

8. R. L. Larson, *Geology* **19**, 547 (1991).
9. R. L. Larson, E. Erba, *Paleoceanography* **14**, 663 (1999).
10. E. Erba, *Mar. Micropaleontol.* **52**, 85 (2004).
11. R. M. Leckie, T. J. Bralower, R. Cashman, *Paleoceanography* **17**, 1041 (2002).
12. E. Erba, F. Tremolada, *Paleoceanography* **19**, 1008 (2004).
13. A. P. Menegatti *et al.*, *Paleoceanography* **13**, 530 (1998).
14. H. C. Jenkyns, *Philos. Trans. R. Soc. London Ser. A* **361**, 1885 (2003).
15. S. Méhay *et al.*, *Geology* **37**, 819 (2009).
16. M. L. G. Tejada *et al.*, *Geology* **37**, 855 (2009).
17. H. Weissert, E. Erba, *J. Geol. Soc. London* **161**, 695 (2004).
18. Materials and methods are available as supporting material on Science Online.
19. A. Malinverno, E. Erba, T. Herbert, *Paleoceanography* **24**, PA2203 (2010).
20. A. Ando, K. Kaiho, H. Kawahata, T. Kakegawa, *Paleogeogr. Palaeoclimatol. Palaeoecol.* **260**, 463 (2008).
21. D. Bernoulli, H. C. Jenkyns, *Sedimentology* **56**, 149 (2009).
22. J. B. Ries, A. L. Cohen, D. C. McCorkle, *Geology* **37**, 1131 (2009).
23. M. I. Millán, H. J. Weissert, P. A. Fernández-Mendiola, J. García-Mondéjar, *Earth Planet. Sci. Lett.* **287**, 392 (2009).
24. J. C. Zachos *et al.*, *Science* **308**, 1611 (2005).
25. G. R. Dickens, M. M. Castillo, J. C. G. Walker, *Geology* **25**, 259 (1997).
26. I. Raffi, B. De Bernardi, *Mar. Micropaleontol.* **69**, 119 (2008).
27. P. Bown, P. Pearson, *Mar. Micropaleontol.* **71**, 60 (2009).
28. S. J. Gibbs, P. R. Bown, J. A. Sessa, T. J. Bralower, P. A. Wilson, *Science* **314**, 1770 (2006).
29. E. Erba *et al.*, *J. Foraminiferal Res.* **29**, 371 (1999).
30. E. Erba, *Paleoceanography* **9**, 483 (1994).
31. E.E. and C.B. were funded through MIUR (Italian Ministry of University and Research), Research Programs of National Interest (PRIN) grant 2007-2007W9B2WE 001. H.J.W. and C.E.K. were funded by the Swiss Science Foundation (project 200021-113687) and by ETH Zurich. E.E. and C.B. investigated calcareous nannofossil assemblages, H.J.W. and C.E.K. analyzed C and O isotopes. All authors contributed to the synthesis and paleoceanographic model. Samples from DSDP site 463 were supplied by the Integrated Ocean Drilling Program.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/329/5990/428/DC1  
Materials and Methods  
Figs. S1 and S2  
Tables S1 to S3  
References

1 March 2010; accepted 18 June 2010  
10.1126/science.1188886

## The Landscape of *C. elegans* 3'UTRs

Marco Mangone,<sup>1\*</sup> Arun Prasad Manoharan,<sup>2\*</sup> Danielle Thierry-Mieg,<sup>3\*</sup> Jean Thierry-Mieg,<sup>3\*</sup> Ting Han,<sup>2\*</sup> Sebastian D. Mackowiak,<sup>4</sup> Emily Mis,<sup>1</sup> Charles Zegar,<sup>1</sup> Michelle R. Gutwein,<sup>1</sup> Vishal Khivansara,<sup>2</sup> Oliver Attie,<sup>1</sup> Kevin Chen,<sup>1,4</sup> Kourosh Salehi-Ashtiani,<sup>5,6</sup> Marc Vidal,<sup>5,6</sup> Timothy T. Harkins,<sup>7</sup> Pascal Bouffard,<sup>8</sup> Yutaka Suzuki,<sup>9</sup> Sumio Sugano,<sup>9</sup> Yuji Kohara,<sup>10</sup> Nikolaus Rajewsky,<sup>4</sup> Fabio Piano,<sup>1,11†</sup> Kristin C. Gunsalus,<sup>1,11†</sup> John K. Kim<sup>2†</sup>

