5926–5938 *Nucleic Acids Research, 2020, Vol. 48, No. 11* *Published online 18 May 2020*
doi: 10.1093/nar/gkaa359

Gene expression dynamics are a proxy for selective pressures on alternatively polyadenylated isoforms

Michal Levin[1,†], Harel Zalts[2,†], Natalia Mostov[2,†], Tamar Hashimshony[2] and Itai Yanai[3,*]

[1]Quantitative Proteomics, Institute of Molecular Biology, Mainz 55128, Germany, [2]Faculty of Biology, Technion – Israel Institute of Technology, Haifa 3200003, Israel and [3]Institute for Computational Medicine, NYU Grossman School of Medicine, New York 10016, USA

Received January 09, 2020; Revised April 11, 2020; Editorial Decision April 26, 2020; Accepted April 27, 2020

ABSTRACT

Alternative polyadenylation (APA) produces isoforms with distinct 3′-ends, yet their functional differences remain largely unknown. Here, we introduce the APA-seq method to detect the expression levels of APA isoforms from 3′-end RNA-Seq data by exploiting both paired-end reads for gene isoform identification and quantification. We detected the expression levels of APA isoforms in individual *Caenorhabditis elegans* embryos at different stages throughout embryogenesis. Examining the correlation between the temporal profiles of isoforms led us to distinguish two classes of genes: those with highly correlated isoforms (HCI) and those with lowly correlated isoforms (LCI) across time. We hypothesized that variants with similar expression profiles may be the product of biological noise, while the LCI variants may be under tighter selection and consequently their distinct 3′ UTR isoforms are more likely to have functional consequences. Supporting this notion, we found that LCI genes have significantly more miRNA binding sites, more correlated expression profiles with those of their targeting miRNAs and a relative lack of correspondence between their transcription and protein abundances. Collectively, our results suggest that a lack of coherence among the regulation of 3′ UTR isoforms is a proxy for selective pressures acting upon APA usage and consequently for their functional relevance.

INTRODUCTION

Alternative polyadenylation (APA) is a crucial regulatory mechanism—widespread and conserved across all eukaryotes—that diversifies post-transcriptional regulation by selective mRNA–miRNA and mRNA–protein interactions (1–6). APA plays an important role in a vast variety of biological processes, such as maternal to zygotic transition, cell differentiation, and tissue specification (7–9) and exerts a large influence over gene expression, the transcript's cellular localization, stability and translation rate (10–14). Since it was first discovered in the immunoglobulin M and the DHFR genes (15–17), technological advances and high-throughput sequencing techniques have led to results revealing that APA is more widespread than initially thought; 30–70% of an organism's genes undergo APA across diverse species (1–5,18,19). During the last decade, widespread APA alterations were detected across different tissues and distinct stages of embryogenesis (4–6,8,20–25). However, beyond these classifications and functional characterization of a short list of single gene APA alterations (26) the global functional significance of APA remains largely elusive.

Caenorhabditis elegans is a convenient model organism for studying APA and gene expression during embryogenesis, since the cell lineage is invariant and has been fully traced (27,28). *C. elegans* was the first multicellular organism to have a fully sequenced genome (29); and its transcriptome has also been well characterized (30–35). During embryogenesis, the *C. elegans* transcriptome is highly dynamic; in early stages it is comprised mostly of maternal transcripts, but as development proceeds, zygotic transcription commences, and maternally supplied transcripts undergo degradation (8,9,36–37). Our previous results detected many genes with a dynamic overall gene expression profile throughout embryogenesis, as well as genes with constitutive levels of expression (38).

CEL-Seq is a sensitive multiplexed single-cell RNA-Seq method (39,40). One important feature of the CEL-Seq method is that it is largely restricted to studying the 3′ends of transcripts and thus measures overall expression levels; typically collapsing the various isoforms produced by a gene to one summary profile. While this has been a useful simplifying criterion, it does ignore possible dynamic profiles across different splicing isoforms of a particular gene. An important advantage of the CEL-Seq 3′end bias though is that this information can be used to detect and quantify al-

*To whom correspondence should be addressed. Tel: +1 646 501 4603; Email: itai.yanai@nyulangone.org
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

ternative polyadenylation patterns, i.e. 3′ UTR isoforms of the same gene.

Here, we studied APA profiles in individual *C. elegans* throughout embryogenesis using APA-seq, an approach to detect alternative polyadenylation profiles at a genomic scale. APA-seq is based on CEL-Seq but further exploits the information from both paired-end reads for gene identification from Read 2 and the exact location of polyadenylation from Read 1. Combining this information empowers quantitative expression level assessment globally at both the gene and APA isoform level. Using this approach, we delineated two groups of genes: those with highly and lowly correlated 3′ UTR isoform groups (HCI and LCI), respectively. We detected unique regulatory features between these groups that support the notion that variants across these two groups are under distinct selective biases. Genes with uncorrelated 3′ UTR isoform expression (LCI) are predicted to have the highest miRNA regulation compared to genes with well-correlated 3′ UTR isoforms. Integrating extensive previously published embryonic mRNA and protein expression datasets (35,41), we also found the lowest correlation between total mRNA transcript and protein levels in genes with dynamic 3′ UTR isoform expression (LCI). Extending this analysis to *Drosophila melanogaster* and *Xenopus laevis* datasets we found a consistent relationship (42–46). Together, our results suggest that genes of the HCI and LCI groups experience distinct regulatory pressures upon their alternatively polyadenylated isoforms.

MATERIALS AND METHODS

Detection of polyadenylation site using CEL-Seq reads

We used our previously published *C. elegans* time-course data (GSE50548) sequenced using the CEL-Seq protocol and paired-end 100 bp sequencing mode. The CEL-Seq Read 2 insert was used to identify the gene by mapping reads to the reference genome (version WS230) using Bowtie 2 version 2.2.3 (47) with default parameters. The htseq-count algorithm (19) coupled with the genomic feature file was used to assign each individual Read 2 to its gene of origin. For each gene we extracted the whole gene coding sequence as well as the 5000 nucleotides downstream of the stop codon, or fewer than 5000 nucleotides if another gene was found in closer proximity. We truncated Read 1, removing the barcode sequence and the polyT stretch, leaving a sequence of ∼70 nucleotides. We then used Bowtie 2 (47) in order to map Read 1 exclusively to the coding sequence of the identified gene. The maximum and minimum mismatch penalty (–mp MX,MN) parameters for Bowtie 2 were set to 2 and 1, respectively. To summarize the data for each sample, we counted all 3′-most mapping locations of truncated Read 1 to the respective genes up to a distance of 20 nucleotides upstream of the polyadenylation site. We predicted APA sites only for those genes passing a threshold of 20 mapped reads in at least two samples. For these genes, we identified the peaks representing the polyadenylation sites by summarizing the last mapping coordinates of all the truncated Read 1 entities that mapped to a specific gene using the 'findpeaks' function in MATLAB. Peaks were required to be separated by at least 20 nucleotides in order to be considered as distinct. The height threshold for a peak was 5 reads, or 1/1000 of the total gene expression in a particular sample. We then filtered out possible spurious peaks that may have resulted from internal priming, by removing any peak whose downstream genomic sequence included any of the following nucleotide combinations: AAAA, AGAA, AAGA or AAAG (48). The exact coordinate of any polyadenylation site was defined as the most 3′ coordinate of the respective peak. To validate the quality of our data, we compared our polyadenylation site annotation with a previously published *C. elegans* 3′ UTR annotation (41). We performed this by measuring the difference between the mapping positions of polyadenylation sites from both data sets (see Figure 1D).

3′ UTR isoform expression throughout *C. elegans* development

Expression data for each sample was obtained by counting all Read 1 sequences whose 3′-most mapping location mapped up to a distance of 20 nucleotides upstream of the polyadenylation site. The raw expression data was then converted to transcripts per million (tpm) by dividing by the total reads and multiplying by one million. We worked with $\log_{10}$ values unless otherwise noted. The profiles were further smoothed by computing a running average over 5 time points. To study dynamics of 3′ UTR isoform usage throughout the *C. elegans* embryonic time-course, we first filtered genes on overall expression, keeping only those in the upper 85 percentile of the sum of expression throughout the time-course. Ratios between the two most highly expressed 3′ UTR isoforms where calculated for all genes and all stages by dividing the expression levels (tpm) of the short by the levels of the long 3′ UTR isoform. Only genes whose 3′ UTR isoform ratios across time differed by at least a factor of three were kept for further analysis ending up with 305 genes. Significance of the changes in the ratios during the transition between successive periods were calculated using Student's *t*-test on all ratios of one and the ratios of the successive period. *P*-values and fold changes are shown in Figure 2B. To generate temporally sorted expression profiles, we used 'ZAVIT' as previously described (49,50).

