Modelling global mRNA dynamics during *Drosophila* embryogenesis reveals a relationship between mRNA degradation and P-bodies

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

Abstract

Regulation of mRNA degradation is critical for a diverse array of cellular processes and developmental cell fate decisions. Although many methods for determining mRNA half-lives rely on transcriptional inhibition or metabolic labelling, these manipulations can influence half-lives or introduce errors in the measurements. Here we use a non-invasive method for estimating mRNA half-lives globally in the early *Drosophila* embryo using 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. Our data suggest that mRNA half-life is not dependent on translation efficiency, but instead we provide evidence that short half-life mRNAs are more strongly associated 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. The abundance of mRNAs in a cell depends not only on the transcription rate but also on the stability of each mRNA. In eukaryotic cells, the stability of mRNAs is controlled by two major pathways of mRNA degradation: Xrn1 endonuclease-mediated 5'-3' decay and exosome catalysed 3'-5' decay (Mugridge et al., 2018; Weick and Lima, 2021). 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 (Ivanov et al., 2019; Standart and Weil, 2018). mRNA stability is also commonly regulated by sequences in the 3' UTR, including binding sites for RNA binding proteins or for miRNAs (Mayya and Duchaine, 2019). 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 (Morris et al., 2021).

Regulation of mRNA degradation is essential for diverse cellular processes including proliferation, differentiation, apoptosis and immune responses (Akira and Maeda, 2021; Akiyama et al., 2021; Fraga de Andrade et al., 2020; Luan et al., 2019; Mugridge et al., 2018). Control of mRNA degradation 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 (Samuels et al., 2020), an *fgf8* mRNA gradient generated by mRNA decay couples differentiation to posterior elongation of the vertebrate embryonic axis (Dubrulle and Pourquié, 2004) and *Hes1* mRNA instability is integral to the Hes1 protein ultradian oscillations that may act as a timer for vertebrate neuronal differentiation (Bonev et al., 2012). In addition to these mRNA-specific examples, a conserved feature of early embryogenesis is that there is bulk degradation of maternal mRNAs around the time of zygotic genome activation (Vastenhouw et al., 2019; Yartseva and Giraldez, 2015). Consistent with the key roles of mRNA stability in cell biology, mutations in many components of the degradation pathways are associated with human diseases (Fraga de Andrade et al., 2020; Pashler et al., 2016).

While the half-lives of strictly maternal mRNAs during embryogenesis can be readily measured genome-wide (Thomsen et al., 2010), measuring the decay of zygotic mRNAs is more difficult due to ongoing transcription. One approach is to inhibit transcription, either chemically or using temperature sensitive alleles, and then follow the decline in mRNA levels over time (Brown and Sagliocco, 1996; Furlan et al., 2021; Tani and Akimitsu, 2012). A disadvantage of this approach is that inhibiting transcription disrupts cellular physiology which can lead to the stabilisation of some mRNAs (Furlan et al., 2021; Tani and Akimitsu, 2012). Other methods involve metabolic labelling of the RNA with a nucleoside analogue, for example in pulse-chase or approach-to-equilibrium experiments (Furlan et al., 2021; Lugowski et al., 2018; Tani and Akimitsu, 2012). Related approaches use computational models to estimate transcription and degradation rates by sequencing both the total and labelled RNA following the pulse (Furlan et al., 2021). However, the choice of pulse-chase labelling times can introduce errors in the half-life estimates (Uvarovskii et al., 2019). In addition, efficient metabolic labelling can be difficult *in vivo*, especially on the short timescales needed for high resolution measurements. Single molecule fluorescent in situ hybridisation (smFISH) imaging based methods for estimating mRNA half-lives have also been described. However, these require steady-state transcript levels (Bahar Halpern and Itzkovitz, 2016) or a natural shut off of transcription (Boettiger and Levine, 2013) and these imaging based approaches are not high throughput.

In this study we generate a high-resolution total RNA-seq time series across early embryogenesis that we then use to estimate half-lives and assign mRNAs into different stability classes. Our data support a role for P-bodies in mRNA degradation, as we find that unstable mRNAs are more highly colocalised with P-bodies and for some mRNAs we can detect fragments with only the 3’ end 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. RNA was isolated from single early *Drosophila* embryos at 10 time points, starting at nuclear cycle (nc) 11 prior to the onset of bulk zygotic transcription through to the beginning of gastrulation (Figure 1A). As in a previous study (Lott et al., 2011), to stage the embryos we used a His2avRFP transgenic line to visualise nuclei during the rapid nuclear cleavage cycles of early *Drosophila* development. Embryos were precisely staged at nc11, nc12, nc13 and nc14 by calculating an internuclear distance ratio (Figure S1A). Single embryos were collected in triplicate at 5 min (nc11, nc12 and nc13) and 15 min (nc13 and nc14) intervals after the division and finally at cephalic furrow formation (Figure 1A). As male and female embryos have differences in X chromosome transcription due to dosage compensation (Lott et al., 2011), we used PCR to determine the sex of each embryo and only included female embryos as biological repeats. 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 are useful for quantifying nascent, unspliced transcripts and also allow us to detect early zygotic expression by distinguishing zygotic transcripts containing introns from maternally loaded spliced mRNAs.

Based on the mapped reads we detected a total of 18159 transcripts during early embryogenesis representing 9026 unique genes. Using a 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 is consistent with the biological age of the embryos explaining the majority of variation within the data rather than other differences between individual 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 time points 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*, *bcd*), maternal and zygotic (*Neu3* and *da*), early zygotic (*upd1* and *dpp*) and late zygotic expression (*wg* and *hnt*) (Figure 1D).

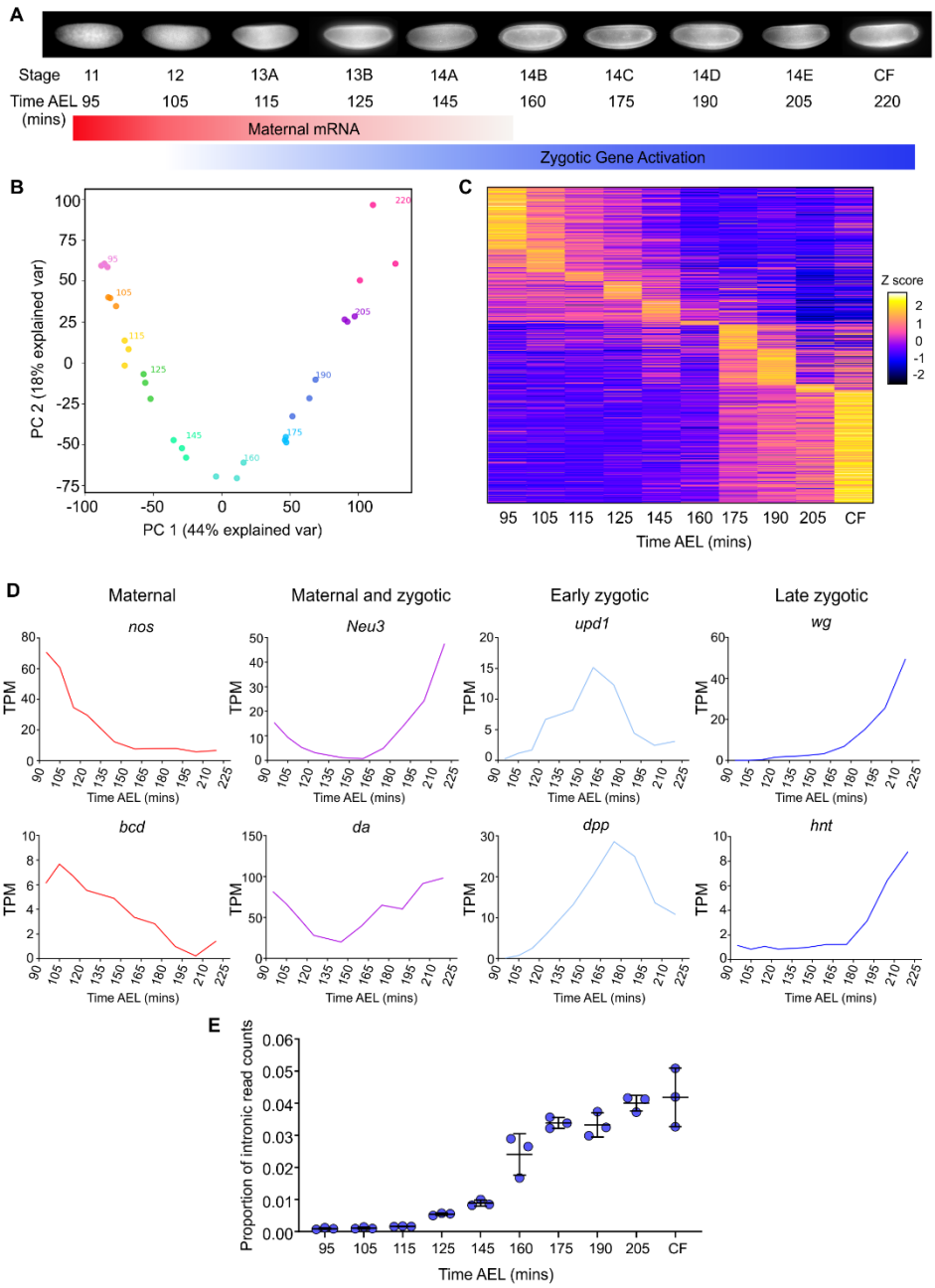

Figure 1. Total RNA-seq captures dynamic gene expression during early *Drosophila* development. (A) RNA-seq samples were collected at 10 different developmental time points spanning early nc11 through to cephalic furrow (CF) formation. Approximate time after egg lay (AEL) at 20°C for each time point is shown, along with a schematic representation of the maternal to zygotic transition where the embryo switches from relying on the 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 of the series. (C) A heat map 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) Examples of transcript dynamics captured by the data include maternal, maternal and zygotic, early zygotic and late zygotic gene expression. (E) Proportion of normalised intron reads (RPKM) throughout the time series.

As we sequenced total RNA, we were able to determine the number of reads for each transcript that mapped to introns (where appropriate) as well as exons. Only a very small proportion of transcripts at time points between 105-125 min have intronic reads (Figure 1E). These data suggest that during these stages (nc12 to early nc13) there is only minor zygotic transcription of intron-containing genes. Previous studies have shown that the earliest zygotic activation of the *Drosophila* genome is biased towards expression of short intronless genes (De Renzis et al., 2007; Heyn et al., 2014), which we cannot distinguish from maternally deposited transcripts at any individual early time point in our data. In addition, the early nuclear cycles are short, limiting the time period of active transcription. Nonetheless, we found that eight genes had detectable levels of intron signal at nc12 and nc13A, indicative of early zygotic transcription (Figure S2A-B).

The proportion of intronic reads increases significantly at 135 min (Figure S2C), then there is a further large increase around mid-nc14 (Figure 1E), when bulk activation of zygotic transcription occurs (Lott et al., 2011). We detect 7276 zygotically expressed genes, which is similar to a previous estimate based on GRO-seq data (Saunders et al., 2013). The benefit of the high temporal resolution of our data can be seen in examples of transient peak expression such as the gene *runt* (*run*) which peaks in early nc13 and is also expressed earlier at nc12 (Figure S2B). *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 (Wheeler et al., 2000). Additionally, we observe temporal changes in mRNA isoforms during development, exemplified by the genes *Meltrin* and *thickveins* (*tkv*) (Figure S3A and B). In both cases, these isoforms have altered coding sequences, which for the BMP receptor Tkv results in a shorter extracellular ligand-binding domain in the zygotically expressed isoform. We are also able to detect the 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

Previous studies have focused on clearance of maternal transcripts (Thomsen et al., 2010), therefore we were interested in the kinetics by which zygotic transcripts 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). Using the profiles of intronic and exonic reads over the time course, we fit a model of mRNA accumulation and degradation in order to estimate the mRNA degradation rate.

We then used a Gaussian process (GP) regression model (Honkela et al., 2015; Lawrence et al., 2006) to estimate zygotic transcript half-lives from the intronic (pre-mRNA) and exonic (mature

mRNA) time series RNA-seq data (Figure 2B). Before fitting the GP regression, we applied a dynamic filter where we a computed 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. From these, we filtered further to select transcripts with a correlation between the mRNA and pre-mRNA above 0.4. 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, which describes the evolution of the mature mRNA levels over time, as a function of the pre-mRNA dynamics, with parameters including splicing and degradation rates which are inferred using the GP regression.