Three-prime untranslated regions (3'UTRs) of metazoan messenger RNAs (mRNAs) contain numerous regulatory elements, yet remain largely uncharacterized. Using polyA capture, 3' rapid amplification of complementary DNA (cDNA) ends, full-length cDNAs, and RNA-seq, we defined ~26,000 distinct 3'UTRs in *Caenorhabditis elegans* for ~85% of the 18,328 experimentally supported protein-coding genes and revised ~40% of gene models. Alternative 3'UTR isoforms are frequent, often differentially expressed during development. Average 3'UTR length decreases with animal age. Surprisingly, no polyadenylation signal (PAS) was detected for 13% of polyadenylation sites, predominantly among shorter alternative isoforms. Trans-spliced (versus non-trans-spliced) mRNAs possess longer 3'UTRs and frequently contain no PAS or variant PAS. We identified conserved 3'UTR motifs, isoform-specific predicted microRNA target sites, and polyadenylation of most histone genes. Our data reveal a rich complexity of 3'UTRs, both genome-wide and throughout development.

The 3' untranslated regions (3'UTRs) of mRNAs contain cis-acting sequences that interact with RNA-binding proteins and/or small noncoding RNAs [such as micro RNAs (miRNAs)] to influence mRNA stability, localization, and translational efficiency (1–3). The differential processing of mRNA 3' ends has evident roles in development, metabolism, and disease (4, 5). Despite these critical roles, genome-wide characterization of 3'UTRs lags far behind that of coding sequences (CDSs). Even in the well-annotated genome of *Caenorhabditis elegans*, nearly half (~47%) of the 20,191 genes annotated in WormBase (release WS190) (6, 7) lack an annotated 3'UTR, and only ~1180 (~5%) are annotated with alternative 3'UTR isoforms (fig. S1, A and B).

We have taken a multifaceted, empirical approach to defining the 3'UTR landscape in *C. elegans* (figs. S2 to S5 and tables S1 to S4) (8). We prepared developmentally staged cDNA libraries composed of mostly full-length clones spanning from 5' capped first base to polyadenylated (polyA) tail, and we annotated 16,659 polyA addition sites in 11,180 genes by manually curating ~300,000 Sanger capillary sequence traces in National Center for Biotechnology Information (NCBI) AceView (9). We developed a method to capture the 3' ends

of polyadenylated transcripts genome-wide by deep sampling and generated a comprehensive developmental profile comprising more than 2.5 million sequence reads from Roche/454 (fig. S2 to S5 and tables S1 to S4). We cloned 3' rapid amplification of cDNA ends (RACE) products directly targeting 3'UTRs for 7105 CDSs (6741 genes) in both the Promoterome (10) and ORFeome (11) collections, and we recovered one or more sequenced isoforms for 85% of the targets (figs. S2 and S5 and tables S1 to S4) (8, 12). Finally, we remapped and annotated polyA addition sites in published RNA-seq data (13, 14).

All data sets were mapped, cross-validated, consolidated, and filtered to eliminate obvious experimental artifacts, including internal priming on A-rich stretches (Fig. 1A) (8). These data sets are not yet saturated: Whereas for most genes (11,516 or 73%), at least one 3'UTR isoform is supported by two or more experimental approaches, 47% of transcripts are observed by only one method (in part due to limitations specific to each protocol) (Fig. 1 and tables S3 and S4) (8). The resulting 130,090 distinct polyA sites, identified at single-nucleotide resolution and supported by more than 3 million independent polyA tags, were clustered into 26,967 representative polyA sites. Due to biological varia-

tion, 86% of tags occur within 4 nucleotides of representative sites, although individual polyA tags may spread over ~20 nucleotides (fig. S6).