Categorization of genes by APA behavior

To determine whether total gene expression is considered static or dynamic throughout development, we used the ratio of minimum to maximum and as validation the interquartile range (IQR) of expression levels across time. To quantify 3′ UTR variant expression deviations, we calculated Pearson correlation for the expression pattern of the two, or more, 3′ UTR isoforms. For genes with more than two isoforms (< % of expressed genes), we used the minimal Pearson correlation between any pair of isoforms. Highly correlated 3′ UTR isoforms (HCI) are those with $r > 0.7$, while lowly correlated 3′ UTR isoforms (LCI) are those with $r < 0.3$.

miRNA target analysis

3′ UTR sequences were identified according to APA-seq. After removing gene coding sequences using WormBase

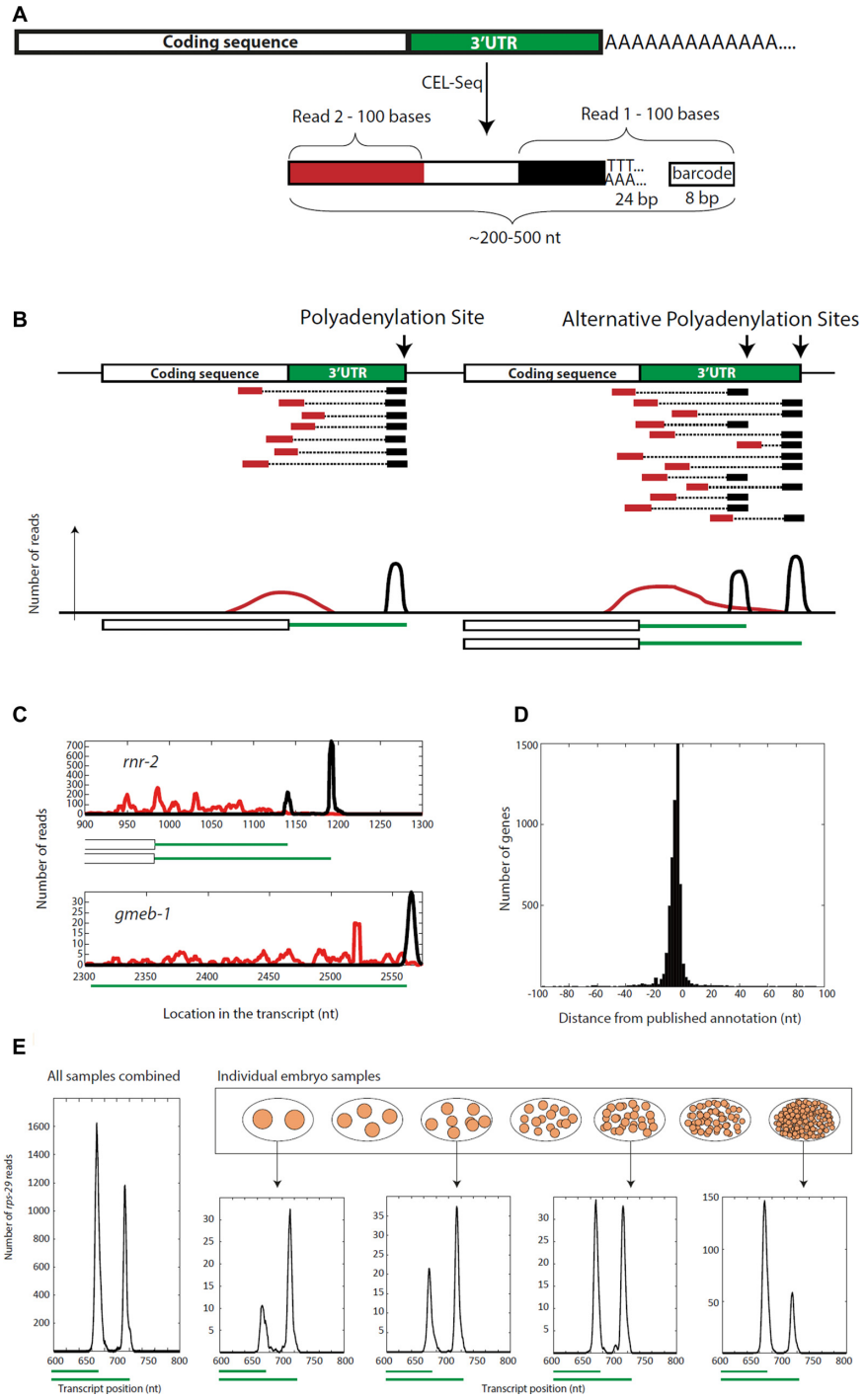

Figure 1. APA-seq measures expression levels of distinct 3′ UTR isoforms in individual *C. elegans* embryos. (**A**) APA-seq identifies gene expression levels for distinct alternatively polyadenylated isoforms. APA-seq is an adaptation of the CEL-Seq method which utilizes paired-end reads: Read 1 contains a sample-specific barcode while Read 2 identifies the transcript. If sequenced long enough (100 bp in our case), Read 1 also provides information on the exact location of the polyadenylation site. APA-seq thus uses Read 2 to identify the expressed gene and then maps Read 1 to the gene specific region, thus enabling unique mapping in spite of the low sequencing quality that results from sequencing through the low-complexity poly-T region. (**B**) Plotting the 3′ends of Read 2 sequences results in a wide distribution within the gene (red peak) due to the fragmentation step of library preparation. However, the 3′ends of Read 1 sequences all map to the site immediately upstream of the poly-A tail thus producing a clear peak (black peak) when mapped to the gene sequence and revealing exact polyadenylation sites. The white boxes at the bottom of the distribution plots mark the coding sequence, while the green lines indicate the determined 3′ UTR regions. (**C**) Using the APA-seq method enables detection of the exact location of polyadenylation sites in *C. elegans*. The green lines at the bottom of each plot mark previously annotated 3′ UTRs (41) for the indicated gene, showing good agreement with the APA-seq Read 1 peaks (black). (**D**) Global comparison of the detected polyadenylation sites using APA-seq to the *C. elegans* UTRome annotation (41) shows high consistency between the two datasets. (**E**) The expression of two unique 3′ UTR isoforms for the *C. elegans* rps-29 gene throughout embryogenesis. Although in total (leftmost panel) the 3′ UTR variants show equal expression, examining expression across developmental stages shows predominant expression of the shorter 3′ UTR variant during early embryogenesis, and the inverse later in development.

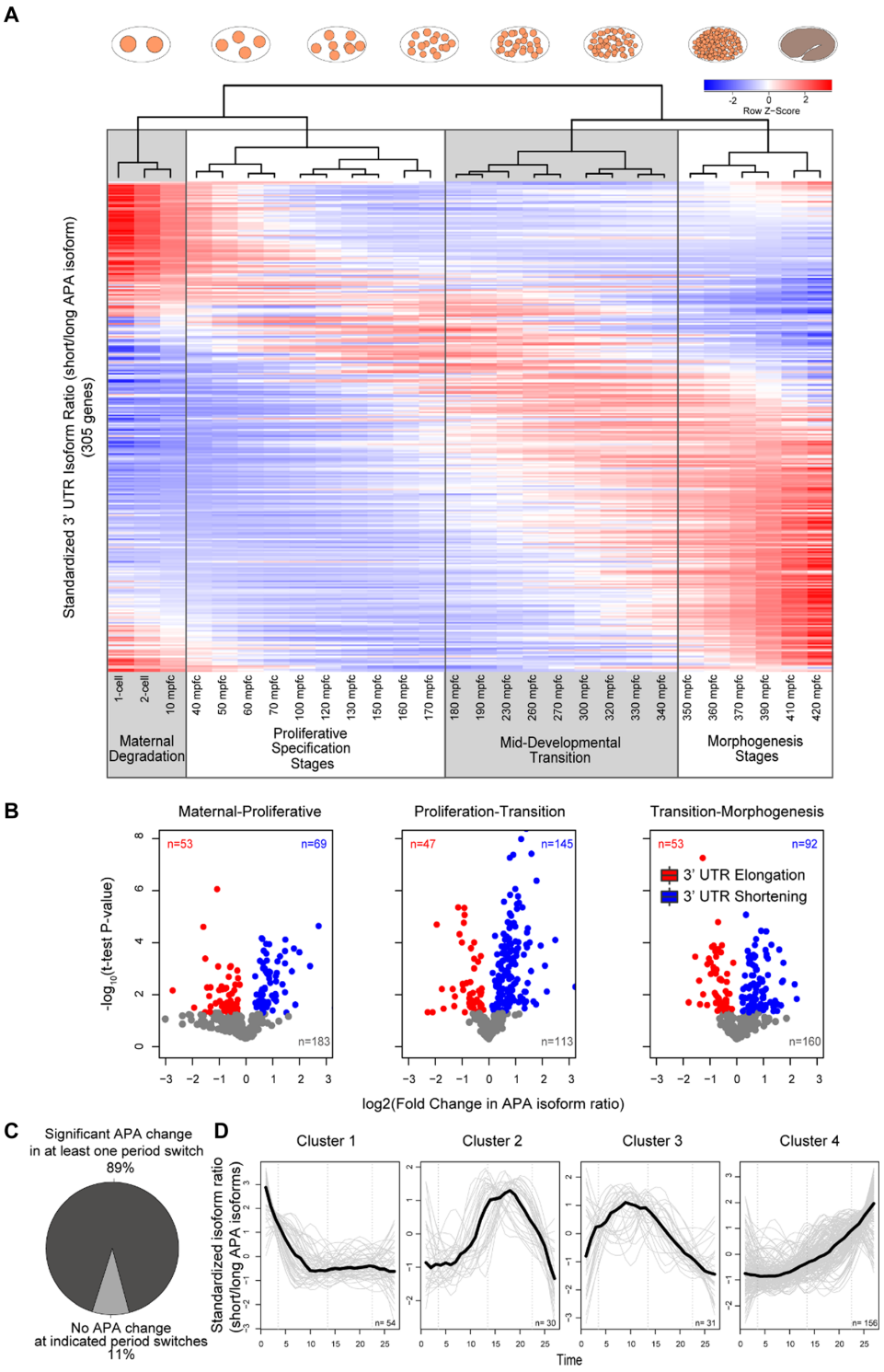

Figure 2. 3′ UTR isoform expression throughout *C. elegans* embryogenesis. (**A**) Heatmap showing the relative expression of 3′ UTR isoforms for 305 genes which passed overall expression and 3′ UTR isoform dynamics threshold throughout *C. elegans* embryogenesis. Each row in the heatmap corresponds to a gene and indicates the ratio between the expression of its short and long isoforms. Red and blue indicate the maximum and minimum 3′ UTR ratio for each gene, respectively. White and grey shadowed boxes indicate sets of developmental stages with similar isoform usage (based on Ward clustering, see clustergram on top of the heatmap) and identify the periods of major isoform switches. mpfc = minutes past four-cell stage. (**B**) The plots indicate the fold-change and *P*-value of difference in the isoform ratios for pairs of successive periods as defined by Ward clustering in (A). Red and blue colouring indicates significant elongation and shortening events, respectively. Grey colouring indicates insignificant changes. (**C**) Most of significant 3′ UTR isoform switches occur between the identified developmental switch periods. 89% (271) of the genes show significant changes in 3′ UTR usage in at least one of the identified period switches. (**D**) Clusters of significant 3′ UTR isoform changes indicate four main patterns—almost constant elongation or shortening over time (Clusters 1 and 4, respectively) or peaking shortening during mid-developmental transition and proliferative periods (Clusters 3 and 4, respectively). The thick black line indicates the mean 3′ UTR isoform ratio profile of all genes in the cluster and individual genes profiles are shown as thin grey lines.