The model provides half-life estimates for 279 zygotic transcripts corresponding to 187 genes (Supplementary Table 1). The distribution of these, coloured by long, medium or short half-life category can be seen in Figure 2C, with the mean half-life at 42 minutes and median at 31 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. *RhoU* mRNA, encoding a Rho GTPase, has a short half-life of just 7 minutes whereas the *cv-2* mRNA, encoding a BMP binding protein, has a long half-life of 85 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 (Bahar Halpern and Itzkovitz, 2016) was unsuitable for validation of half-lives in this particular system. An alternative method exploited the aborting 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 (Boettiger and Levine, 2013). 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-D). As a result, any reduction due to degradation is masked by high embryo to embryo variation, as has previously been observed for mRNA numbers for other genes in the *Drosophila* embryo (Calvo et al., 2021). The *snail* mRNA numbers are tightly controlled by negative autoregulation (Boettiger and Levine, 2013), suggesting that the *snail* mRNA may be uniquely suited to this method for calculating half-life.

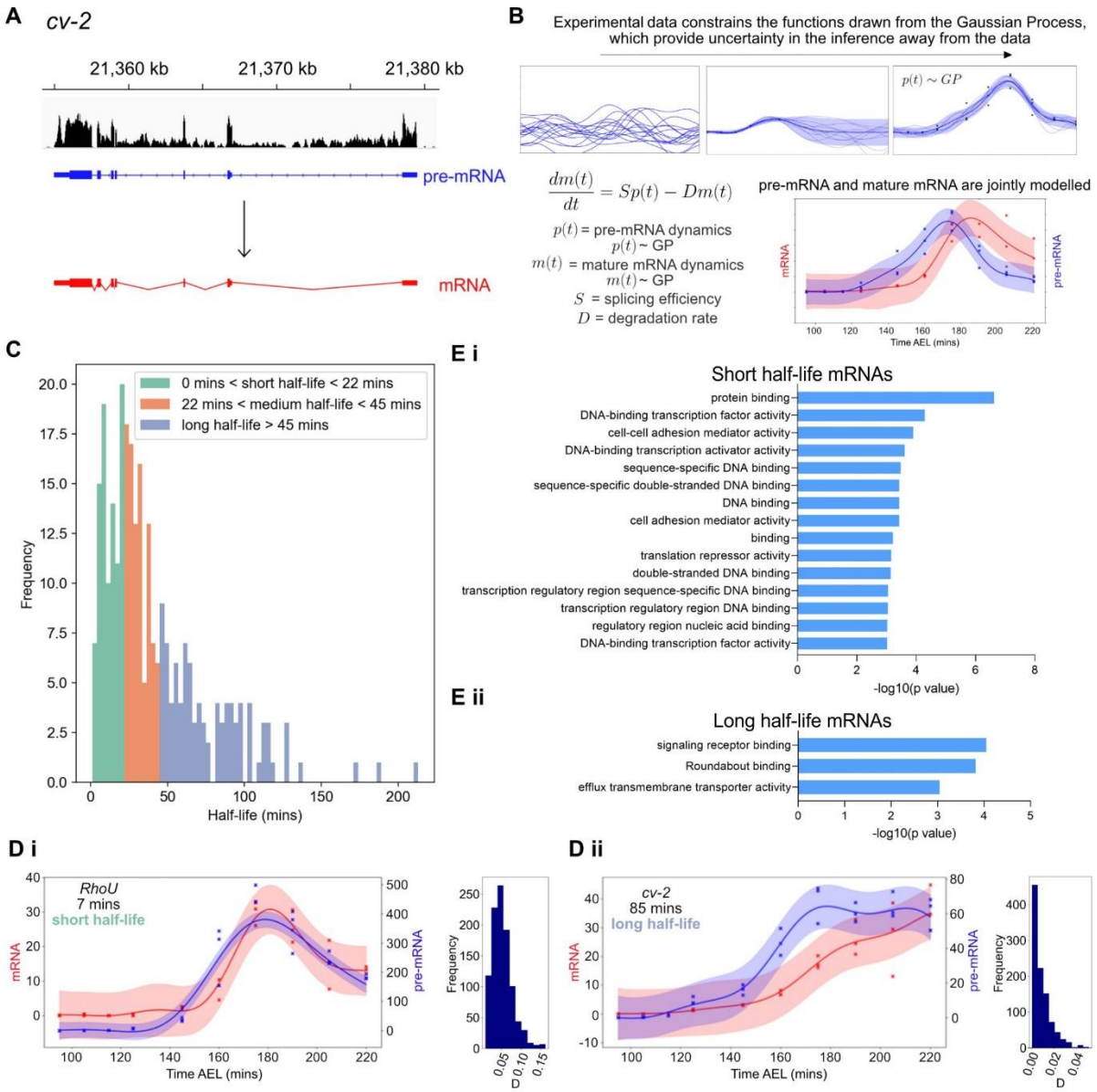

Figure 2: Gaussian process regression provides estimates of transcript half-lives in early embryogenesis. (A) Reads aligned to intronic regions of genes are used to represent pre-mRNA levels (blue). Transcript-level exonic reads represent mature mRNA levels (red). (B) Schematic of Gaussian Process (GP) regression and the model implemented in this study. The ODE underlying the model is displayed, showing how 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 279 transcripts estimated using the Gaussian process model. Transcripts are divided into short, medium and long half-lives from the quantiles of the data and coloured accordingly. (D) Examples of data for a short (i) and a long half-life mRNA (ii), *RhoU* and *cv-2*, 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. (E) Gene ontology analysis of groups of short (i) and long (ii) half-life mRNAs compared to the full set of dynamic genes in the dataset.

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 that are compatible with their inferred stabilities. We therefore used gene ontology (GO) analysis to investigate the functions of proteins encoded by the sets of short and long half-life mRNAs (Figure 2Ei). By comparing the GO terms associated with short half-life mRNAs compared to all dynamic transcripts in the RNA-seq data, we find an enrichment of mRNAs encoding DNA-binding proteins, transcription factors and cell adhesion proteins. Transcription factors and other DNA binding proteins have previously been identified as encoded by unstable mRNAs (Burow et al., 2015; Edgar et al., 1989; Thomsen et al., 2010). The GO analysis also suggests that stable mRNAs encode signalling receptors and transmembrane transporters (Figure 2Eii). Overall, this approach has allowed the classification of transcripts into short, medium and long half-life categories where groups of mRNAs are enriched for protein functions reflected by their different stabilities.

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 time series data were clustered using GPclust (Figure S5), a package specifically designed for clustering noisy time series data using Gaussian processes (Hensman et al., 2015). From the set of intronic clusters (Figure S5), 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, another well populated cluster, also display a range of mature mRNA dynamics and are shown in Figure S6. The Gaussian process 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.

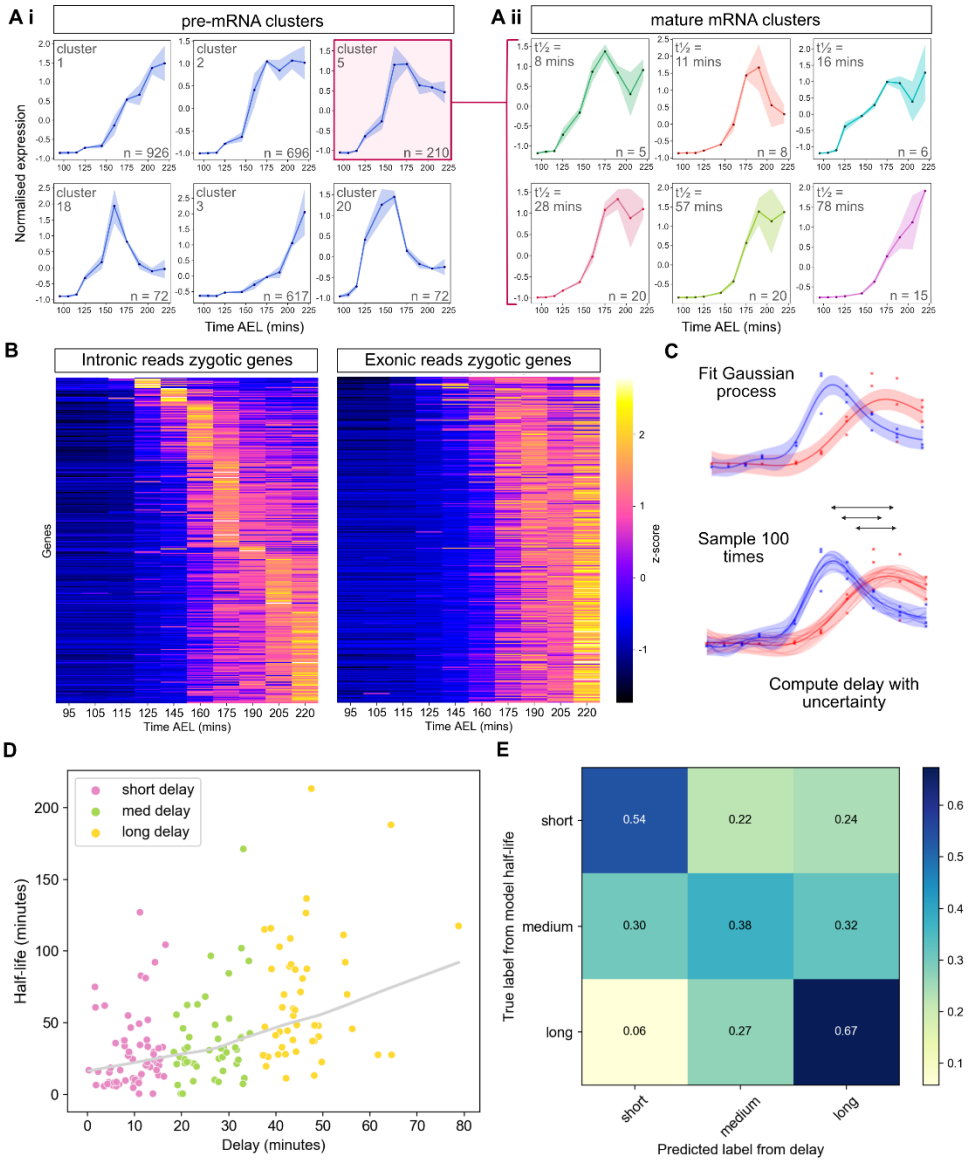

Figure 3: Degradation regulates mature mRNA dynamics and can be approximated using the time delay between peak pre-mRNA and mature mRNA expression. (A) Plots show the clustering of pre-mRNA (Ai) and mature (Aii) mRNA time series using GPclust. Plots show the mean expression value at each time point for 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 of the plot, the number of transcripts in each cluster is 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 delays between pre-mRNA and mature mRNA peak (arrows) by fitting and sampling (n = 100) from a Gaussian process with a squared exponential kernel to give estimates with uncertainty. (D) LOWESS fit of the delay between the peak of pre-mRNA and mature mRNA against inferred half-life. Pearson's correlation coefficient of 0.49 with p = 3.2e-10 for testing non-correlation. Points representing transcripts are coloured by 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 mRNA 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 (by combining reads for all isoforms of a gene) 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). The pre-mRNA and mature mRNA data were modelled using a Gaussian process which was then sampled with n = 100, so that the delay between the peaks could be estimated and the uncertainty in the estimate quantified (Figure 3C). The relationship between delay and half-life estimated from the model, for each transcript that has been modelled, is shown in Figure 3D. There is a moderately positive yet significant correlation between the two variables (Pearson's R = 0.49, p = 3.2e-10). 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 that delay is indicative of half-life; 54% of short delay genes have short half-lives and 67% of long delay genes have long half-lives. The picture is less clear in the medium delay category where the proportions of genes with short, medium and long half-lives are comparable, however the highest proportion (38%) of genes is in the medium half-life category. 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.

mRNA stability is related to conformation

The degradation of an mRNA in the cytoplasm can be closely linked to its translation (Mugridge et al., 2018). We therefore sought to investigate 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 mRNA 3'UTR (Mayr, 2017). 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 (Burow et al., 2015). 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 (Hanson and Coller, 2018), we investigated whether there was a relationship between half-life and both the translation efficiency and codon usage across the transcripts within our dataset. Using a previously published dataset obtained by ribosome footprint profiling of 2-3 hour embryos (Eichhorn et al., 2016) we plotted the translation efficiency and half-life for each of the transcripts within our dataset (Figure 4A). We observed a very slight but significant negative correlation between translation efficiency and half-life (Pearson's R = -0.141, p = 0.028).

