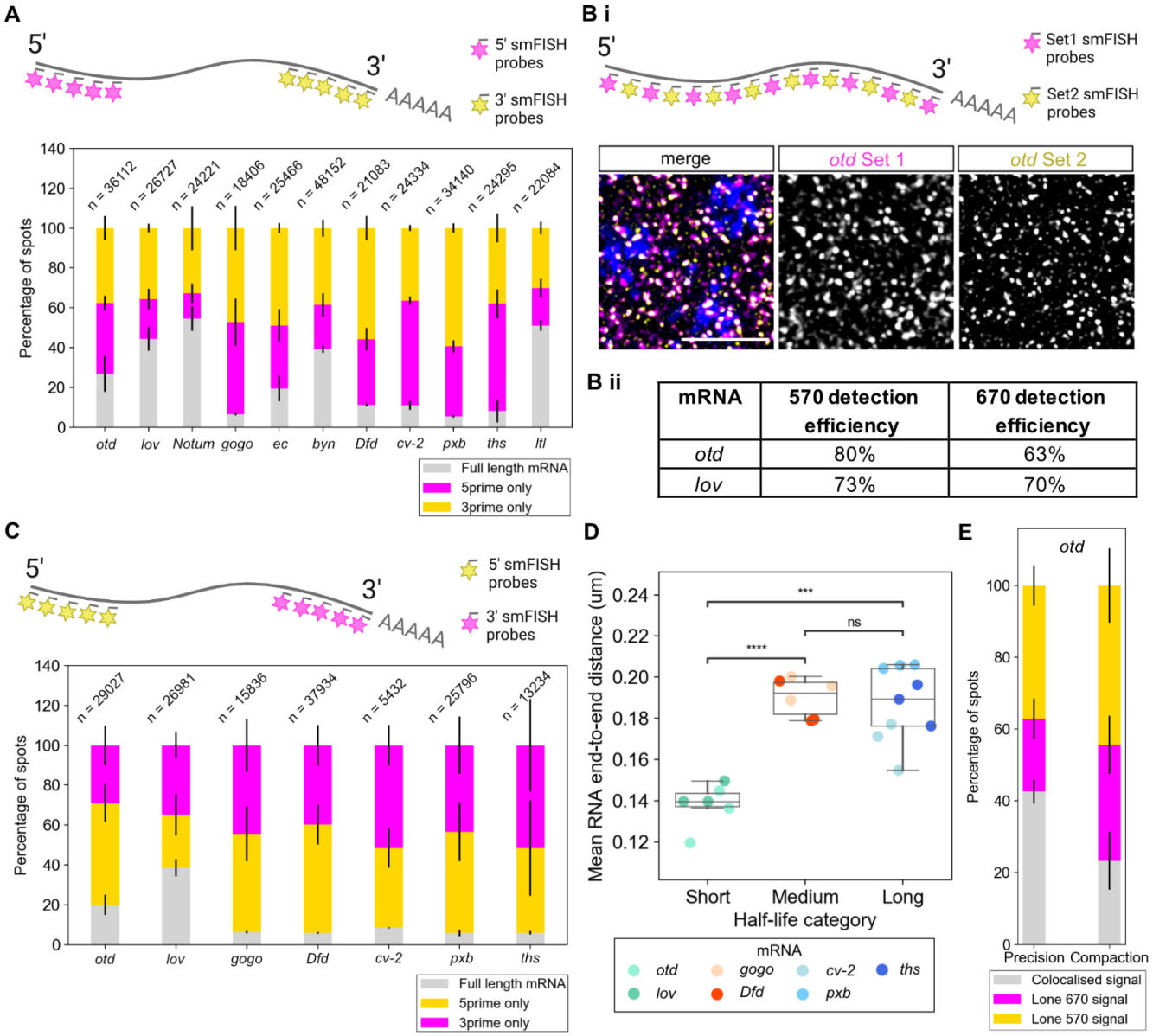


Figure S8: Compaction of mRNAs in early embryos. (A) A table of the transcripts used in smFISH experiments. (B) Confocal images of fixed embryos showing smFISH detection of the 5' (magenta) and 3' (yellow) ends of the mRNAs used in the compaction experiments quantitated in Figure 4F.