annotation (51), we annotated miRNA binding sites by searching for the basic seed match sequence within the 3′ UTR, i.e. the complementary sequence for nucleotides 2–7 of the miRNA (52,53). We disregarded other parameters such as type of seed match or complementarity outside of the seed region. We then counted the number of miRNA binding sites for each gene subgroup (LCI and HCI genes), and used Student's t-test to determine the significance of the difference in the number of miRNA binding sites between the LCI and HCI groups. To determine the effect specific miRNAs have on their targets, in genes with multiple 3′ UTR isoforms, we calculated Pearson correlation between the isoform expression ratio of the two most highly expressed isoforms, and the miRNA expression profile (54). For this analysis, we examined only dynamically expressed miRNAs, whose expression is >250 transcripts overall across the timepoints. To compare the miRNA dataset with our APA-seq dataset, we examined only matching timepoints to our time-course and used the Matlab 'imresize' function followed by smoothing to stretch the miRNA expression data.

mRNA-protein correlation analysis

RNA-Seq and Silac data of different developmental stages in *C. elegans* was downloaded from Grün *et al.* (35). Replicates were averaged and data was normalized by division by the sum of all genes for the specific samples. Correlation between rpkm (mRNA) and Silac (protein) expression values using Pearson correlation was computed only for genes with rpkm and silac values for at least three stages. A similar approach was used to calculate the correlation between RNA and protein levels for the Drosophila melanogaster time-course. APA data of an embryonic time-course was downloaded from Sanfilippo *et al.* (44); 3′ UTR isoform ratio was analysed similarly to our time-course data yielding correlation between 3′ UTR isoforms for all multi-isoform genes. Highly correlated 3′ UTR isoforms (HCI) are those with $r > 0.7$, while lowly correlated 3′ UTR isoforms (LCI) are those with $r < 0.3$, coherent with the thresholds used for our dataset. Total mRNA and protein expression levels were downloaded from Graveley *et al.* (42) and Casas-Vila *et al.* (43), respectively. The two datasets were integrated by correlating mRNA (rpkm) and protein (lfq) expression levels using Pearson correlation. The same procedure was used to analyse the data for *Xenopus laevis*. mRNA RNA-Seq and protein LFQ data for embryonic time points were downloaded from (46) and APA data from (45). For the datasets from all three species, the Mann–Whitney test was used to determine if LCI genes show different mRNA-protein expression correlations from HCI genes.

RESULTS

APA-seq identifies expression levels of alternative polyadenylation isoforms

Paired-end reads generated by CEL-Seq and CEL-Seq2 contain the sample-specific barcode on Read 1 and the sequence identifying the transcript on Read 2 (39,40) (Figure 1A). Thus, typically only Read 2 is used for measuring gene expression levels. However, since CEL-Seq sequences the 3′ ends, it can be used in principle to identify 3′ UTR isoforms. Using Read 2 for this purpose is not accurate due to the uneven sizes of the inserts in the sequencing library, producing a smear of mapped reads, which makes distinguishing different 3′ UTR isoforms of the same transcript impossible in many cases (Figure 1B, red peak). We noted though that Read 1 includes the actual 3′end of the transcript (Figure 1A), located just upstream of the polyadenylation site (Figure 1B, black peak). In CEL-Seq, this region follows 24 'T's used for capturing the polyA tail of the transcript, and the sequencing quality is relatively poor after this low complexity region, rendering conventional mapping impossible (Supplementary Figure S1A and B). To overcome this, we found that the sequence is still of sufficient quality when mapping is restricted to a particular region of the genome, using relaxed parameters. Thus, while permissive Read 1 mapping matches many genomic loci due to its poor quality, it maps accurately when restricted to the sequence of a particular gene, whose identity is detected by using Read 2.

We exploited this approach to devise the APA-seq method to study 3′ UTR isoforms using both Read 1 and Read 2 information. We applied APA-Seq to study the expression of alternative polyadenylated isoforms during early embryogenesis in the nematode *C. elegans*. For this, we used a dataset previously published by our lab in which embryos were individually collected and sequenced throughout embryogenesis (49). Pooling together all samples, we detected distinct 3′ UTR isoforms for 4,871 genes using APA-seq, after removing possible artifacts caused by internal priming (48) (see Methods). These were not biased according to expression level (Supplementary Figure S2). For example, the *C. elegans* genes *gmeb-1* and *rnr-2* show a smear of Read 2′s 3′-most mapping locations (Figure 1C, red lines), while Read 1′s 3′-most mapping locations form distinct peaks (black lines) identifying previously characterized polyadenylation sites (green lines) (41) (see also Supplementary Figure S1C).

Overall, we detected multiple 3′ UTR isoforms for 14% of the expressed genes in our dataset (699 out of 4871 genes). Of these, <1% have more than two isoforms (Supplementary Figure S1D). This set is not comprehensive to all possible 3′ UTR isoforms produced by the *C. elegans* genome given that we only examined 430 min during embryogenesis, and the stringent thresholds set in our bioinformatics pipeline (in terms of mapping parameters, mapping level filtering, minimal distance between isoforms, and spurious site removal, see Materials and Methods). Moreover, this rate of genes with multiple 3′ UTR isoforms is comparable to that observed in other studies (5,18,23,24). We assayed the accuracy of the isoforms detected by comparing our polyadenylation sites with a known repository of 3′ UTR annotations in *C. elegans* (41) and found highly concordant profiles, with 95% of sites corresponding to well-established annotated sites (Figure 1D, Supplementary Table S1). The remaining 5% show significantly lower expression levels than the annotation overlapping 3′ UTR isoforms (Supplementary Figure S1E) and we therefore excluded these from further analyses by expression level filtering. To study the temporal dynamics of 3′ UTR isoform expression, we classified the reads according to their embryo of origin (Figure 1E, Supplementary Table S1). As an

example, two isoforms were identified for *rps-29*, which encodes a ribosomal protein subunit (55). Interestingly, while the sum of expression of both *rps-29* isoforms is roughly constant over time, the long 3′ UTR isoform is predominantly expressed in the early stages while the shorter is expressed in later embryos (Figure 1E). Correlating the expression levels of all 3′ UTR isoforms across stages, we found that successive stages (near-replicates) show high correlations thus highlighting the reproducibility of the data (Supplementary Figure S3). We conclude that while CEL-Seq is typically used to assay overall expression levels with the mapping location typically ignored, processing the reads using APA-seq can identify the exact locations of alternative polyadenylation sites and the expression levels of distinct 3′ UTR isoforms.

Alternative polyadenylation profiles throughout *C. elegans* embryogenesis

Our dataset enabled us to study overall patterns of 3′ UTR isoform usage throughout early development in individual embryos. For each gene, we first computed the ratio of expression between its 3′ UTR isoforms throughout the time-course. Interestingly, the clustering of stages according to these profiles revealed that adjacent developmental stages show concordant 3′ UTR isoform usage patterns which group into distinct periods with diverging APA dynamics (Figure 2A). The four groups correspond to previously characterized developmental periods: maternal degradation, early period of extensive proliferation and specification, the mid-embryonic transition period and the subsequent period of morphogenesis (49). The observation that different periods have different corresponding patterns of 3′ UTR isoforms may reflect that each of these periods has a distinct functional requirement. Between adjacent periods we found that the direction and level of change of polyadenylation site usage generally shows a broad burst of shortening especially between the proliferative and mid-embryonic transition periods (Figure 2B). This is consistent with previous work indicating that proliferative states show a general trend for 3′ UTR shortening (14). Overall, we identified that 89% of dynamic 3′ UTR isoform genes show APA switches in at least one of the period switches indicated (Figure 2C).

Clustering the 3′ UTR isoform ratio profiles throughout the time-course, we identified four main distinct clusters (Figure 2D, Supplementary Table S1): genes whose 3′ UTRs elongate or shorten over time (Clusters 1 and 4, respectively) and genes whose 3′ UTRs shorten transiently during the proliferation period (Clusters 2 and 3). The majority of genes, however, exhibited continuous shortening of their 3′ UTR regions with progressing development (Cluster 4). Interestingly, these genes show enrichments for MAP kinase cascade, morphogenesis and neuronal differentiation (Supplementary Figure S4). All these functions are crucial components of the switch from germ cell biology to processes involved in proliferation and differentiation. Other isoform patterns may be present throughout embryogenesis, requiring more high-resolved data for their characterization. In summary, we show that the 3′ UTR dynamics detected by APA-seq reflect characteristic functionalities of embryonic development, though the significance of individual events is unclear. The contribution of APA events to the general regulatory states involved in these events might be crucial or, alternatively, they may be a neutral consequence of the vast changes occurring across developmental periods at the other levels of gene expression regulation.