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 (Burow et al., 2018) and their occurrence within the CSC plot. Although the five codons with the highest CSC values in our dataset are classified as optimal codons the proportion of optimal codons is not enriched within the positive and negative CSC groups (40% vs. 32%, $p = 0.5946$) (Figure 4B). There was 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 (Adivarahan et al., 2018; Khong and Parker, 2018; Vinter et al., 2021), raising the possibility that a particular conformation may also influence mRNA stability. To investigate this, we selected a set of 6 zygotic mRNAs, with different half-lives ranging from 6 to 95 minutes (Figure S8A). We used dual-colour smFISH probes to visualise their 5' and 3’ ends, and quantitate the distance between them, in fixed embryos (Figure 4Di). A representative smFISH image for one of the test mRNAs, *Deformed* (*Dfd*), is shown in Figure 4Dii, images for the other mRNAs tested are shown in Figure S8B.

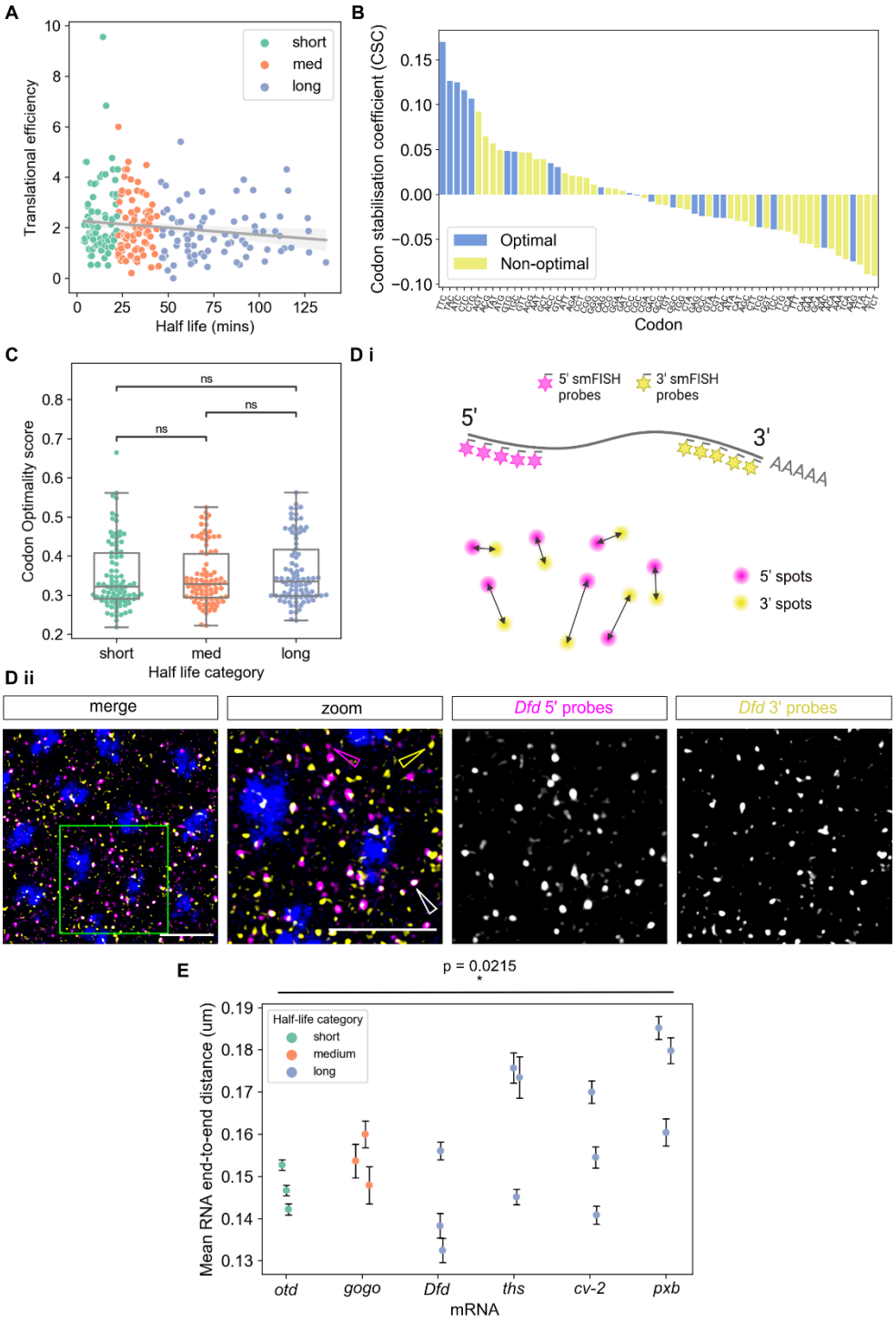

Figure 4: mRNA properties and stability. (A) Estimated half-lives of each transcript (x-axis) versus the translation efficiency from 2-3 hour embryos (Eichhorn et al., 2016). Points representing transcripts are coloured by half-life category (Pearson's R = -0.141, p = 0.028). (B) Codon stabilisation coefficient of the codons calculated from our estimated half-lives. Optimal codons are shown in blue and non-optimal codons are shown in yellow (chi-squared test p = 0.5946). (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 as tested by independent t-test (short vs med p = 1; short vs

long p = 1; med vs long p = 0.8). (Di) Schematic showing detection of the 5' (magenta) and 3' (yellow) ends of each mRNA with different smFISH probe sets. Spots belonging to the same mRNA are matched (see Methods). (Dii) Maximum projection of 6 slices from a confocal image showing smFISH detection of the 5' and 3' ends of *Dfd* mRNAs. Scale bars: 5 µm. A higher magnification view of the boxed region is shown with lone 5' ends, 3' ends and colocalised ends labelled by magenta, yellow and white arrowheads, respectively. (E) Graph shows the end-to-end distances of mRNAs with different stabilities, n = 3 embryos, data are mean +/- standard error. Genes are ordered by their model half-life along the x-axis and coloured by half-life category (p = 0.0215, tested by ordinal regression).

For each image, the number and position of all the 5' and 3' probe signals were collected and pairs were identified by treating the data as a paired assignment problem (Figure 4Di). For each pair, the distance between the 5' and 3' signals was then measured; only ends with a distance less than 300 nm were considered to belong to the same mRNA (Vinter et al., 2021). The distributions of end-to-end distances for each of the mRNAs tested reveal that there is a significant (p = 0.0215) trend whereby shorter half-life mRNAs are more compact, based on a shorter distance between their 5' and 3' ends (Figure 4E). We also identified unpaired mRNA ends (see later), which were further apart than the 300 nm distance threshold used. Similar data were obtained from quantitation of control smFISH experiments in which the fluorophore dyes on each set of probes were switched (Figure S8C-D). In addition, a precision control for the smFISH dual labelling is included for the *otd* mRNA, based on detection of two different probe sets, labelled with different fluorophores, that hybridise alternately to the *otd* mRNA (Figure S8E). This control confirms that the end-to-end distances we measure are significantly above the detection threshold (Figure S8E-F). In contrast, we find no correlation between mRNA compaction and mRNA length, as *Dfd* and *golden goal* (*gogo*) 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 linked to translation efficiency or optimal codon content, however unstable mRNAs tend to be more compact than long half-life transcripts.

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

As we did not detect any strong relationship between mRNA translation and half-life, we next tested whether mRNA localisation influences stability. To this end, we focused on cytoplasmic P-bodies, which have been implicated in mRNA degradation and storage in *Drosophila* (Wang et al., 2017). Therefore, we investigated whether mRNAs with distinct stabilities are differentially localised to P-bodies in the early *Drosophila* embryo. We visualised P-bodies using Me31B, which is a general marker of P-bodies, including in *Drosophila* (Patel et al., 2016). To detect Me31B we made use of a fly stock carrying a GFP-Me31B exon trap in which GFP is inserted into the endogenous Me31B locus (Buszczak et al., 2007).

Using smFISH we quantified single mRNAs and P-bodies labelled by GFP-Me31B in fixed embryos. The same set of six 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 S9A). These foci have a mean radius of 200 nm (Figure S9B), consistent with a previous observation that P-bodies in the embryo are smaller than those in the oocyte (Sankaranarayanan et al., 2021).

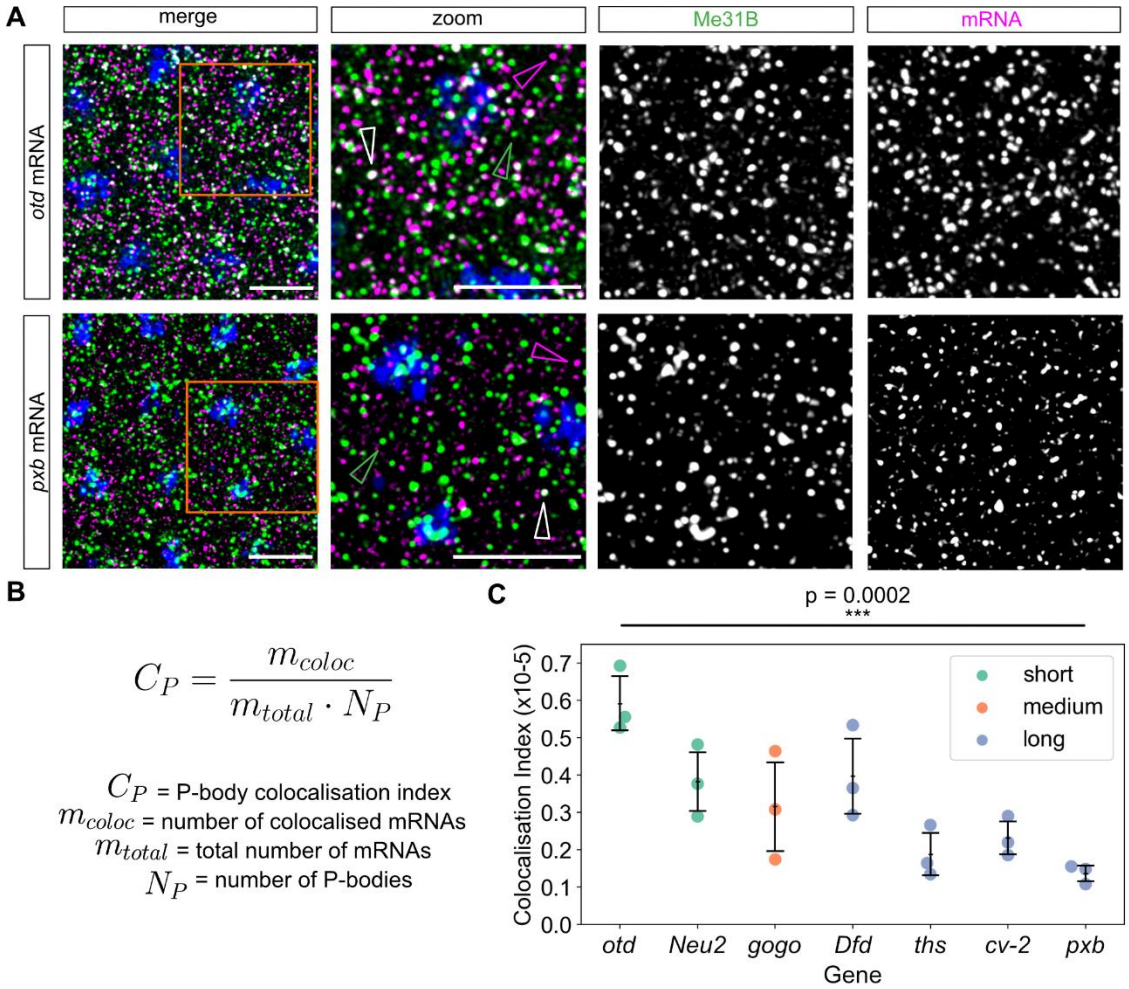

Figure 5: Short half-life mRNAs are more colocalised with P-bodies in the early embryo.
(A) Confocal images of fixed, early nc14 embryos stained with smFISH probes for the indicated mRNAs (magenta) and showing labelled GFP-Me31B P-bodies (green). Scale bars: 5 µm. Image is a maximum projection of 7 slices, along with a higher magnification image shown as both a merged image and single channels. 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 shows the P-body colocalisation index for the indicated mRNAs in early nc14. mRNAs are ordered by their model half-life and coloured by half-life category. Points represent individual embryos and overlays display mean and standard deviation across replicates. Unstable mRNAs are significantly more colocalised with P-bodies, tested by ordinal regression (p = 0.0002).