mRNA	570 detection efficiency	670 detection efficiency
otd	80%	63%
lov	73%	70%

Figure S9: Control experiments for smFISH. (A) Proportions of 5' (670) and 3' (570) ends and whole mRNAs detected in smFISH experiments for all the compaction forward data described in Figure 4D-F. Numbers represent total mRNAs in image (whole, 5' only and 3' only). Schematic illustrates the forward staining scheme. (B) (i) Schematic shows full length *otd* mRNAs detected using probes with alternating labels used as a precision control. Confocal images of fixed embryos stained with alternating smFISH probes. All images are maximum projections of 7 Z slices. Scale bars: 5 µm. (ii) Table of detection efficiencies for the two mRNAs used in precision control experiments. (C) Proportions of 5' (570) and 3' (670) ends and whole mRNAs detected in smFISH experiments for reverse data (with switched fluorophores). Numbers represent total mRNAs in image as in A. (D) Graph shows the end-to-end distances of mRNAs with different stabilities in the reverse fluorophore experiment to Figure 4D, data are shown in boxplots for each half-life category with n = 3 embryos for each transcript. (E) Proportion of *otd* signals detected in either the 570, 670 or colocalised signal in the precision or both compaction experiments.

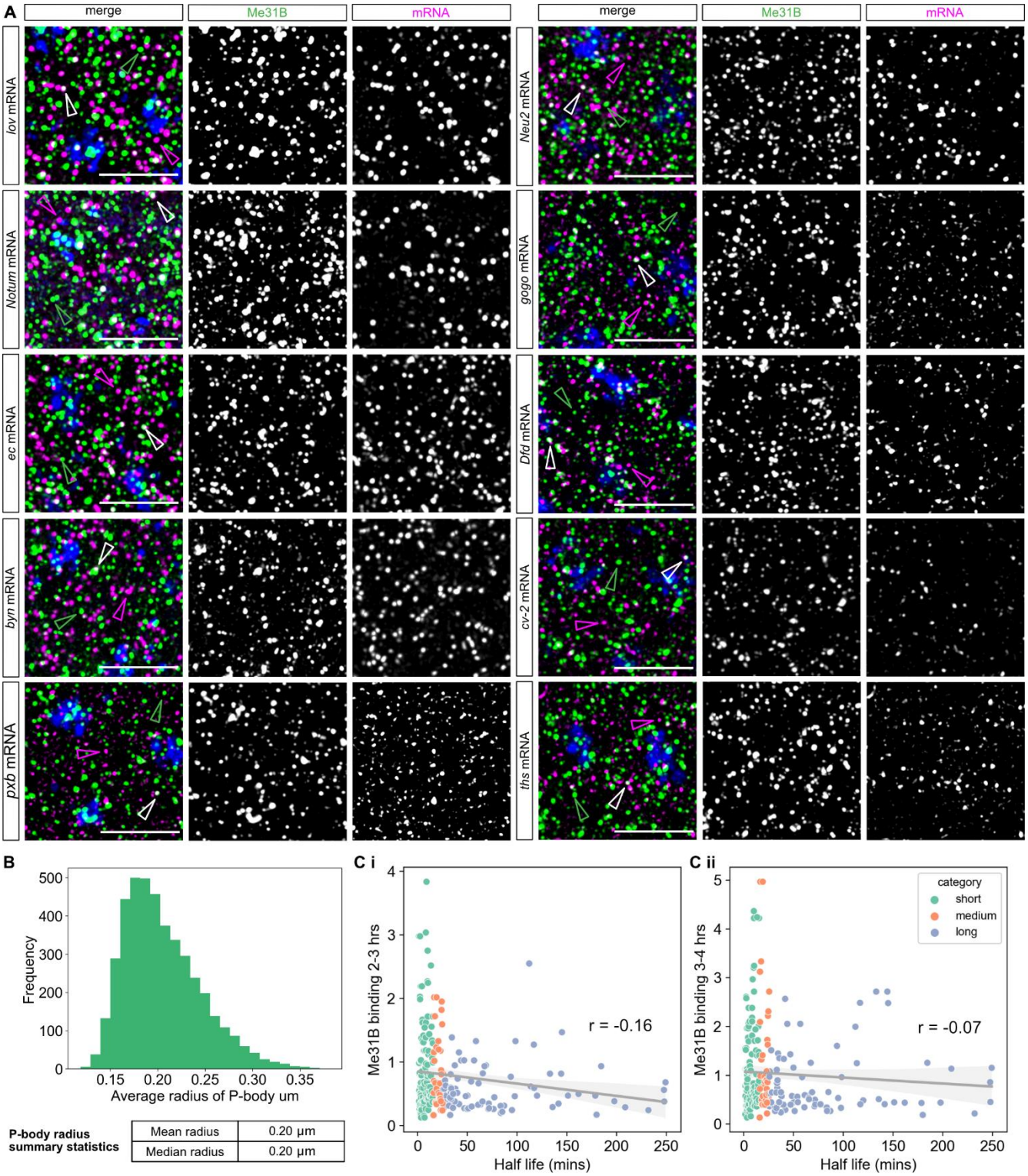


Figure S10: Confocal images of mRNAs with various stabilities and their colocalisation with the P-body marker Me31B. (A) Confocal images of early nc14 Me31B-GFP embryos showing smFISH staining of the indicated test mRNAs. DAPI marks the nuclei (blue), mRNAs are shown in magenta (magenta arrowhead), GFP-Me31B marking P-bodies in green (green arrowhead) and mRNAs

colocalising with P-bodies in white (white arrowhead). All images are MIP of 7 Z slices of a confocal image. Scale bar: 5µm (B) Analysis of P-body sizes reveals an average radius of 0.2 µm. (C) Analysis of correlation between Me31B binding from (i) 2-3 hour embryos and (ii) 3-4 hr embryos with model half-life. A significant negative correlation ($r = -0.16$, $p = 0.014$) is found at 2-3 hours but no significant relationship is found at 3-4 hours ($r = -0.07$, $p = 0.30$).