Genes with constitutive total expression are enriched for dynamic 3′ UTR isoform expression

Examining the temporal profiles, we found that 3′ UTR isoforms of a particular gene may exhibit striking dynamics throughout development, while the total expression for the gene may be uniform (Figure 3A). To study this systematically, we defined the dynamic range of a gene's overall expression profile as the fold differences between the maximum and minimum expression values throughout the time-course. We found a positive correlation between the dynamic overall expression range of a gene (binned) and the correlation among its isoforms (Figure 3B; $r = 0.94$, $P = 0.03$, second degree polynomial regression test, $N = 746$). Similar results were obtained using the interquartile range (IQR) of the time-course overall expression levels as a proxy for expression dynamicity ($r = 0.98$, $P = 0.04$, third degree polynomial regression test, $N = 746$). This result provides evidence for the notion that apparently constitutive genes can be highly dynamic at the level of their individual 3′ UTR isoforms.

Gene expression levels may explain this correlation, if genes with less dynamic behavior are also lowly expressed, and in turn may have noisier and uncorrelated 3′ UTR isoform profiles. To control for this possibility, we asked if there is a trend in expression levels according to the correlation among isoforms (Figure 3C). Interestingly, genes with non-congruent APA behavior (LCI genes) also show higher expression levels (Figure 3C; $r = 0.99$, $P < 0.002$, second degree polynomial regression test, $N = 746$). This analysis effectively eliminates expression noise as a confounding factor between the lowly and highly correlated isoforms.

To further examine this phenomenon, we visualized the dynamics of 3′ UTR isoform expression and total expression in genes with highly or lowly correlated isoforms. The 746 genes with multiple 3′ UTR isoforms displayed a variety of isoform expression correlations (Supplementary Figure S5). More than 70% of the genes (545 genes) have highly correlated 3′ UTR isoform expression ($r > 0.7$), while 11% of the profiles (79 genes) are lowly correlated ($r < 0.3$). We henceforth refer to the two gene sets of highly and lowly correlated 3′ UTR isoforms, as HCI and LCI, respectively. As the heat maps in Figure 3D and E show, we found dynamic expression for both the HCI and LCI groups at the 3′ UTR isoform level. As expected, the total gene expression for the correlated isoforms is dynamic, recovering the dynamics of the isoforms. Conversely, the total expression of the LCI genes mostly corresponds to profiles with constitutive overall expression. Such profiles are frequently attributed as housekeeping profiles, however as our analysis reveals, at the 3′ UTR isoform level they may be very dynamic. Thus, by de-convolving the total expression into profiles of distinct 3′ UTR isoforms we were able to extract a new layer of information from this dataset.

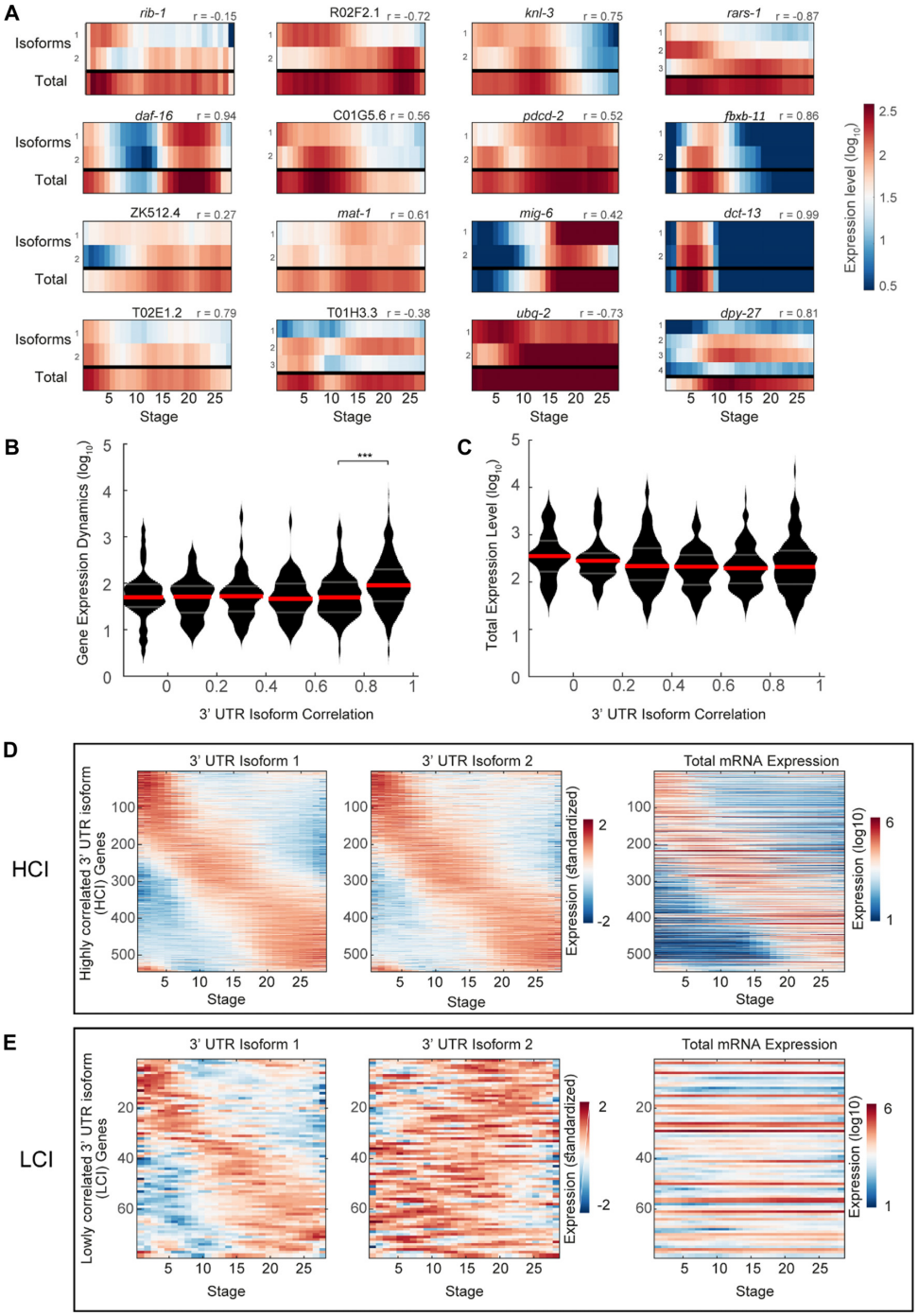

Figure 3. Genes with constitutive overall expression have dynamically expressed 3′ UTR isoforms. (**A**) Expression heatmaps indicating 3′ UTR isoform expression for 16 genes with multiple isoforms displaying varying levels of correlation between the expression of their 3′ UTR isoforms. The Pearson correlation coefficient r for each of the genes displayed is indicated at the top of each heatmap. (**B**) Relationship between 3′ UTR isoform expression correlations and the overall expression dynamics. Genes whose 3′ UTR isoform expression levels are correlated are more dynamic in their overall expression ($r = 0.94$, $P = 0.03$, second degree polynomial regression test). Dynamics for each gene is defined as the fold differences between its maximum and minimum expression values throughout the time-course. Red and gray horizontal bars represent the median and the interquartile ranges of the data, respectively. The last bin with highest 3′ UTR isoform correlations exhibit significantly higher overall expression dynamics than the preceding bin ($P < 10^{-20}$, Mann–Whitney test). (**C**) Relationship between 3′ UTR isoform expression correlations and the respective overall total mRNA expression levels. Genes whose 3′ UTR isoform expression levels are uncorrelated show significantly higher overall expression levels ($r = 0.99$, $P = 0.002$, second degree polynomial regression test). Red and grey horizontal bars represent the median and the interquartile ranges of the data, respectively. (**D**) Heatmaps in the two leftmost panels show standardized expression of the 3′ UTR isoforms of 545 genes belonging to the group of genes whose 3′ UTR isoform's expression show high correlation (HCI genes). The right panel shows a heatmap of the total mRNA expression (on a $\log_{10}$ scale) of the same genes confirming that genes with highly correlated isoform expression are also dynamically expressed. The time points are the same as indicated in Figure 2A. (**E**) Same as D for 79 genes belonging to the group of genes whose 3′ UTR isoform's expression show low correlation (LCI genes). Genes with distinct expression profiles of 3′ UTR isoforms appear constitutively expressed when examined at the total gene expression level.

Delineating the LCI and HCI groups, led us to hypothesize that 3′ UTR isoforms from the former group are under stronger selective pressures, and are consequently more likely to be functionally different. We thus set out to test this hypothesis by studying the post-transcriptional and translational regulatory characteristics of these two gene sets.

LCI genes contain more miRNA binding sites and correlate with miRNA expression

The 3′ UTR region is known to be a locus of considerable post-transcriptional regulation (56,57), and the role of miRNAs in this regulation is well evidenced (52,53). We thus searched for evidence that our detected HCI and LCI 3′ UTR isoform expression profiles are regulated by miRNAs. Specifically, we asked if genes with a different number of 3′ UTR variants (single versus multiple) and different 3′ UTR isoform expression correlations (HCI versus LCI) have distinguishing sequence properties related to miRNA regulation (Figure 4A). We first counted the number of basic miRNA seed matches in the 3′ UTR sequence of the different groups (58). We found that genes with more than one 3′ UTR variant have significantly more miRNA binding sites than genes with a single 3′ UTR isoform ($P < 10^{-12}$, Mann–Whitney test, Figure 4B). Genes with multiple variants also have significantly longer 3′ UTRs ($P < 10^{-28}$, Mann–Whitney test, Figure 4C), though the number of miRNA binding sites per base is not different across the groups (Supplementary Figure S6). Thus, as expected from previous work, genes with multiple variants are predicted to be regulated more actively by miRNAs.