For each mRNA tested, a proportion of the individual mRNA signals colocalise with P-bodies (Figure 5A, Figure S9A). As seen in Figure 5A, orthodenticle (*otd*) (also called *ocelliless*) mRNAs appear to be more highly colocalised with P-bodies than *pxb* mRNAs. As *otd* has a much shorter half-life than *pxb* (6 mins and 95 mins respectively) we examined whether this was a general trend across the set of test mRNAs. To quantitate colocalisation, we used a colocalisation index which controls for variation in both mRNA and P-body numbers between embryos (Figure 5B). This analysis reveals that short half-life mRNAs are significantly more colocalised with P-bodies than mRNAs with long half-lives (Figure 5C, $p = 0.0002$), suggesting that P-bodies may be sites of mRNA degradation.

We noticed in our dual-colour smFISH images that there is a proportion of unpaired 5' and 3' mRNA ends suggestive of degradation intermediates (Figure 4Dii). Therefore, to further investigate whether P-bodies may be sites of mRNA degradation, we performed dual colour smFISH with 5' and 3' probe sets, in addition to GFP-Me31B as a P-body marker. These data allowed us to identify unpaired 5' and 3' ends and assess whether there is an enrichment of either end in P-bodies. A representative image of an early nc14 embryo is shown for the *Dfd* mRNA in Figure 6A, which reveals that complete mRNAs (orange arrowhead) and lone 3' ends (yellow arrowhead) can be seen to be 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, the data is shown for the other test mRNAs in S10A). Quantitation of the proportion of single 5' and 3' signals that localise to P-bodies reveals a general trend that there are more unpaired 3' ends in P-bodies, which is significant for half of the mRNAs investigated (Figure 6C). Similar results are obtained when the fluorophores on the *Dfd* 5' and 3' probes are reversed (Figure S10B). The detection of these lone 3' signals is consistent with them being 5' to 3' mRNA decay intermediates. Together, these data suggest that mRNA degradation occurs within P-bodies in the early *Drosophila* embryo.

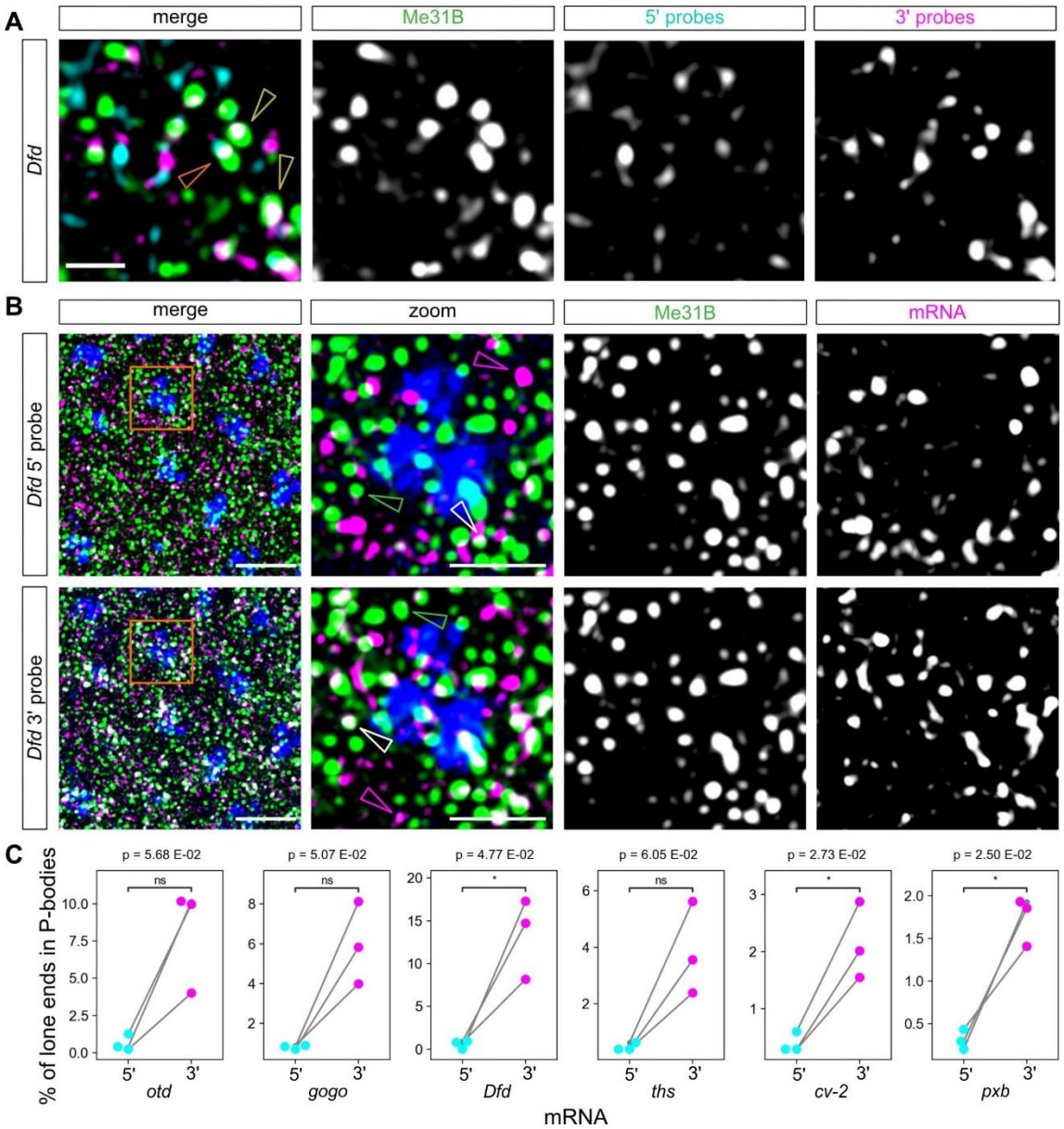

Figure 6: mRNA 3' end fragments are more associated with P-bodies than 5' ends.
(A) Maximum projection (2 slices) of a confocal image showing smFISH staining of an early nc14 embryo with probes for *Dfd* 5' (cyan) and 3' (magenta) ends, GFP-Me31B (green) and DAPI staining (blue). Examples where both 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) 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. Overlapping data points have been separated for clarity. n=3 embryos, paired t-test used to determine significance.

Discussion

Here, using total RNA-seq time series data and Gaussian process regression, half-lives of ~300 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 (Graveley et al., 2011; Lott et al., 2011). 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, where the estimates can be affected by the labelling time chosen (Furlan et al., 2021; Tani and Akimitsu, 2012). 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 (Furlan et al., 2020). 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.

Due to strict filtering we have only derived a proportion of transcript half-lives from the dataset. These filters could be loosened to provide half-lives for more mRNAs, but potentially with lower confidence in the estimates. As we were unable to study some mRNAs because they lack introns, a Pol II ChIP-seq time series could be generated to use instead of intronic reads for the transcription profiles. In addition, we show how the stability of transcripts can be classified using the delay between the peak of the pre-mRNA and mature mRNA, which represents a simpler approach for estimating stabilities that does not require modelling.

The half-lives we estimated for ~300 zygotic transcripts in the early embryo have a median of 31 minutes. This is similar to the median half-lives of ~30-40 mins reported for maternal mRNAs degraded by only the maternal decay pathway or by both the maternal and zygotic pathways, based on microarray analysis of fertilised and unfertilised *Drosophila* embryos (Thomsen et al., 2010). In addition, the range of half-lives we observe is consistent with 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 (Boettiger and Levine, 2013; Edgar et al., 1986; Little et al., 2013). Moreover, the wide range of half-lives we observe in the embryo is consistent with mRNA stability being an important checkpoint of control in the regulation of gene expression.

The median half-life we estimate in the early embryo 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 (Burow et al., 2015). While it is possible that the pulse-labelling timing skews some of the half-life estimates (Uvarovskii et al., 2019), the shorter median half-life in the early embryo may reflect its rapid initial development. Early embryogenesis is characterised by short mitotic cycles (Ferree et al., 2016) and fast rates of transcription (Fukaya et al., 2017) and translation (Dufourt et al., 2021), with the resulting localised gene expression patterns specifying three tissues along the dorsal-ventral axis in a time period of only 90 mins (Levine and Davidson, 2005). 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, whereas stable mRNAs encode signalling receptors and transmembrane transporters. This is consistent with transient localised expression of key transcription factors in the early embryo and the mRNAs encoding transcription factors commonly being unstable (Burow et al., 2015; Edgar et al., 1989; Thomsen et al., 2010). In contrast, signalling receptors tend to be more ubiquitously expressed, with signalling activation often regulated by ligand tissue specific expression and/or availability (Umulis et al., 2009). 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 (Stathopoulos and Levine, 2005).

Previous studies have shown that mRNAs exist in a more open conformation during translation, while untranslated mRNAs are more compact (Adivarahan et al., 2018; Khong and Parker, 2018; Vinter et al., 2021) regardless of whether they are stress granule associated (Adivarahan et al., 2018; Khong and Parker, 2018). 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 (Ermolenko and Mathews, 2021). Alternatively, the shorter distance between 5’ and 3’ ends could reflect a transient interaction associated with degradation for all mRNAs, 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 (Hanson and Coller, 2018). 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 (Hanson and Coller, 2018). 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 (Bazzini et al., 2016). 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 (Burow et al., 2018). 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, a dependence on a particular distribution of non-optimal codons 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 (Gobet et al., 2020).

Our data show that short half-life mRNAs, which are enriched for those encoding transcription factors, are more likely to colocalise with the P-body marker Me31B than more stable mRNAs in the early embryo. This is consistent with enrichment of mRNAs encoding RNA polymerase II regulators in P-bodies purified from mammalian cells, although in these cells P-bodies are associated with mRNA storage (Hubstenberger et al., 2017). 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 they may be sites of 5' to 3' mRNA decay in the early *Drosophila* embryo. In contrast, we detect weaker colocalization of 5' end fragments with P-bodies, suggesting that 3' to 5' mRNA degradation by the exosome does not occur in P-bodies. In support of this, components of the exosome are largely absent from P-bodies (Standart and Weil, 2018).

A role for P-bodies in 5' to 3' decay is consistent both with early studies in yeast following the discovery of P-bodies (Sheth and Parker, 2003) and with later work in *Drosophila* suggesting that P-bodies are involved in mRNA degradation in the embryo following zygotic genome activation (Wang et al., 2017) and in intestinal stem cells (Buddika et al., 2021). In addition, the Xrn1 exonuclease localises with P-bodies in yeast, *Drosophila* and mammalian cells (Jones et al., 2012). However, P-bodies have been implicated in mRNA storage and translational repression in mature *Drosophila* oocytes (Sankaranarayanan et al., 2021) and embryos prior to zygotic genome activation (Wang et al., 2017). 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 (Hubstenberger et al., 2017) or based on a live imaging approach (Horvathova et al., 2017), mRNA degradation when P-body formation is disrupted (Hubstenberger et al., 2017) and the ability of P-body mRNAs to re-enter translation (Bhattacharyya et al., 2006; Brengues et al., 2005).

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 mature oocyte to smaller, more dynamic structures in the early embryo (Sankaranarayanan et al., 2021). Moreover, at the maternal-to-zygotic transition some P-body proteins are degraded, including the Cup

translational repressor protein, which may facilitate a switch from a translational repression/storage role to an mRNA degradation function prevailing in P-bodies (Wang et al., 2017). Our data suggest that zygotic mRNAs exploit the degradation function of P-bodies in their post-transcriptional regulation. Future studies exploiting the method developed for determining the protein and RNA contents of purified P-bodies (Hubstenberger et al., 2017), 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

Fly stocks

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

Staging and collection of embryos for RNA-seq

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

Ten timepoints were collected spanning early *Drosophila* embryonic development from nc11 through to cephalic furrow formation (Table 1). Embryos were collected 5 minutes after nuclear division for nc11 and nc12, 5 and 15 minutes following the nc13 nuclear division and every 15 minutes following

the nc14 nuclear division as well as embryos that showed clear cephalic furrow formation. This yielded samples covering every 10-15 minutes through development from nc11 to cephalic furrow formation. The internuclear distance of 15-20 nuclei pairs per embryo was measured in Fiji and normalised to the whole embryo length to obtain an average internuclear distance per embryo (Figure S1A and B). This was compared to the internuclear distance of embryos of known stages to accurately confirm the nuclear cleavage stage and age of embryos. All embryos were collected at 20°C with approximate time after egg lay in minutes shown in Table 1.