Linking polyA sites to their parent genes proved to be a challenge, as many previous gene models were incomplete or incompatible with our new data. Using all available empirical evidence, we reannotated in AceView the *C. elegans* gene models (9). Of the 15,683 protein-coding genes with both polyA sites and cDNA support, 57% confirm the structure of WormBase WS190 gene models. The remainder encode different proteins, usually representing different cDNA-supported splice patterns: ~25% share the same stop codon, ~12% use a different stop (hundreds of those correspond to fusions or splits of earlier gene models), and ~6% are not yet annotated in WormBase (supporting data sets S1 and S2).

This integrated collection, herein called the 3'UTRome (fig. S1 and data set S2), provides evidence supporting 3'UTR structures for ~74% of all *C. elegans* protein-coding genes in WormBase WS190, including previously unannotated isoforms for ~7397 genes (fig. S1, A to D). The length distribution of 3'UTRs parallels that in WormBase (fig. S1D), with a mean of 211 nu-

<sup>1</sup>Center for Genomics and Systems Biology, Department of Biology, New York University, 1009 Silver Center, New York, NY 10003, USA. <sup>2</sup>Life Sciences Institute, Department of Human Genetics, University of Michigan, 210 Washtenaw Avenue, Ann Arbor, MI 48109-2216, USA. <sup>3</sup>National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA. <sup>4</sup>Max Delbrück Centrum für Molekulare Medizin, Robe Rössle-Strasse 10, Berlin-Buch, 13092 Berlin, Germany. <sup>5</sup>Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute (DFCI), Boston, MA 02115, USA. <sup>6</sup>Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. <sup>7</sup>Roche Applied Science, Indianapolis, IN 46250, USA. <sup>8</sup>454 Life Sciences, Roche Applied Science, Branford, CT 06405, USA. <sup>9</sup>Department of Medical Genome Sciences, Graduate School of Frontier Sciences, University of Tokyo, 4-6-1 Shirokanedai, Minatoku, Tokyo 108-8639, Japan. <sup>10</sup>Genome Biology Lab Center for Genetic Resource Information, National Institute of Genetics, Mishima 411-8540, Japan. <sup>11</sup>New York University, Abu Dhabi, United Arab Emirates.

\*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: fp1@nyu.edu (F.P.); kcg1@nyu.edu (K.C.G.); jnkim@umich.edu (J.K.K.)

cleotides (nt) (median = 140 nt). The 3'UTRome matches 61% of WormBase 3'UTRs within  $\pm 10$  nt (6714 polyA ends for 6563 genes) and contains thousands of longer or shorter isoforms (fig. S1A). We identified 6177 polyA ends for 4466 genes with no previous 3'UTR annotation and discovered 1490 polyA ends for 1031 genes not yet represented in WormBase (fig. S1A and data sets S1 to S3).

We annotate more than one 3'UTR isoform for 43% of 3'UTRome genes (figs. S1 and S7). Of these, a majority (65%) reflects alternative 3'-end formation at distinct locations in the same terminal exon for proteins using the same stop; the remainder use distinct stops in the same last exon or distinct last exons. Very rarely (79 examples), an intron within the 3'UTR is excised or retained (fig. S8), potentially affecting functional sequence content elements (fig. S8C). Indeed, putative binding sites for miRNAs (this study) or ALG-1 (15) were identified in the variable regions of some of these transcripts. About 2% of genes possess five or more 3'UTR isoforms (Fig. 1A and figs. S1B and S7).