Comparing the number of binding sites between the HCI and LCI groups, we found that the latter group has significantly more miRNA binding sites in their 3′ UTR ($P < 0.02$, Mann–Whitney test, Figure 4B). We further studied this group of genes by examining the sequence that is unique to the longer 3′ UTR isoform (Figure 4A). Comparing between the HCI and LCI groups, we found that the latter have significantly more miRNA binding sites in their unique 3′ UTR region ($P < 10^{-3}$, Mann–Whitney test, Figure 4D). This is not because LCI genes have denser distribution of miRNAs, rather their unique 3′UTR regions are significantly longer than in HCI genes ($P < 10^{-4}$, Mann–Whitney test, Figure 4E), thus they tend to carry more potential miRNA binding sites. These results implicate a role for miRNAs in the differences we see between the genes sets of correlated and uncorrelated isoform genes.

To further examine the effect that miRNA regulation exerts on HCI and LCI genes, we computed correlations between the expression profiles of all dynamically expressed miRNAs (54) and the 3′ UTR isoform expression ratio of the genes they regulate. For example, *cel-miR-1-5p* is conserved in *C. briggsae* as well as in *Drosophila*. Expression of *cel-miR-1-5p* has been detected during early morphogenesis (54), and interestingly, its profile has a correlation of 0.94 ($P < 10^{-13}$) with the ratio of the 3′ UTR isoform expression of its target gene *F07C6.4*, a gene which is enriched in the germ line, germline precursor cell, the body wall musculature and in the PVD and OLL neurons (59,60) (Figure 4F). Figure 4G shows the distributions of correlation coefficients for dynamically expressed miRNAs with the 3′ UTR isoform expression ratio of their targets in the HCI and LCI groups. These ten miRNAs were selected based upon the most significant difference between the target correlations of the HCI and LCI gene groups (out of 64 miRNAs with dynamic expression- statistics of all miRNAs can be found in Supplementary Table S2). We found that in all ten the correlations are significantly higher in LCI than in HCI genes. For example, *cel-miR-2-5p* shows significantly higher correlations with the expression ratio of the isoforms of all the genes it potentially binds and regulates in the LCI genes ($P < 0^{-3}$, Mann–Whitney test, Figure 4G). This analysis suggests that miRNAs play a major role in regulating genes with lowly correlated 3′ UTR isoforms (LCI) during *C. elegans* embryogenesis.

LCI genes exhibit lower mRNA–protein correspondences

Beyond regulation by miRNA, control of translation efficiency constitutes another level of post-transcriptional regulation. 3′ UTR regions are known preferential targets for RNA binding proteins that regulate translation in terms of localization and efficiency (61–63). Indeed, mRNA and protein levels are notoriously lowly correlated (35). We reasoned that one possible explanation for the low correspondence of some genes may follow from the fact that different 3′ UTR isoforms may exhibit different translation efficiencies. Thus, we predicted that LCI genes would have a worse correspondence—relative to HCI genes – between transcription and protein levels. To test this, we turned to a previously published mRNA and protein *C. elegans* time-course (35) and examined the distribution of correlations between mRNA and protein abundances across the HCI and LCI groups. Our data suggested (but did not provide significant evidence) that LCI genes show lower correlations between mRNA and protein abundances, relative to the HCI genes (Figure 5A; $P = 0.07$, Mann–Whitney test; $N = 188$, 30 for HCI and LCI, respectively). This limited significance may be due the low number of detected LCI genes following the shortness of the time-course, and the restricted protein data (only ∼25% of the RNA-Seq detected transcripts were detected at the protein level). To further test the prediction we turned to Drosophila where an available high-resolution time-course allowed us to detect more LCI and HCI genes. Coupling the 3′ UTR isoform expression throughout embryogenesis (44) with total mRNA (42) and protein expression data (43), we delineated LCI and HCI genes (see Materials and Methods) and studied the correlation between their transcription and protein levels in a *D. melanogaster* embryonic time-course. As in *C. elegans*, we found that LCI genes exhibited reduced mRNA to protein expression correlation relative to HCI genes (Figure 5B; $P = 0.00038$, Mann–Whitney test; $N = 571$, 189 for HCI and LCI, respectively). We further analyzed three extensive embryonic mRNA, protein and APA datasets from *Xenopus laevis* (45,46) and observed similarly highly significant trends for this vertebrate species (Figure 5C; $P < 0^{-6}$, Mann–Whitney test; $N = 1953$, 996 for HCI and LCI, respectively). These results suggest that weak correlations between mRNA and protein may be in part explained by the existence of LCI genes with isoforms with distinct translation efficiencies.