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' (Lott et al., 2011) 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 stranded library prep kit (Illumina) according to the manufacturer's protocol. Unique dual index adaptors were used for each library and they were pooled in equimolar concentration and run across 8 lanes on the flow cell of the HiSeq 4000 to obtain paired end sequence reads. The average number of reads obtained per library was 105 million reads.

Embryo fixation and smFISH

Flies were allowed to lay on apple juice agar plates in small cages for 2 hours at 25°C. After ageing for another 2 hours, 2-4 hour old embryos were dechorinated in 50% bleach for 3 minutes and washed thoroughly in distilled water. Embryos were fixed as previously described (Kosman et al., 2004) and stored in methanol at -20°C until required. Fixed embryos were placed in Wheaton vials (Sigma, Z188700-1PAK) for the smFISH reaction as described previously (Hoppe et al., 2020). mRNA targets were detected in embryos using smiFISH probes designed to exonic sequence with 5' end X flap sequences (Tsanov et al., 2016) and using secondary detection probes labelled with Quasar 570 or Quasar 670 fluorophore dyes (LGC Biosearch Technologies). Probe sequences are listed in Supplementary Table 2. DAPI (500µg/ml) was added to the third of the final four washes of the protocol at a concentration of 1:1000 and embryos were mounted onto slides in Prolong Diamond to set overnight before imaging. To visualise the membrane to age the embryos a mouse α-Spectrin antibody (DSHB, 3A9 (323 or M10-2)) with an Alexa Fluor 647 Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody (Thermo Fisher Scientific, A-31571) was used or a brightfield image was taken.

For compaction experiments at least 24 probes were designed to each end of the mRNA (5' and 3') separated by at least 1.3kb. As a control, fluorophore dyes were switched and the images from stained embryos analysed and quantified. Additional controls for *otd* used adjacent probes with alternating Quasar dyes to determine the precision of detection of single mRNAs.

Confocal microscopy of fixed embryos

A Leica TCS SP8 gSTED confocal was used to acquire images of the transcription sites (TSs), single mRNAs and P-bodies within cells of fixed embryos using a 100x/ 1.3 HC PI Apo Cs2 objective with 3X line accumulation and 3X zoom for compaction and P-body colocalisation experiments, and 2X zoom for quantifying mRNAs for the half-life validation. Confocal settings were ~0.6 airy unit pinhole, 400 Hz scan speed with bidirectional line scanning and a format of 2048 x 2048 or 4096 x 4096 pixels. Laser detection settings were collected as follows: PMT DAPI excitation at 405nm (collection: 417-474nm); Hybrid Detectors: AlexaFluor 488 excitation at 490nm (collection: 498-548nm), Quasar 570 excitation at 548nm (collection: 558-640nm) and Quasar 670 excitation at 647nm (657-779nm) with 1-6ns gating.

All images were collected sequentially and optical stacks were acquired at system optimised 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).

Image analysis

The spot detection algorithm Airlocalize (Trcek et al., 2017) 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 (Boettiger and Levine, 2013), embryos were imaged at various time points during the 13th nuclear division (Figure S4) using the DAPI channel and reference movies of His-RFP (Hoppe et al., 2020) 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 (Bray et al., 2016) 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 (BinTayyash et al., 2021). 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. 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.

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 (Honkela et al., 2015; Lawrence et al., 2006). 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 – 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 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. 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 from the 33% and 66% quantiles of the data, in this instance 22 and 45 minutes.

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

Clustering

The intronic data (n = 5035) was z-transformed and clustered using GPclust (Hensman et al., 2015). Intronic clusters of interest exhibiting a range of expression profiles were selected (clusters 2 and 5) (Figure S5). The zygotic transcripts (n = 593) corresponding to the genes in each selected intronic cluster were then normalised and clustered (Figure S6). Summary statistics for the half-lives of the genes in each zygotic cluster were then computed for clusters with >2 transcripts with estimated half-lives. A list of the transcripts for intronic clusters 2 and 5 and their corresponding zygotic clusters can be found in Supplementary Table 3.

Analysis of time delays

Delays between peak of pre-mRNA and mature mRNA time series from the zygotic set (n = 593) were estimated by fitting a Gaussian process with RBF kernel to each time series. 100 samples from each GP were taken and the delay between the peak of each sampled function of premature and mature mRNA were computed to provide estimates of the delay with uncertainty. Any transcripts with delays ≤ 0, or with mature mRNA profiles peaking at the final timepoint (t = 220), were removed. Transcripts were then grouped into short, medium and long delay groups, setting the boundaries from the 33% and 66% quantiles of the data (17.55 and 36.16 mins respectively). A gene was classified as short if there was 90% probability that the delay for that gene was in the short delay interval. All statistical analysis was carried out in Python using the scipy, sklearn and statannot libraries.

Gene ontology analysis

Gene ontology analysis was conducted using Gorilla (Eden et al., 2009). Enrichment of short and long half-life genes was performed using the half-life set as the target set and the entire group of dynamic genes from the RNA-seq dataset (n = 8791) as the reference set with default parameters.

Codon usage and translation efficiency analysis

For further analysis we removed transcripts with half-lives of less than a minute and greater than 150 minutes. The codon stabilisation coefficient (CSC) value was calculated for each codon as previously described (Burow et al., 2018; Presnyak et al., 2015). The CSC is equivalent to the Pearson correlation coefficient, calculated by plotting the frequency of each codon per transcript within our dataset against its half-life. Classification of optimal *Drosophila* codons used are as in (Burow et al., 2018). A chi-square test of association between optimal and non-optimal codons in positive and negative CSC groups was determined. The codon optimality score was determined by adding the proportion of optimal codons within each transcript. Transcripts were grouped 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 (Eichhorn et al., 2016). 3'UTR and transcript lengths were obtained from Flybase (Larkin et al., 2021).

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 (Vinter et al., 2021). 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, 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. 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).

Author Contributions

Conceptualization, 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.

References

Adivarahan S, Livingston N, Nicholson B, Rahman S, Wu B, Rissland OS, Zenklusen D. 2018. Spatial Organization of Single mRNPs at Different Stages of the Gene Expression Pathway. *Molecular Cell* **72**:727-738.e5. doi:10.1016/j.molcel.2018.10.010

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

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

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

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

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

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

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

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

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

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

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

Buddika K, Huang Y-T, Ariyapala IS, Butrum-Griffith A, Norrell SA, O'Connor AM, Patel VK, Rector SA, Slovan M, Sokolowski M, Kato Y, Nakamura A, Sokol NS. 2021. Coordinated repression of pro-differentiation genes via P-bodies and transcription maintains Drosophila intestinal stem cell identity. *Current Biology* S0960982221015918. doi:10.1016/j.cub.2021.11.032

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

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

Buszczak M, Paterno S, Lighthouse D, Bachman J, Planck J, Owen S, Skora AD, Nystul TG, Ohlstein B, Allen A, Wilhelm JE, Murphy TD, Levis RW, Matunis E, Srivali N, Hoskins RA, Spradling AC. 2007. The Carnegie Protein Trap Library: A Versatile Tool for Drosophila Developmental Studies. *Genetics* **175**:1505–1531. doi:10.1534/genetics.106.065961

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, Brown JB, Cherbas L, Davis CA, Dobin A, Li R, Lin W, Malone JH, Mattiuzzo NR, Miller D, Sturgill D, Tuch BB, Zaleski C, Zhang D, Blanchette M, Dudoit S, Eads B, Green RE, Hammonds A, Jiang L, Kapranov P, Langton L, Perrimon N, Sandler JE, Wan KH, Willingham A, Zhang Y, Zou Y, Andrews J, Bickel PJ, Brenner SE, Brent MR, Cherbas P, Gingeras TR, Hoskins RA, Kaufman TC, Oliver B, Celniker SE. 2011. The developmental transcriptome of Drosophila melanogaster. *Nature* **471**:473–479. doi:10.1038/nature09715

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

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

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

Honkela A, Peltonen J, Topa H, Charapitsa I, Matarese F, Grote K, Stunnenberg HG, Reid G, Lawrence ND, Rattray M. 2015. Genome-wide modeling of transcription kinetics reveals patterns of RNA production delays. *Proc Natl Acad Sci U S A* **112**:13115–13120. doi:10.1073/pnas.1420404112

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

Horvathova I, Voigt F, Kotrys AV, Zhan Y, Artus-Revel CG, Eglinger J, Stadler MB, Giorgetti L, Chao JA. 2017. The Dynamics of mRNA Turnover Revealed by Single-Molecule Imaging in Single Cells. *Molecular Cell* **68**:615-625.e9. doi:10.1016/j.molcel.2017.09.030

Hubstenberger A, Courel M, Bénard M, Souquere S, Ernoult-Lange M, Chouaib R, Yi Z, Morlot J-B, Munier A, Fradet M, Daunesse M, Bertrand E, Pierron G, Mozziconacci J, Kress M, Weil D. 2017. P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. *Molecular Cell* **68**:144-157.e5. doi:10.1016/j.molcel.2017.09.003

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

Jones CI, Zabolotskaya MV, Newbury SF. 2012. 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* **3**:455–468. doi:10.1002/wrna.1109

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

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

Larkin A, Marygold SJ, Antonazzo G, Attrill H, dos Santos G, Garapati PV, Goodman JL, Gramates LS, Millburn G, Strelets VB, Tabone CJ, Thurmond J, FlyBase Consortium, Perrimon N, Gelbart SR, Agapite J, Broll K, Crosby M, dos Santos G, Falls K, Gramates LS, Jenkins V, Longden I, Matthews B, Sutherland C, Tabone CJ, Zhou P, Zytkovicz M, Brown N, Antonazzo G, Attrill H, Garapati P, Larkin A, Marygold S, McLachlan A, Millburn G, Pilgrim C, Ozturk-Colak A, Trovisco V, Kaufman T, Calvi B, Goodman J, Strelets V, Thurmond J, Cripps R, Lovato T. 2021. FlyBase: updates to the *Drosophila melanogaster* knowledge base. *Nucleic Acids Research* **49**:D899–D907. doi:10.1093/nar/gkaa1026

Lawrence ND, Sanguinetti G, Rattray M. 2006. Modelling Transcriptional Regulation Using Gaussian ProcessesProceedings of the 19th International Conference on Neural Information Processing Systems, NIPS'06. Cambridge, MA, USA: MIT Press. pp. 785–792.