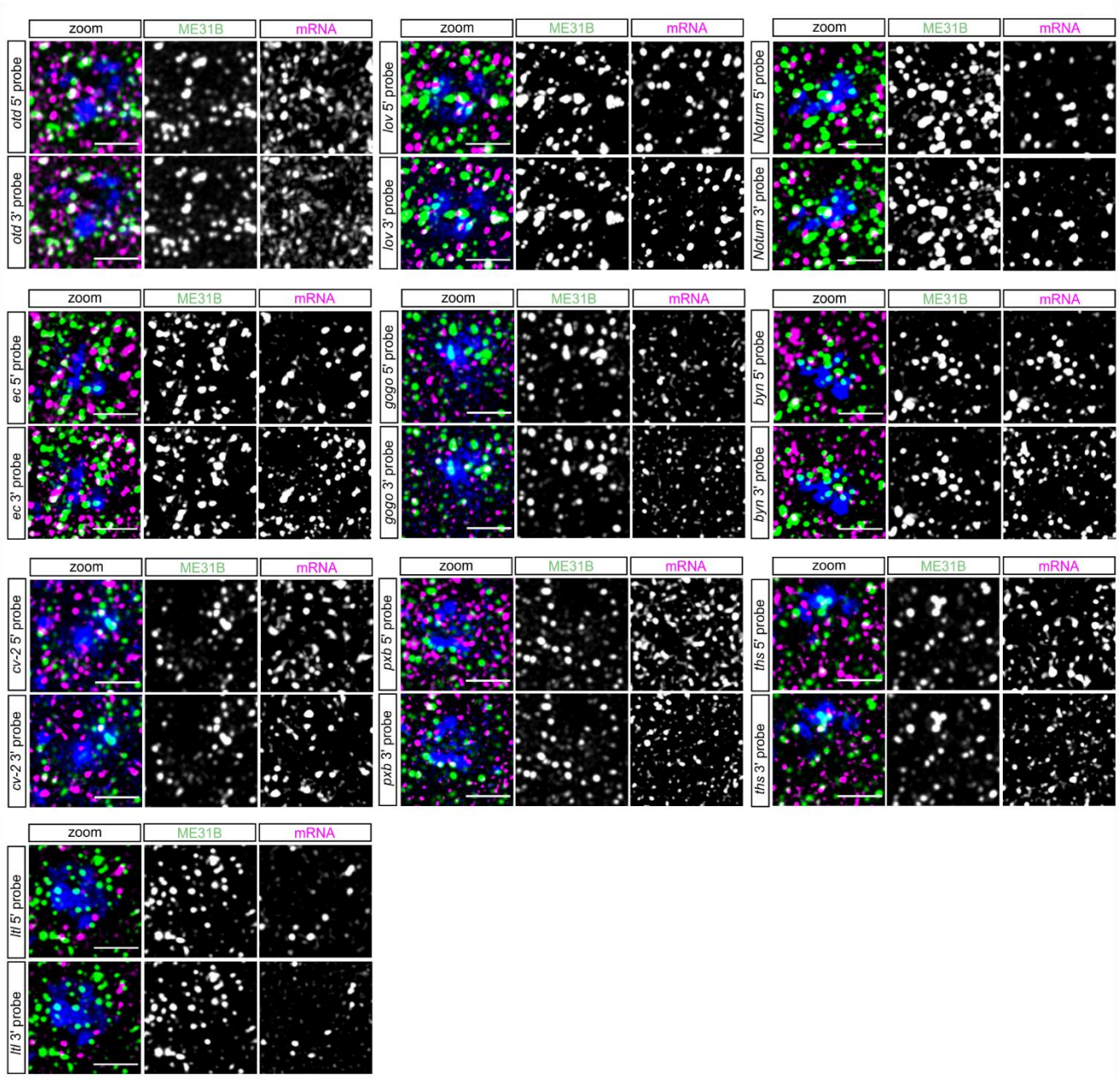


Figure S11: Confocal images of mRNAs and their 5' and 3' end colocalisation with the P-body marker Me31B. Confocal images of early nc14 Me31B-GFP embryos showing smFISH staining of the indicated test mRNAs with 5' and 3' probe sets. For each test mRNA, the same region of the embryo is shown with the 5' and 3' mRNA probe sets separately for clarity. mRNAs are shown in magenta, GFP-Me31B marking P-bodies in green and DAPI labelling nuclei in blue. All images are MIP of 7 Z slices of a confocal image. Scale bar: 2µm.