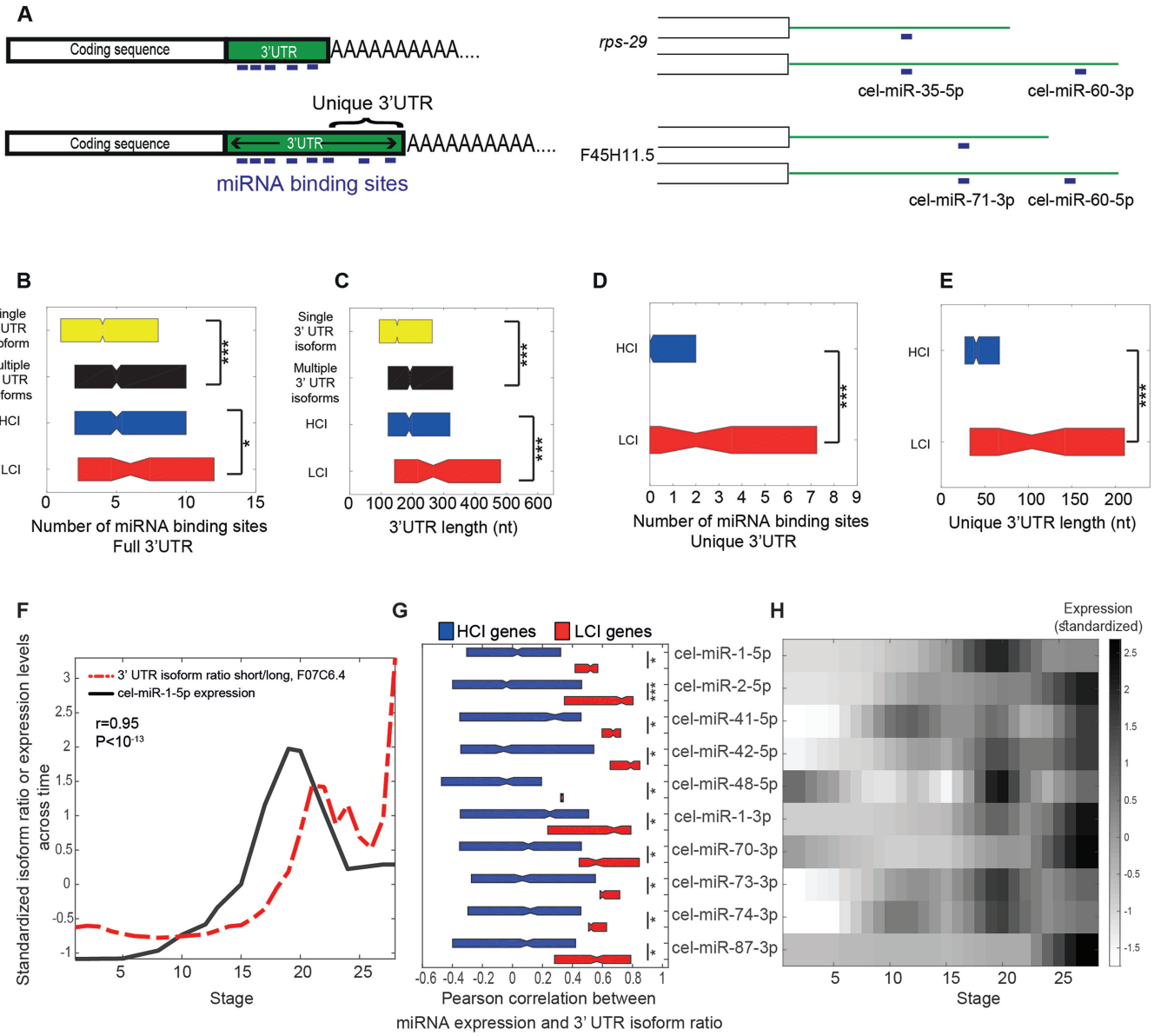

Figure 4. Genes with multiple, lowly correlated 3′ UTR variants show evidence for increased miRNA regulation. (**A**) The left panel shows a schematic representation of the miRNA analysis. The length of the 3′ UTR region, the number of basic miRNA seed matches, and the 3′ UTR region unique to the longer 3′ UTR isoform were considered. The right panel shows two examples of genes and their miRNA binding sites. (**B**) Boxplots indicating the number of miRNA binding sites in the full 3′ UTR regions of genes with single or multiple 3′ UTR isoforms, highly correlating (HCI) or lowly correlating isoforms (LCI). Genes with multiple 3′ UTR variants have significantly more miRNA targets than genes with a single variant ($P < 10^{-12}$, Mann–Whitney test). Between multiple 3′ UTR isoform genes, LCI genes have significantly more miRNA binding sites than HCI genes ($P < 0.02$, Mann–Whitney test). (**C**) Same as B for the full 3′ UTR lengths of genes. Genes with multiple 3′ UTR variants have significantly longer 3′ UTRs than single isoform genes ($P < 10^{-28}$, Mann–Whitney test). Within multiple isoform genes, LCI genes have significantly longer 3′ UTRs than HCI genes ($P < 10^{-3}$, Mann–Whitney test). (**D**) Boxplots indicating the number of miRNA binding sites in the unique 3′ UTR region of the longer 3′ UTR isoform across HCI and LCI genes. LCI genes have significantly more miRNA binding sites in their unique 3′ UTR region than HCI genes ($P < 10^{-3}$, Mann–Whitney test). (**E**) Boxplots indicating the length of the unique 3′ UTR region across HCI and LCI genes. LCI genes have significantly longer unique 3′ UTR regions than HCI genes ($P < 10^{-4}$, Mann–Whitney test). (**F**) Correlating expression of miRNA expression with expression ratio dynamics between 3′ UTR isoforms. The black line shows the expression profile of *cel-miR-1–5p* miRNA throughout the developmental time-course. Depicted in red is the ratio between the expression profiles of the two 3′ UTR isoforms (short/long) of the *F07C6.4* gene. The Pearson correlation coefficients between 3′ UTR isoform ratio and miRNA expression of all target genes were used for the analysis shown in G and H. (**G**) For the indicated miRNAs, the boxplots show the distribution of correlations between miRNA expression and 3′UTR isform ratio showing a significant difference between the HCI and LCI gene groups (out of 64 miRNAs with dynamic expression, see Supplementary Table S2 for statistics for all miRNAs). All ten exhibit a positive median correlation between miRNA expression and the 3′ UTR isoform expression ratio of genes it is predicted to bind (as in F). Further, this correlation is significantly higher in LCI than in HCI genes. (**H**) Heatmap of the standardized expression dynamics of the ten indicated miRNAs which show differences between HCI and LCI genes across development.

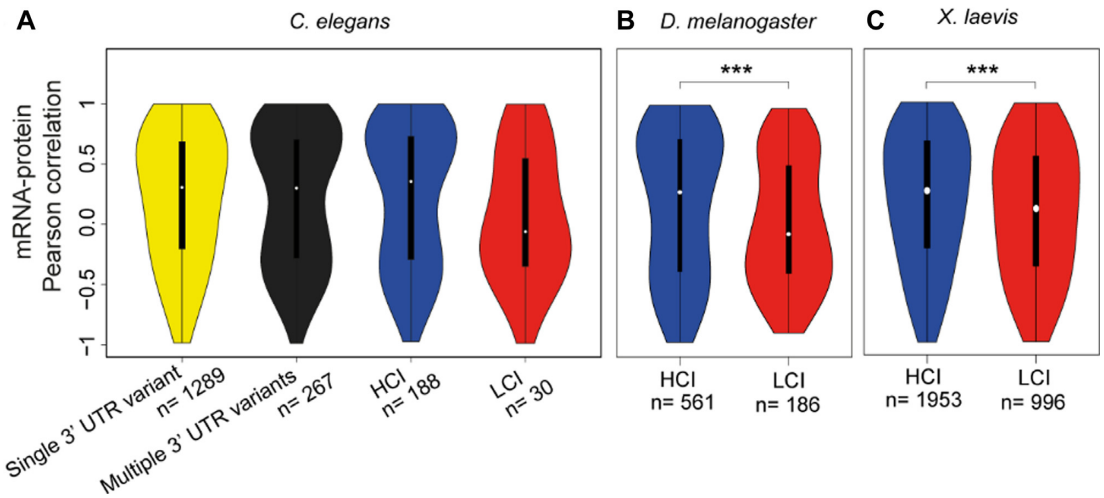

Figure 5. Genes with lowly correlated 3′ UTR isoforms (LCI genes) have a lower correspondence between total mRNA and protein expression. (**A**) Violin plots indicate the distribution of Pearson correlation coefficients between total mRNA and protein expression levels across developmental stages for *C. elegans* genes with one and multiple 3′ UTR isoforms, highly correlating (HCI), and lowly correlating isoforms (LCI). (B, C) Same as A, for LCI and HCI genes in *Drosophila melanogaster* (**B**) and *Xenopus laevis* (**C**) embryonic development.

DISCUSSION

The APA-seq approach allows for the extraction of an additional layer of data from CEL-Seq data. In addition to quantifying the expression of each gene, APA-seq reveals the alternative polyadenylation dynamics on a transcriptomic basis. APA-seq uses the CEL-Seq Read 2 to identify the gene of origin and then maps Read 1 to the respective gene sequence, instead of to the whole genome, enabling the use of even extremely low-quality sequencing data resulting from reading through protocol-conditioned poly-T stretches. Hence, by performing a subtle modification in data analysis without altering the actual CEL-Seq protocol, we were able to switch from analysis at the gene to the APA level. The accuracy of APA-Seq for detecting isoforms is evidenced by comparisons with other datasets (Figure 1), previous observations of isoform shortenings over developmental time, and predicted miRNA regulation. Although the presented data is based on CEL-Seq data, our method is in principle applicable to any RNA-Seq method that uses the polyA tail as anchor and performs paired end sequencing; including inDrop, 10X, and SMART-Seq (64–66). Other 3′ end sequencing methods have enabled important insights into the biological significance and mechanistic aspects of APA, though their experimental procedures require either relatively large amount of starting material or complex protocols combining several amplification steps (IVT and many PCR cycles) (3,19,67–75).

In addition to APA-seq, two methods are available for assessing APA in low-input samples: BATSeq (19) and ScISOr-Seq (74). BATSeq achieves single-cell APA isoform measurements. An important advantage of APA-seq relative to BATSeq is the formers high sensitivity, deriving from its use of CEL-Seq data; employing far less amplification and clean-up steps and PCR cycles, and a simple protocol and analysis. ScISOr-Seq also constitutes pioneering single-cell isoform work with the advantage of revealing the complete isoform due to its reliance on long-read sequencing. Relative to ScISOr-Seq, APA-Seq has the advantage of higher statistical confidence due to its reliance on deep Illumina sequencing. We highlight that APA-seq is not aimed towards the identification of polyA sites *de novo* but rather we present it as an efficient method for quantifying the expression profiles of previously mapped isoforms. Furthermore, while we applied APA-seq here to single-embryo CEL-Seq data, it could in principle also be applied to single-cell or other low-input RNA samples. As the protocol is relatively straight-forward omitting unnecessary clean-up and size-selection steps, the efficiency, complexity and accuracy is as high as that in CEL-Seq (39,40). Applying APA-seq at single-cell resolution has many interesting applications, such as the study of population of cells undergoing cell fate specification, differentiation, and during tumorigenesis.

Here, our goal has been to use APA-seq to reveal what expression dynamics may reveal regarding the selective pressures on 3′ UTR isoforms. When studying the expression of alternative 3′ UTR isoforms throughout development, we found that the global transitions of isoform usage correspond to distinct developmental periods (Figure 2). Our results are consistent with those of Lianoglou *et al.*, who studied malignant transformation across human cell lines, and revealed a change in mRNA abundance levels of genes with a single 3′ UTR isoform, and, in genes with multiple 3′ UTR isoforms, a change in 3′ UTR isoform ratios (70). A similar pattern emerged while comparing embryonic stem cells within differentiated tissues (70). Our own results add an interesting layer to this field, by dissecting the temporal components of normal *C. elegans* embryonic development, providing insight into APA dynamics across distinct developmental stages. More generally, results are consistent with the accumulating evidence indicating that the APA regulatory mechanism is highly conserved across all eukaryotes, regardless of their morphological complexity (5,6,19,21,76,77).

3′ UTR isoforms are ubiquitous and considerable effort has been addressed towards understanding the distinct

functional roles of different isoforms of the same gene (14). Here, we report two general gene classes with 3′ UTR isoforms: those whose 3′ UTR isoforms correlate in their expression profiles across time and those that do not; although the overall expression levels between these gene groups is comparable (Figure 3). Most genes (>70%) with alternatively polyadenylated isoforms exhibit a high correlation among their 3′ UTR isoforms. These highly correlated isoform genes (HCI) may be dominantly regulated at the level of overall transcription. In other words, the main factor influencing the distribution of the 3′ UTR isoforms usage is the intrinsic strength of their polyadenylation sites. Genes belonging to this class are often referred to as 'dynamic genes', which we previously showed to be enriched for developmental functions such as specification and differentiation (38,49). The HCI genes display a relative paucity of miRNA binding sites and higher concordance between total mRNA and protein levels (Figures 4 and 5).

A small fraction of genes with multiple 3′ UTR isoforms (11%) show lowly correlated isoform expression across time. We have named these LCI genes and provide evidence that this gene class is under unique regulation. LCI genes show overall less dynamic total expression profiles; however, at the level of individual 3′ UTR isoforms they are highly dynamic (Figure 3). LCI genes also exhibit several features which indicate a higher level of post-transcriptional regulation of 3′ UTR isoform usage. Post-transcriptional 3′ UTR-isoform regulation processes include miRNA mediated degradation and RNA-binding protein mediated stabilization or destabilization of mRNA molecules and control of translation efficiencies. 3′ UTRs of the LCI genes comprise significantly more miRNA binding sites and the 3′ UTR isoform ratio correlates well with the expression of a sub-group of miRNAs, many of which are known regulators of embryogenesis (Figure 4). Consistently with these findings, the correlation between total mRNA and protein abundances is lower in LCI genes relative to HCI genes indicating that the 3′ UTR usage of this group of genes is tightly regulated.