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

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

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

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

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

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

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

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

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

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

Patel PH, Barbee SA, Blankenship JT. 2016. GW-Bodies and P-Bodies Constitute Two Separate Pools of Sequestered Non-Translating RNAs. *PLoS ONE* **11**:e0150291. doi:10.1371/journal.pone.0150291

Presnyak V, Alhusaini N, Chen Y-H, Martin S, Morris N, Kline N, Olson S, Weinberg D, Baker KE, Graveley BR, Coller J. 2015. Codon Optimality Is a Major Determinant of mRNA Stability. *Cell* **160**:1111–1124. doi:10.1016/j.cell.2015.02.029

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

Sankaranarayanan M, Emenecker RJ, Wilby EL, Jahnel M, Trussina IREA, Wayland M, Alberti S, Holehouse AS, Weil TT. 2021. Adaptable P body physical states differentially regulate bicoid mRNA storage during early Drosophila development. *Developmental Cell* **56**:2886-2901.e6. doi:10.1016/j.devcel.2021.09.021

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

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

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

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

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

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

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

Tsanov N, Samacoits A, Chouaib R, Traboulsi A-M, Gostan T, Weber C, Zimmer C, Zibara K, Walter T, Peter M, Bertrand E, Mueller F. 2016. smiFISH and FISH-quant – a flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Res* **44**:e165–e165. doi:10.1093/nar/gkw784

Umulis D, O'Connor MB, Blair SS. 2009. The extracellular regulation of bone morphogenetic protein signaling. *Development* **136**:3715–3728. doi:10.1242/dev.031534

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

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

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

Wang M, Ly M, Lugowski A, Laver JD, Lipshitz HD, Smibert CA, Rissland OS. 2017. ME31B globally represses maternal mRNAs by two distinct mechanisms during the Drosophila maternal-to-zygotic transition. *eLife* **6**:e27891. doi:10.7554/eLife.27891

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

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

Supplementary Data

Supplementary Figures S1-S10

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

Supplementary Figures

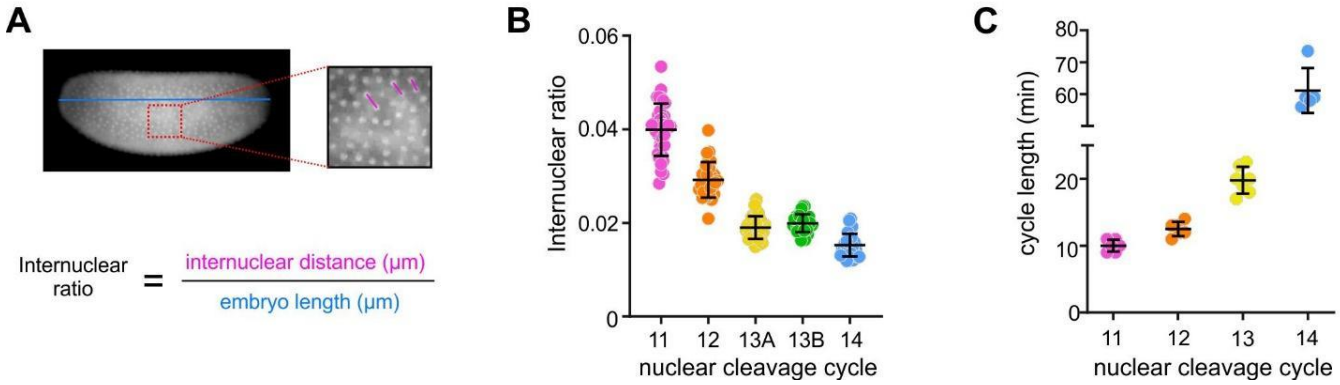

Figure S1. Staging of early embryos for RNA-seq. (A) Images of embryos were captured immediately prior to collection and the internuclear distance and embryo length were measured for each to give an internuclear ratio. (B) The internuclear ratio at nuclear cleavage cycles (nc) can be used to accurately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

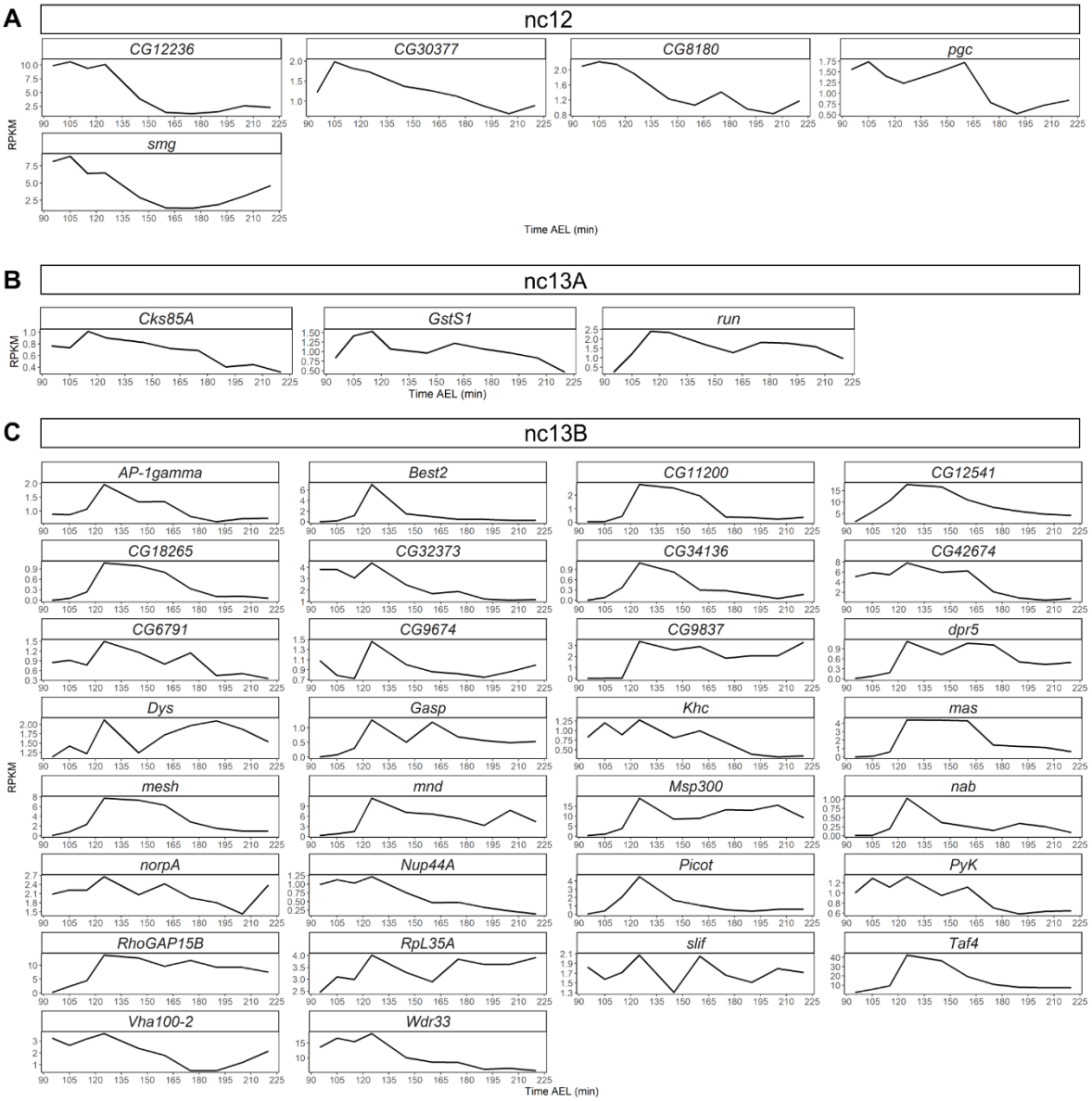

Figure S2. Intronic read counts of early zygotic genes. Intronic read counts (RPKM) of genes that show early transcription in our dataset, grouped depending on peak expression at (A) nc12 (105 min), (B) nc13A (115 min) or (C) nc13B (125 min), based on timings at 20°C.

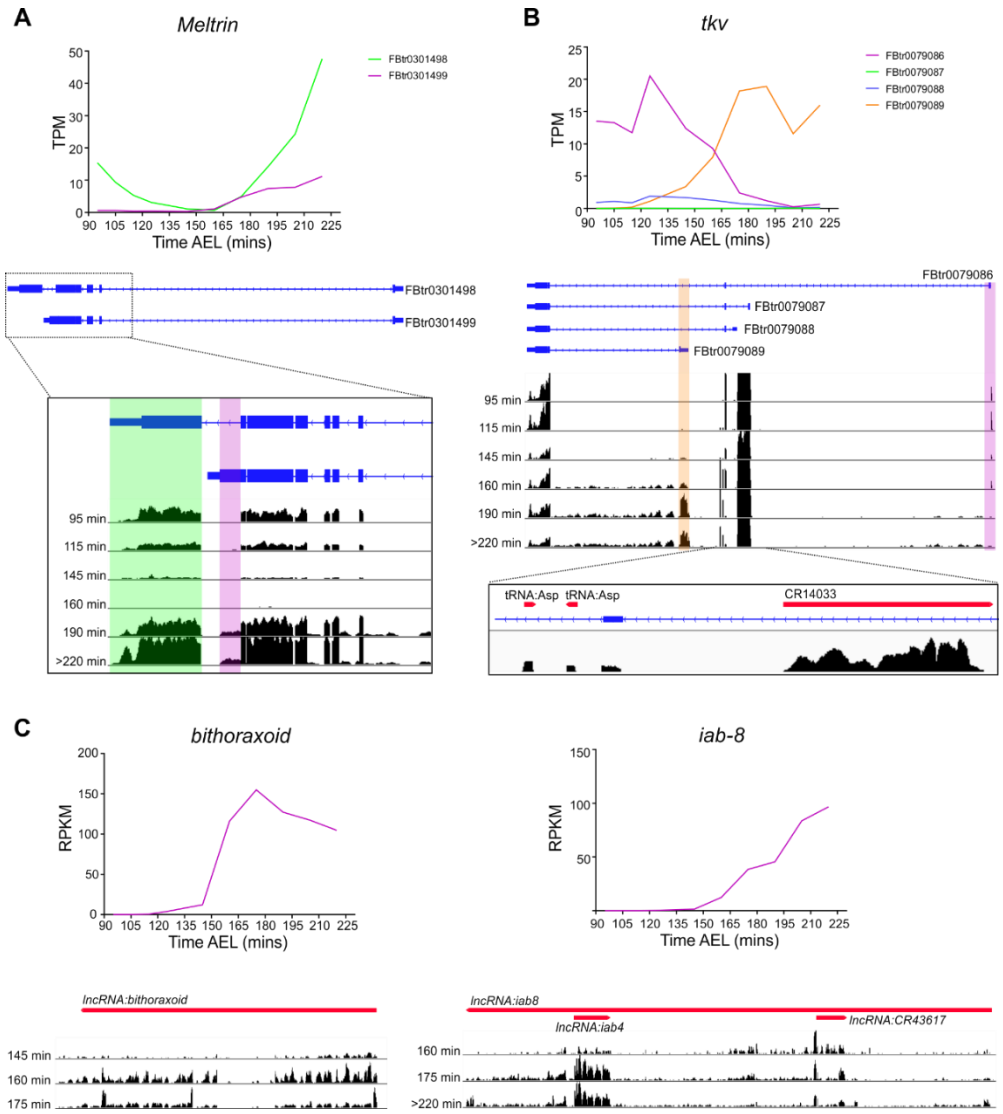

Figure S3. Examples of alternative isoform and non coding RNA expression during development in *Drosophila* embryos. (A) Quantification of the two different transcripts (TPM) of the *Meltrin* gene. The FBtr0301499 isoform (purple) is not detected in embryos <160 min AEL but is detected at later time 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	Half-life (mins)	Lower credible interval (mins)	Upper credible interval (mins)
RhoU	FBtr0110782	0.098	7	4	20
cv-2	FBtr0071610	0.0082	85	29	inf

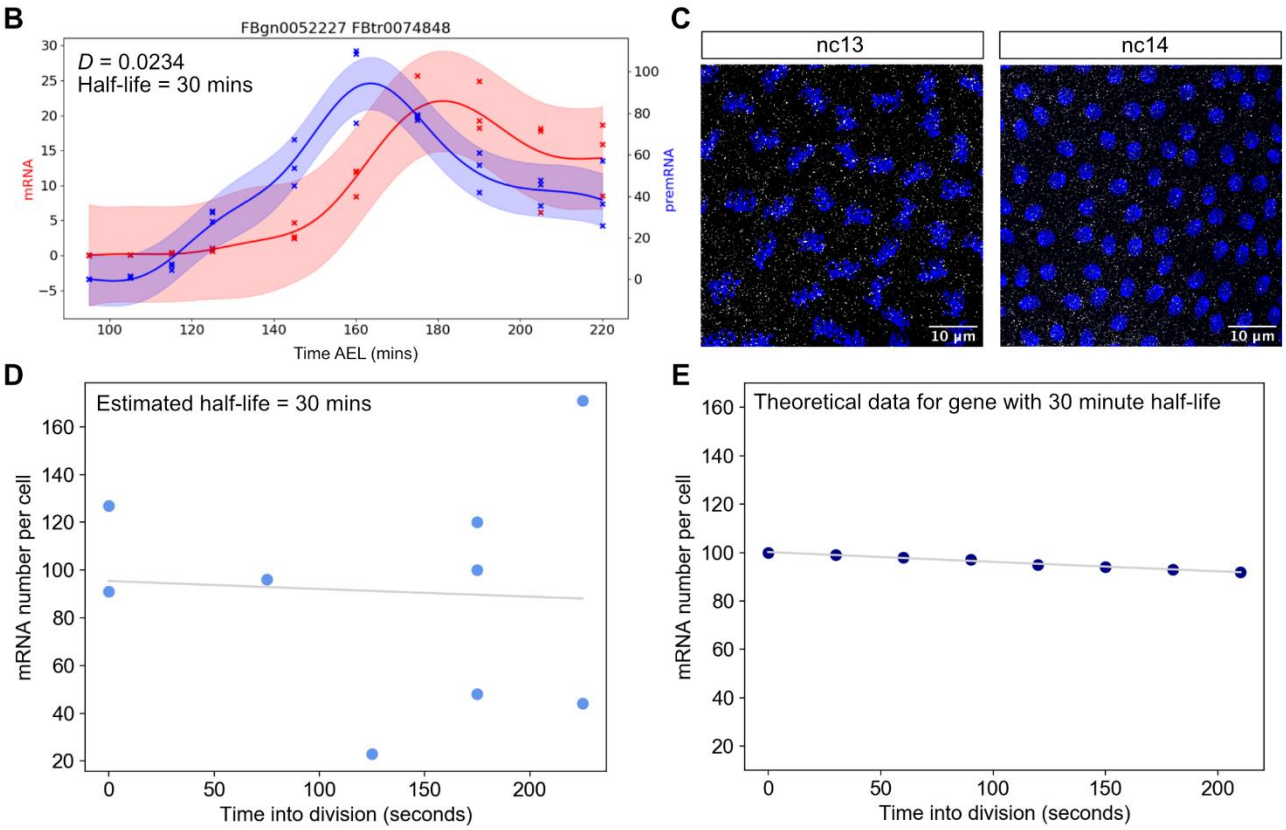

Figure S4: High embryo to embryo variation in transcript numbers masks variation due to degradation over cell division. (A) Parameter estimates for degradation rates and credible regions of short and long half-life genes shown in Figure 2D. (B) RNA-seq data for *gogo* fitted with the Gaussian process model. A half-life of 30 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 gene with a 30 min half-life over a 220 second time frame, which is a reduction of less than 10%.