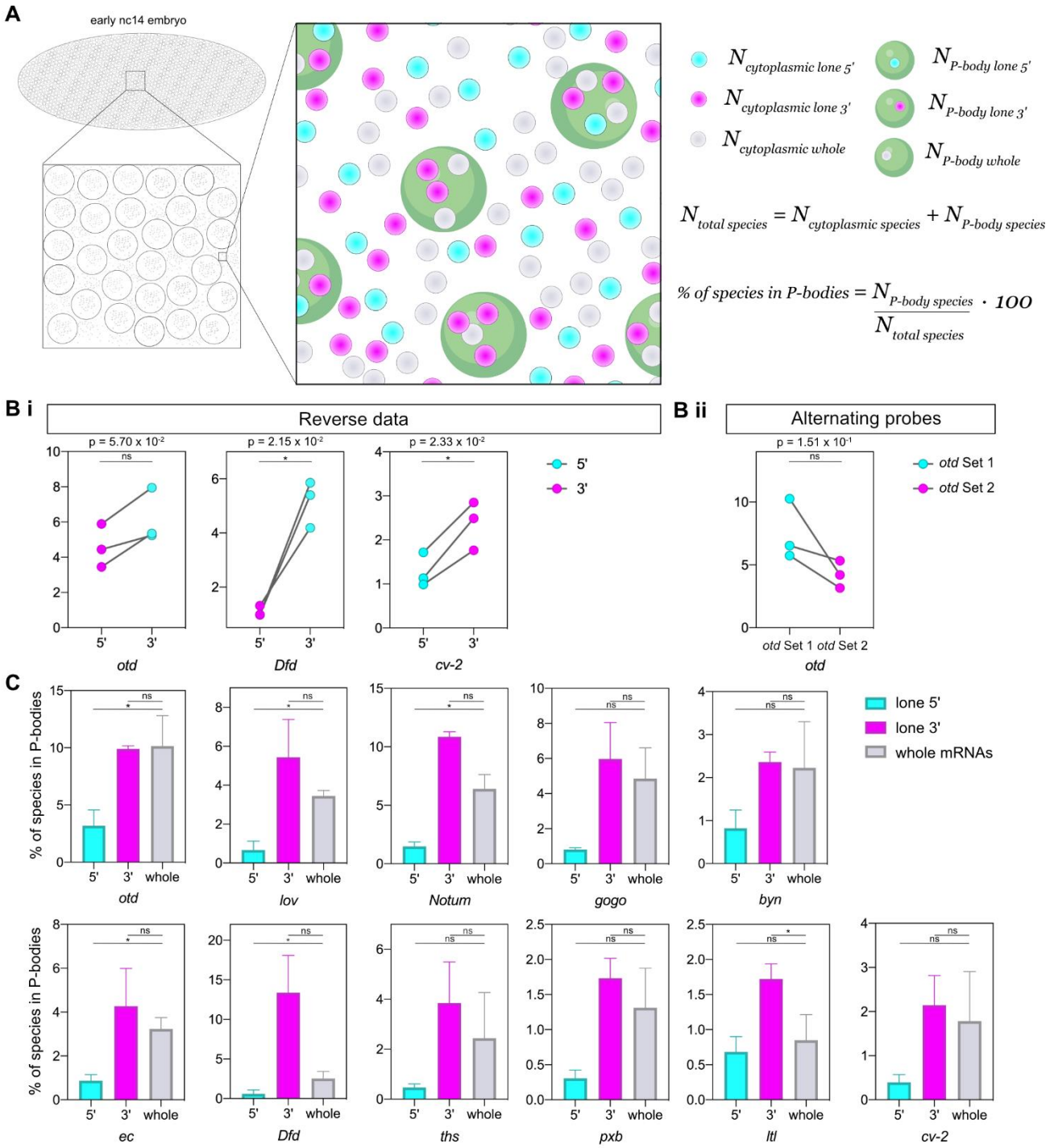


Figure S12: Proportions of lone 5', lone 3' and whole mRNAs colocalised with P-bodies. (A) Schematic demonstrating quantitation of the proportion of each species in P-bodies. 5' and 3' spots are detected and then paired (see methods) to give 3 species - whole mRNAs (grey), lone 5' ends (cyan) and lone 3' ends (magenta). The number of each species that colocalises with P-bodies is divided by the total number of that species to give a percentage enrichment in P-bodies. (Bi) Quantification of the percentage of unpaired mRNA 5' and 3' ends in P-bodies relative to the total number of lone 5' or 3' ends in the switched probe fluorophore experiments for *otd, Dfd* and *cv-2* (see also Figure 6C). Paired t-test used to determine significance. (Bii) (C) Percentages of each species (whole, lone 5' and lone 3')

in P-bodies across the test set of mRNAs. mRNAs are ordered by their half-life from the shortest (*otd*) to the longest (*cv-2*) half-life. One-way repeated measures ANOVA used to determine significance with $\alpha = 0.05$.