Our results reveal principles of selective pressures on alternative polyadenylation. By studying APA dynamics over developmental time, we revealed two classes of genes with alternatively polyadenylated isoforms, the HCI and LCI. The LCI show the hallmarks of strong regulation on their 3′ UTR isoforms in the form of miRNA. Thus, we revealed here that a powerful litmus test for a functional distinction among 3′ UTR isoforms during a biological process (such as embryogenesis) is their discordant expression across time. As a corollary, genes with 3′ UTR isoforms showing correlated expression may represent biological noise of no functional consequence, as is common for other processes such as alternative splicing (78). Collectively, our results characterize the regulatory principles of alternative polyadenylation and provide a context for the incorporation of specific posttranscriptional regulators such as miRNAs in the modelling of biological pathways.

DATA AVAILABILITY

APA-seq scripts are available in the GitHub repository (https://github.com/yanailab). We used our previously published *C. elegans* time-course data available at GEO (GSE50548).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Eitan Winter and Martin Feder for their assistance throughout this project. We also acknowledge assistance from the Technion Genome Center.
Author contributions: N.M., T.H. and I.Y. conceived and designed the project. N.M. led the development of the APA-seq approach. M.L., H.Z., N.M. and I.Y. analysed the 3′ UTR isoform data. H.Z. contributed the miRNA analysis and comparison with previous annotations. M.L. contributed the developmental APA dynamics, mRNA-protein correlation and RNA-binding-protein analyses. M.L., H.Z., N.M. and I.Y. led the interpretation of the data. M.L., H.Z., N.M. and I.Y. drafted the manuscript. M.L. is supported by a Humboldt-Bayer research fellowship.

FUNDING

European Research Council grant (EvoDevoPaths). Funding for open access charge: Institutional start-up funds.
Conflict of interest statement. None declared.

REFERENCES

1. Ozsolak,F., Kapranov,P., Foissac,S., Kim,S.W., Fishilevich,E., Monaghan,A.P., John,B. and Milos,P.M. (2010) Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell*, **143**, 1018–1029.
2. Jan,C.H., Friedman,R.C., Ruby,J.G. and Bartel,D.P. (2011) Formation, regulation and evolution of Caenorhabditis elegans 3′UTRs. *Nature*, **469**, 97–101.
3. Derti,A., Garrett-Engele,P., Macisaac,K.D., Stevens,R.C., Sriram,S., Chen,R., Rohl,C.A., Johnson,J.M. and Babak,T. (2012) A quantitative atlas of polyadenylation in five mammals. *Genome Res.*, **22**, 1173–1183.
4. Smibert,P., Miura,P., Westholm,J.O., Shenker,S., May,G., Duff,M.O., Zhang,D., Eads,B.D., Carlson,J., Brown,J.B. *et al.* (2012) Global patterns of tissue-specific alternative polyadenylation in Drosophila. *Cell Rep.*, **1**, 277–289.
5. Ulitsky,I., Shkumatava,A., Jan,C.H., Subtelny,A.O., Koppstein,D., Bell,G.W., Sive,H. and Bartel,D.P. (2012) Extensive alternative polyadenylation during zebrafish development. *Genome Res.*, **22**, 2054–2066.
6. Hu,W., Li,S., Park,J.Y., Boppana,S., Ni,T., Li,M., Zhu,J., Tian,B., Xie,Z. and Xiang,M. (2017) Dynamic landscape of alternative polyadenylation during retinal development. *Cell. Mol. Life Sci.*, **74**, 1721–1739.
7. Wormington,M. (1994) Unmasking the role of the 3′ UTR in the cytoplasmic polyadenylation and translational regulation of maternal mRNAs. *Bioessays*, **16**, 533–535.
8. Tadros,W., Westwood,J.T. and Lipshitz,H.D. (2007) The mother-to-child transition. *Dev. Cell*, **12**, 847–849.
9. Tadros,W. and Lipshitz,H.D. (2009) The maternal-to-zygotic transition: a play in two acts. *Development*, **136**, 3033–3042.
10. Mazumder,B., Seshadri,V. and Fox,P.L. (2003) Translational control by the 3′-UTR: tThe ends specify the means. *Trends Biochem. Sci.*, **28**, 91–98.
11. Keene,J.D. (2007) RNA regulons: coordination of post-transcriptional events. *Nat. Rev. Genet.*, **8**, 533–543.
12. Lutz,C.S. and Moreira,A. (2011) Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. *Wiley Interdiscip. Rev. RNA*, **2**, 22–31.

13. Elkon,R., Ugalde,A.P. and Agami,R. (2013) Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.*, **14**, 496–506.
14. Tian,B. and Manley,J.L. (2017) Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.*, **18**, 18–30.
15. Alt,F.W., Bothwell,A.L.M., Knapp,M., Siden,E., Mather,E., Koshland,M. and Baltimore,D. (1980) Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3′ ends. *Cell*, **20**, 293–301.
16. Rogers,J., Early,P., Carter,C., Calame,K., Bond,M., Hood,L. and Wall,R. (1980) Two mRNAs with different 3′ ends encode membrane-bound and secreted forms of immunoglobulin mu chain. *Cell*, **20**, 303–312.
17. Setzer,D.R., McGrogan,M., Nunberg,J.H. and Schimke,R.T. (1980) Size heterogeneity in the 3′ end of dihydrofolate reductase messenger RNAs in mouse cells. *Cell*, **22**, 361–370.
18. Mangone,M., Manoharan,A.P., Thierry-Mieg,D., Thierry-Mieg,J., Han,T., Mackowiak,S.D., Mis,E., Zegar,C., Gutwein,M.R., Khivansara,V. *et al.* (2010) The landscape of C. elegans 3′UTRs. *Science*, **329**, 432–435.
19. Velten,L., Anders,S., Pekowska,A., Järvelin,A.I., Huber,W., Pelechano,V. and Steinmetz,L.M. (2015) Single-cell polyadenylation site mapping reveals 3′ isoform choice variability. *Mol. Syst. Biol.*, **11**, 812.
20. Tian,B., Hu,J., Zhang,H. and Lutz,C.S. (2005) A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.*, **33**, 201–212.
21. Wang,E.T., Sandberg,R., Luo,S., Khrebtukova,I., Zhang,L., Mayr,C., Kingsmore,S.F., Schroth,G.P. and Burge,C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470–476.
22. Ji,Z., Lee,J.Y., Pan,Z., Jiang,B. and Tian,B. (2009) Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 7028–7033.
23. Blazie,S.M., Geissel,H.C., Wilky,H., Joshi,R., Newbern,J. and Mangone,M. (2017) Alternative polyadenylation directs Tissue-Specific miRNA targeting in Caenorhabditis elegans somatic tissues. *Genetics*, **206**, 757–774.
24. Blazie,S.M., Babb,C., Wilky,H., Rawls,A., Park,J.G. and Mangone,M. (2015) Comparative RNA-Seq analysis reveals pervasive tissue-specific alternative polyadenylation in Caenorhabditis elegans intestine and muscles. *BMC Biol.*, **13**, 4.
25. Khraiwesh,B. and Salehi-Ashtiani,K. (2017) Alternative poly(A) tails meet miRNA targeting in Caenorhabditis elegans. *Genetics*, **206**, 755–756.
26. Chen,W., Jia,Q., Song,Y., Fu,H., Wei,G. and Ni,T. (2017) Alternative polyadenylation: methods, findings, and impacts. *Genomics. Proteomics Bioinformatics*, **15**, 287–300.
27. Sulston,J.E. and Horvitz,H.R. (1977) Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Dev. Biol.*, **56**, 110–156.
28. Kimble,J. and Hirsh,D. (1979) The postembryonic cell lineages of the hermaphrodite and male gonads in Caenorhabditis elegans. *Dev. Biol.*, **70**, 396–417.
29. 1998) Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science*, **282**, 2012–2018.
30. McKay,S.J., Johnsen,R., Khattra,J., Asano,J., Baillie,D.L., Chan,S., Dube,N., Fang,L., Goszczynski,B., Ha,E. *et al.* (2003) Gene expression profiling of cells, tissues, and developmental stages of the nematode C. elegans. In: *Cold Spring Harbor Symposia on Quantitative Biology*. Vol. **68**, pp. 159–169.
31. Shin,H., Hirst,M., Bainbridge,M.N., Magrini,V., Mardis,E., Moerman,D.G., Marra,M.A., Baillie,D.L. and Jones,S.J.M. (2008) Transcriptome analysis for Caenorhabditis elegans based on novel expressed sequence tags. *BMC Biol.*, **6**, 30.
32. Hillier,L.W., Reinke,V., Green,P., Hirst,M., Marra,M.A. and Waterston,R.H. (2009) Massively parallel sequencing of the polyadenylated transcriptome of C. elegans. *Genome Res.*, **19**, 657–666.
33. Ramani,A.K., Nelson,A.C., Kapranov,P., Bell,I., Gingeras,T.R. and Fraser,A.G. (2009) High resolution transcriptome maps for wild-type and NMD mutant C. elegans through development. *Genome Biol.*, **10**, R101.
34. Lamm,A.T., Stadler,M.R., Zhang,H., Gent,J.I. and Fire,A.Z. (2011) Multimodal RNA-seq using single-strand, double-strand, and CircLigase-based capture yields a refined and extended description of the C. elegans transcriptome. *Genome Res.*, **21**, 265–275.
35. Grün,D., Kirchner,M., Thierfelder,N., Stoeckius,M., Selbach,M. and Rajewsky,N. (2014) Conservation of mRNA and protein expression during development of C.elegans. *Cell Rep.*, **6**, 565–577.
36. Newman-Smith,E.D. and Rothman,J.H. (1998) The maternal-to-zygotic transition in embryonic patterning of Caenorhabditis elegans. *Curr. Opin. Genet. Dev.*, **8**, 472–480.
37. Walser,C.B. and Lipshitz,H.D. (2011) Transcript clearance during the maternal-to-zygotic transition. *Curr. Opin. Genet. Dev.*, **21**, 431–443.
38. Levin,M., Hashimshony,T., Wagner,F. and Yanai,I. (2012) Developmental milestones punctuate gene expression in the Caenorhabditis embryo. *Dev. Cell*, **22**, 1101–1108.
39. Hashimshony,T., Wagner,F., Sher,N. and Yanai,I. (2012) CEL-Seq: single-cell RNA-seq by multiplexed linear amplification. *Cell Rep.*, **2**, 666–673.
40. Hashimshony,T., Senderovich,N., Avital,G., Klochendler,A., de Leeuw,Y., Anavy,L., Gennert,D., Li,S., Livak,K.J., Rozenblatt-Rosen,O. *et al.* (2016) CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.*, **17**, 77.
41. Mangone,M., Macmenamin,P., Zegar,C., Piano,F. and Gunsalus,K.C. (2008) UTRome.org: a platform for 3′UTR biology in C. elegans. *Nucleic Acids Res.*, **36**, D57–D62.
42. Graveley,B.R., Brooks,A.N., Carlson,J.W., Duff,M.O., Landolin,J.M., Yang,L., Artieri,C.G., van Baren,M.J., Boley,N., Booth,B.W. *et al.* (2011) The developmental transcriptome of Drosophila melanogaster. *Nature*, **471**, 473–479.
43. Casas-Vila,N., Bluhm,A., Sayols,S., Dinges,N., Dejung,M., Altenhein,T., Kappei,D., Altenhein,B., Roignant,J.Y. and Butter,F. (2017) The developmental proteome of Drosophila melanogaster. *Genome Res.*, **27**, 1273–1285.
44. Sanfilippo,P., Wen,J. and Lai,E.C. (2017) Landscape and evolution of tissue-specific alternative polyadenylation across Drosophila species. *Genome Biol.*, **18**, 229.
45. Zhou,X., Zhang,Y., Michal,J.J., Qu,L., Zhang,S., Wildung,M.R., Du,W., Pouchnik,D.J., Zhao,H., Xia,Y. *et al.* (2019) Alternative polyadenylation coordinates embryonic development, sexual dimorphism and longitudinal growth in Xenopus tropicalis. *Cell. Mol. Life Sci.*, **76**, 2185–2198.
46. Peshkin,L., Wühr,M., Pearl,E., Haas,W., Freeman,R.M., Gerhart,J.C., Klein,A.M., Horb,M., Gygi,S.P. and Kirschner,M.W. (2015) On the relationship of protein and mRNA dynamics in vertebrate embryonic development. *Dev. Cell*, **35**, 383–394.
47. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
48. Gruber,A.J., Schmidt,R., Gruber,A.R., Martin,G., Ghosh,S., Belmadani,M., Keller,W. and Zavolan,M. (2016) A comprehensive analysis of 3′ end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. *Genome Res.*, **26**, 1145–1159.
49. Levin,M., Anavy,L., Cole,A.G., Winter,E., Mostov,N., Khair,S., Senderovich,N., Kovalev,E., Silver,D.H., Feder,M. *et al.* (2016) The mid-developmental transition and the evolution of animal body plans. *Nature*, **531**, 637–641.
50. Zalts,H. and Yanai,I. (2017) Developmental constraints shape the evolution of the nematode mid-developmental transition. *Nat. Ecol. Evol.*, **1**, 0113.
51. Bolt,B.J., Rodgers,F.H., Shafie,M., Kersey,P.J., Berriman,M. and Howe,K.L. (2018) Using WormBase ParaSite: An Integrated Platform for Exploring Helminth Genomic Data. In: *Methods in Molecular Biology (Clifton, N.J.)*. Vol. **1757**, pp. 471–491.
52. Ambros,V. (2004) The functions of animal microRNAs. *Nature*, **431**, 350–355.
53. Bartel,D.P. (2004) MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
54. Avital,G., Starvaggi França,G. and Yanai,I. (2017) Bimodal evolutionary developmental miRNA program in animal embryogenesis. *Mol. Biol. Evol.*, **35**, 646–654.
55. Kamath,R.S., Fraser,A.G., Dong,Y., Poulin,G., Durbin,R., Gotta,M., Kanapin,A., Le Bot,N., Moreno,S., Sohrmann,M. *et al.*