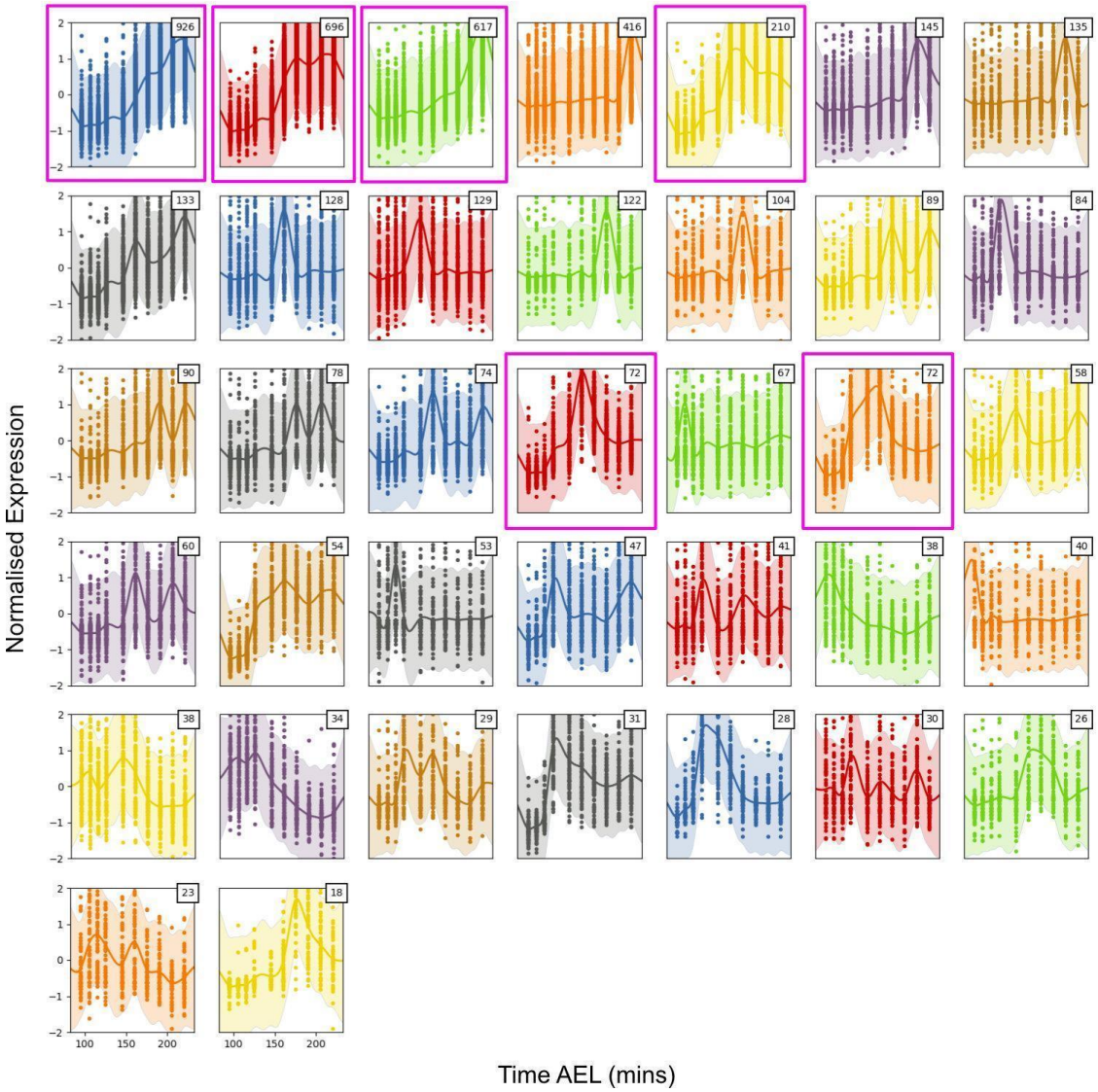

Figure S5: Output from clustering of 5035 intronic transcripts using GPclust. Data for all pre-mRNAs in the cluster are shown with shaded credible regions and inferred function as a solid line. The number of pre-mRNAs in each cluster is shown in the top right corner of each plot and graphs are arbitrarily 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.

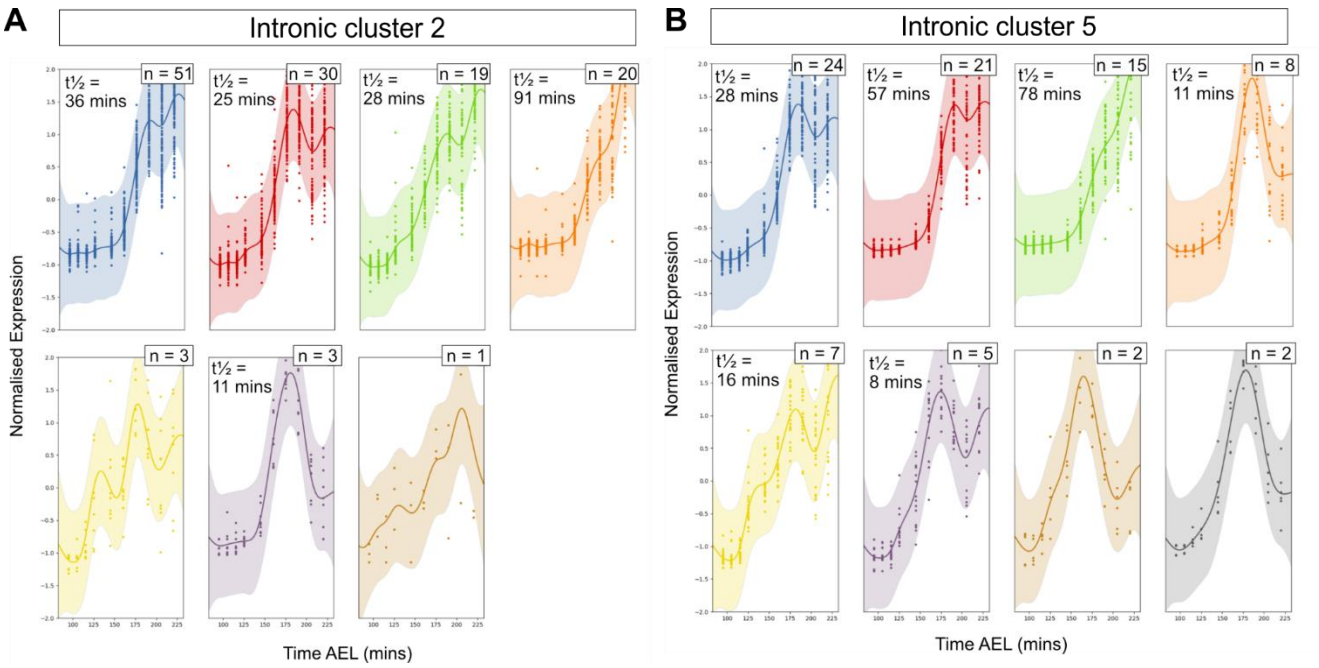

Figure S6: Sub-clustering of zygotic transcripts with intronic data in intronic cluster 2 and 5 using GPclust. Data for all mRNAs in the cluster are shown with shaded credible regions and inferred function as a solid line for (A) intronic cluster 2 and (B) intronic cluster 5. The number of mRNAs in each 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.

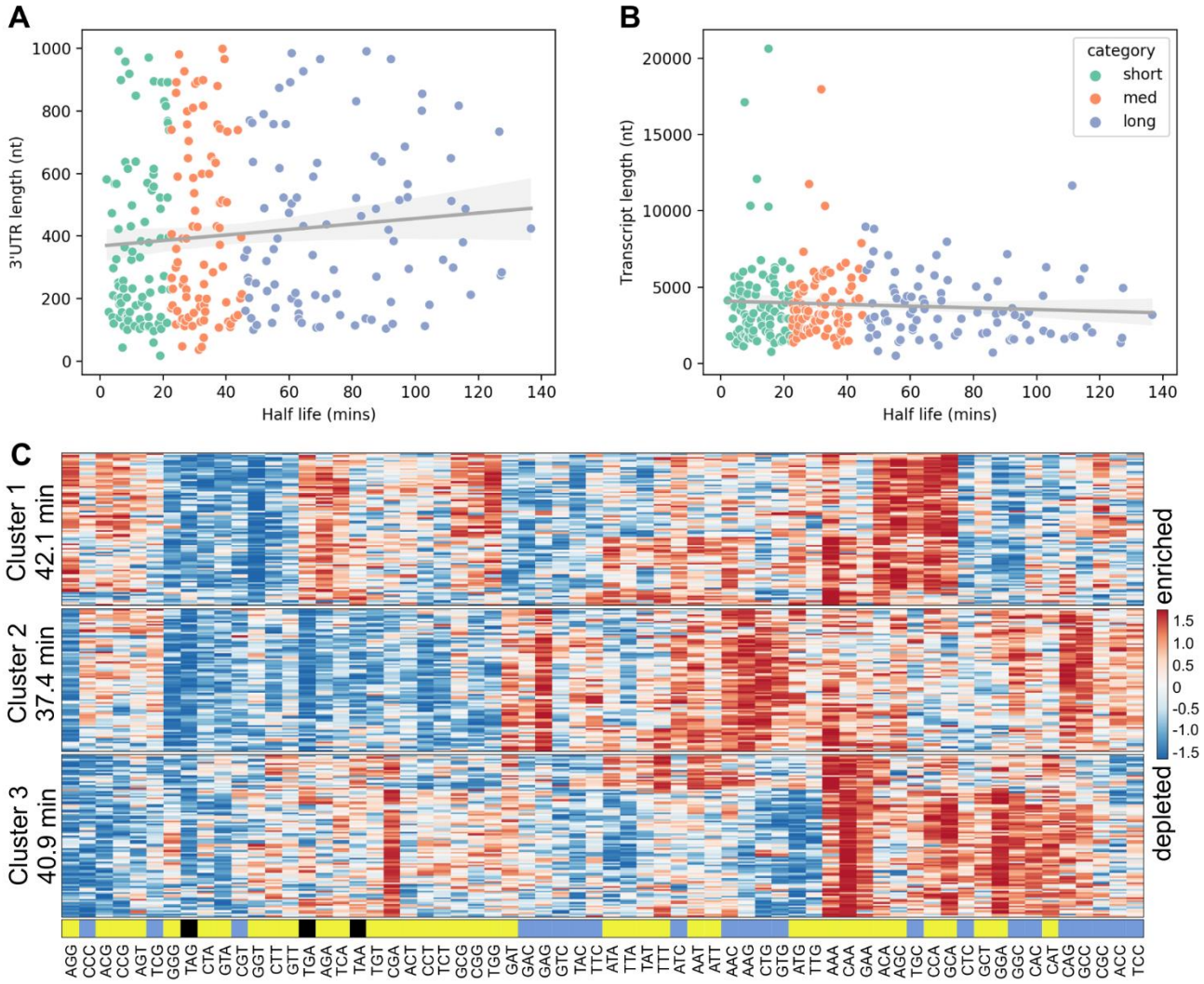

Figure S7: mRNA properties and stability. (A) Graph showing the correlation between mRNA half-life and 3' UTR length. Data fit with linear regression model, Pearson's R = 0.1, p = 0.1 (B) Correlation between mRNA total length and half-life. Transcripts are coloured by half-life category short (green), medium (orange) or long (blue). Data fit with linear regression model, Pearson's R = -0.07, p = 0.3 (C) Transcripts were clustered for transcripts with 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. 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	6	short
Neu2	FBtr0332718	7	short
gogo	FBtr0074848	30	medium
Dfd	FBtr0081621	52	long
ths	FBtr0088030	76	long
cv-2	FBtr0071610	85	long
pxb	FBtr0083162	95	long