(2003) Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature*, **421**, 231–237.
56. Barrett,L.W., Fletcher,S. and Wilton,S.D. (2012) Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell. Mol. Life Sci.*, **69**, 3613–3634.
57. Pichon,X., Wilson,L.A., Stoneley,M., Bastide,A., King,H.A., Somers,J. and Willis,A.E. (2012) RNA binding protein/RNA element interactions and the control of translation. *Curr. Protein Pept. Sci.*, **13**, 294–304.
58. Peterson,S.M., Thompson,J.A., Ufkin,M.L., Sathyanarayana,P., Liaw,L. and Congdon,C.B. (2014) Common features of microRNA target prediction tools. *Front. Genet.*, **5**, 23.
59. Smith,C.J., Watson,J.D., Spencer,W.C., O'Brien,T., Cha,B., Albeg,A., Treinin,M. and Miller,D.M. (2010) Time-lapse imaging and cell-specific expression profiling reveal dynamic branching and molecular determinants of a multi-dendritic nociceptor in C. elegans. *Dev. Biol.*, **345**, 18–33.
60. Lee,C-Y.S., Lu,T. and Seydoux,G. (2017) Nanos promotes epigenetic reprograming of the germline by down-regulation of the THAP transcription factor LIN-15B. *Elife*, **6**, e30201.
61. Szostak,E. and Gebauer,F. (2013) Translational control by 3′-UTR-binding proteins. *Brief. Funct. Genomics*, **12**, 58–65.
62. Zhao,W., Pollack,J.L., Blagev,D.P., Zaitlen,N., McManus,M.T. and Erle,D.J. (2014) Massively parallel functional annotation of 3′ untranslated regions. *Nat. Biotechnol.*, **32**, 387–391.
63. Berkovits,B.D. and Mayr,C. (2015) Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. *Nature*, **522**, 363–367.
64. Klein,A.M., Mazutis,L., Akartuna,I., Tallapragada,N., Veres,A., Li,V., Peshkin,L., Weitz,D.A. and Kirschner,M.W. (2015) Droplet barcoding for Single-Cell transcriptomics applied to embryonic stem cells. *Cell*, **161**, 1187–1201.
65. Picelli,S., Björklund,Å.K., Faridani,O.R., Sagasser,S., Winberg,G. and Sandberg,R. (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods*, **10**, 1096–1098.
66. Zheng,G.X.Y., Terry,J.M., Belgrader,P., Ryvkin,P., Bent,Z.W., Wilson,R., Ziraldo,S.B., Wheeler,T.D., McDermott,G.P., Zhu,J. *et al.* (2017) Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.*, **8**, 14049.
67. Gruber,A.R., Martin,G., Müller,P., Schmidt,A., Gruber,A.J., Gumienny,R., Mittal,N., Jayachandran,R., Pieters,J., Keller,W. *et al.* (2014) Global 3′ UTR shortening has a limited effect on protein abundance in proliferating T cells. *Nat. Commun.*, **5**, 5465.
68. Shepard,P.J., Choi,E.A., Lu,J., Flanagan,L.A., Hertel,K.J. and Shi,Y. (2011) Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA*, **17**, 761–772.
69. Yao,C., Biesinger,J., Wan,J., Weng,L., Xing,Y., Xie,X. and Shi,Y. (2012) Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 18773–18778.
70. Lianoglou,S., Garg,V., Yang,J.L., Leslie,C.S. and Mayr,C. (2013) Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev.*, **27**, 2380–2396.
71. Wilkening,S., Pelechano,V., Järvelin,A.I., Tekkedil,M.M., Anders,S., Benes,V. and Steinmetz,L.M. (2013) An efficient method for genome-wide polyadenylation site mapping and RNA quantification. *Nucleic Acids Res.*, **41**, e65.
72. Nam,J-W., Rissland,O.S., Koppstein,D., Abreu-Goodger,C., Jan,C.H., Agarwal,V., Yildirim,M.A., Rodriguez,A. and Bartel,D.P. (2014) Global analyses of the effect of different cellular contexts on microRNA targeting. *Mol. Cell*, **53**, 1031–1043.
73. Ye,C., Long,Y., Ji,G., Li,Q.Q. and Wu,X. (2018) APAtrap: identification and quantification of alternative polyadenylation sites from RNA-seq data. *Bioinformatics*, **34**, 1841–1849.
74. Gupta,I., Collier,P.G., Haase,B., Mahfouz,A., Joglekar,A., Floyd,T., Koopmans,F., Barres,B., Smit,A.B., Sloan,S.A. *et al.* (2018) Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat. Biotechnol.*, **36**, 1197–1202.
75. Li,W., You,B., Hoque,M., Zheng,D., Luo,W., Ji,Z., Park,J.Y., Gunderson,S.I., Kalsotra,A., Manley,J.L. *et al.* (2015) Systematic profiling of poly(A)+ transcripts modulated by core 3′ end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation. *PLoS Genet.*, **11**, e1005166.
76. Ara,T., Lopez,F., Ritchie,W., Benech,P. and Gautheret,D. (2006) Conservation of alternative polyadenylation patterns in mammalian genes. *BMC Genomics*, **7**, 189.
77. Shi,Y. (2012) Alternative polyadenylation: new insights from global analyses. *RNA*, **18**, 2105–2117.
78. Grishkevich,V. and Yanai,I. (2014) Gene length and expression level shape genomic novelties. *Genome Res.*, **24**.1497–1503.