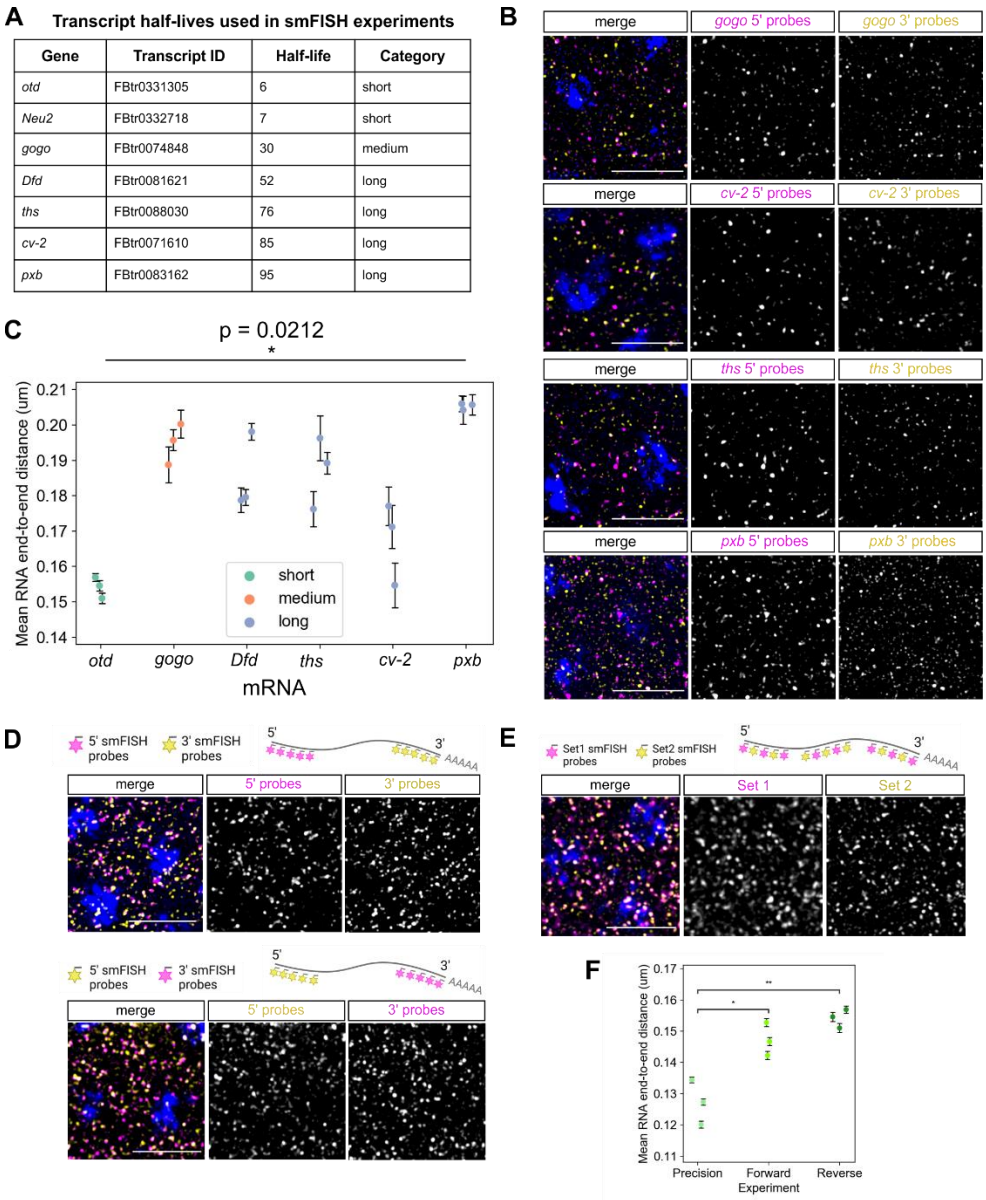

Figure S8: Compaction of mRNAs in early embryos. (A) A table of the transcripts used in smFISH experiments. (B) Confocal images of fixed embryos showing smFISH detection of the 5' (magenta) and 3' (yellow) ends of the *gogo*, *cv-2*, *ths* and *pxb* mRNAs used in compaction experiments. (C) Graph shows the end-to-end distances of mRNAs with different stabilities in the reverse fluorophore experiment to Figure 4E, n = 3 embryos, data are mean ± standard error. (D-E) Confocal images of fixed embryos stained with smFISH probes for *otd* mRNAs. Full length mRNAs were detected using probes for 5' and 3' ends with switched fluorophores (D), or with alternating labels used as a precision control (E). The schematics above each set of confocal images show the experimental probes and relevant colours in the images. (F) Quantification of the *otd* mRNA compaction data in D-E, data are mean ± standard error, independent unpaired t-test: * $p < 0.05$, ** $p < 0.01$. All images are maximum projections of 7 Z slices. Scale bars: 5 µm.

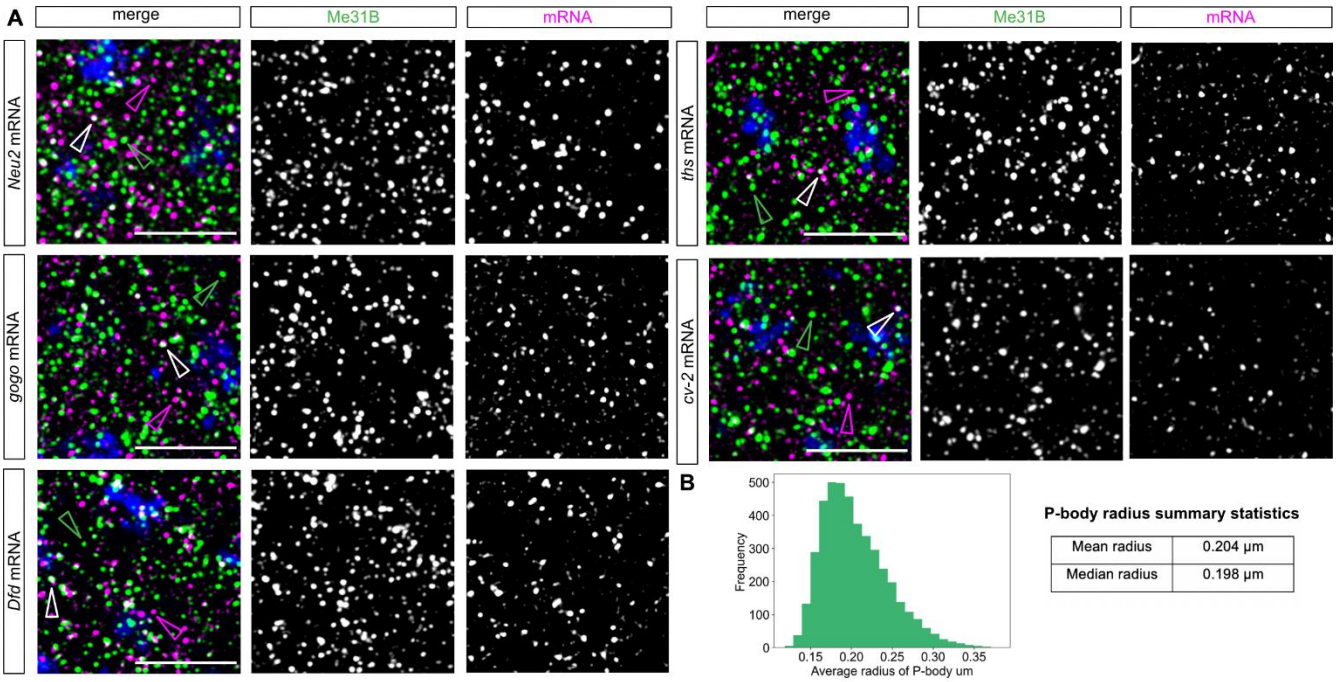

Figure S9: Confocal images of mRNAs with various stabilities and their colocalisation with the P-body marker Me31B. (A) Confocal images of early nc14 Me31B-GFP embryos showing smFISH staining of the indicated test mRNAs. DAPI marks the nuclei (blue), mRNAs are shown in magenta (magenta arrowhead), GFP-Me31B 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: 2µm (B) Analysis of P-body sizes reveals an average radius of 0.2 µm.

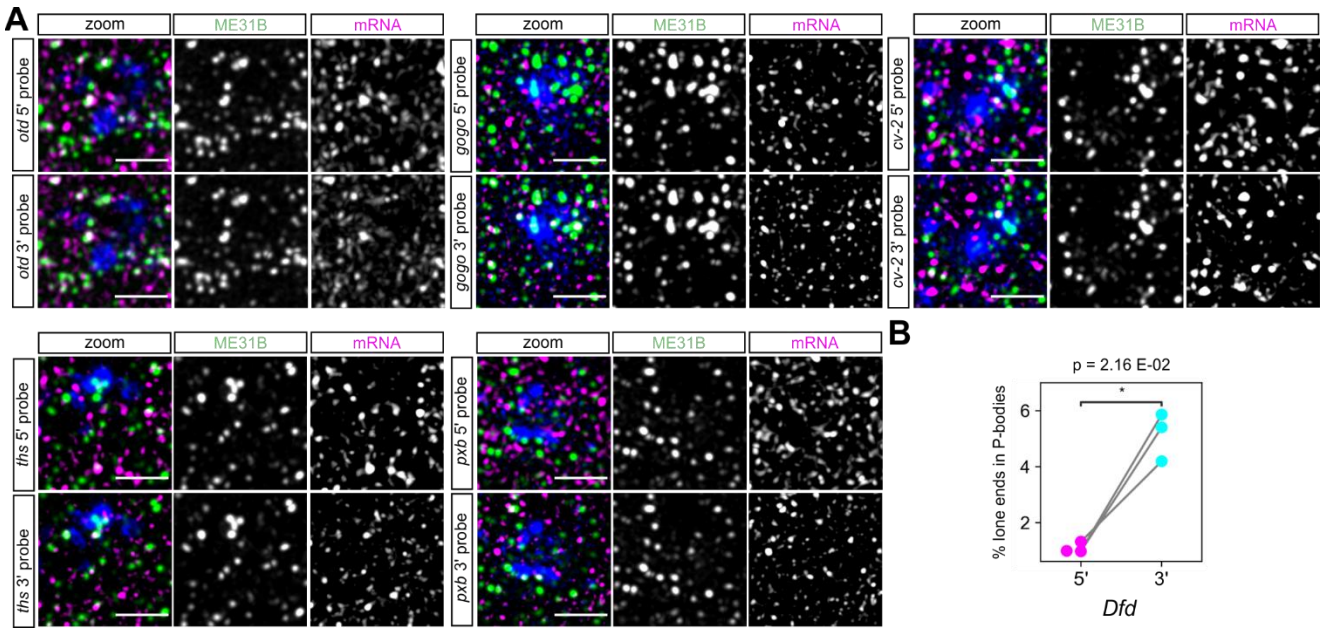

Figure S10: Confocal images of mRNAs and their 5' and 3' end colocalisation with the P-body marker Me31B. (A) Confocal images of early nc14 Me31B-GFP embryos showing smFISH staining of the indicated test mRNAs with 5' and 3' probe sets. For each test mRNA, the same region of the embryo is shown with 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. (B) 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 in the switched probe fluorophore experiment for *Dfd* (see also Figure 6C). Paired t-test used to determine significance.