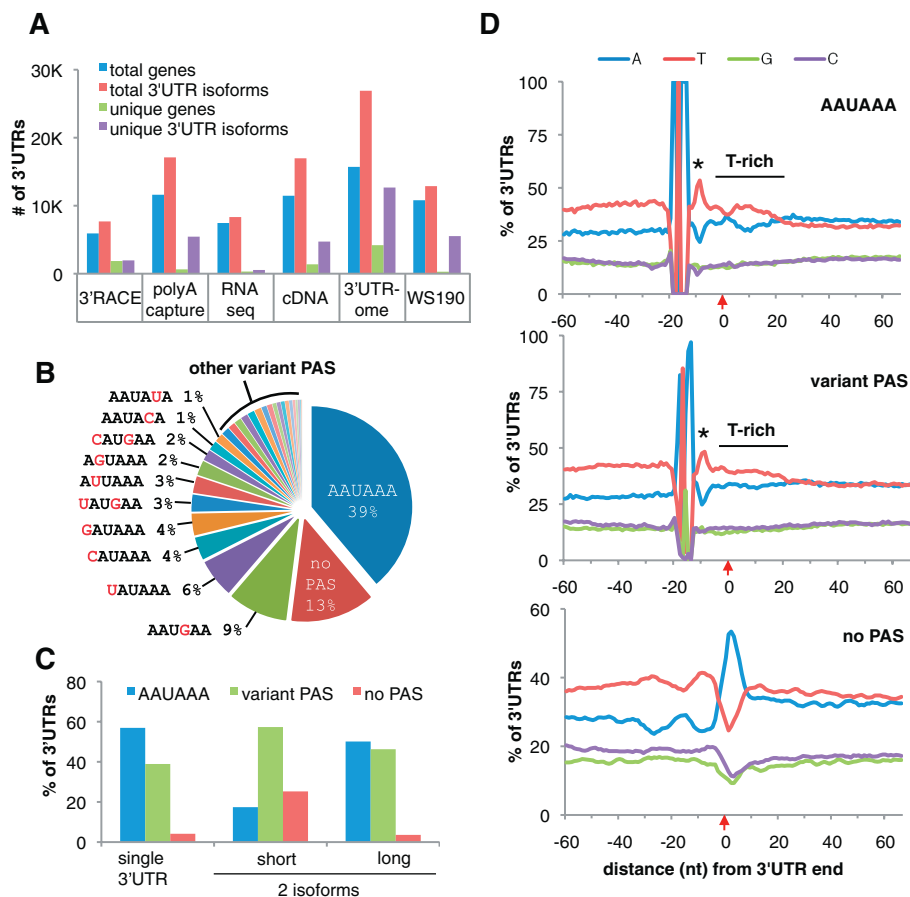
To identify putative cis-acting sequences that may play a role in 3'-end formation, we scanned the 50 nt upstream of the cleavage and polyA addition sites for all possible 5- to 10-mers and assigned the most likely polyadenylation signal (PAS) motif to each 3'UTR using an iterative procedure based on enrichment and centering of the *k*-mers. The canonical PAS motif AAUAAA (seen in 39% of 3' ends) and many variants differing by 1 to 2 nt are detected, with distributions all peaking 19 nt upstream of the polyA site (figs. S9, S10, and table S5) (8). The canonical signal predominates in genes with unique 3'UTRs (57%). However, many high-quality 3'UTRs (3658) lack a detectable PAS motif altogether (Fig. 1, B and C). All PAS variants are embedded within a T-rich region that spikes 5 nt downstream of the PAS motif and extends about 20 nt beyond the cleavage site (Fig. 1D). 3'UTRs with no PAS tend to be T-rich throughout, except for a very A-rich eight-nucleotide region just after the cleavage site (Fig. 1D). Thus, a functional PAS motif with strict sequence specificity appears dispensable for 3'-end formation in *C. elegans*.

Among genes with alternative 3'UTRs, successive polyA sites show a marked asymmetry: The longest isoform prefers a PAS, whereas shorter isoforms more often show no PAS (Fig. 1C and fig. S11). The distance between alternative polyA sites peaks at  $\sim 40$  nt, with resonances at  $\sim 80$  and  $\sim 140$  nt (fig. S11A). This regularity suggests that a physical constraint (possibly queuing transcription complexes) could contribute to cleavage and polyA addition at some upstream sites, which may, therefore, depend less on instructive cues from signal sequences.

Because many *C. elegans* genes undergo splicing of a splice leader (SL) to the 5' end of a nascent transcript (16), we asked whether any properties of transcript 5' and 3' ends correlate (Fig. 2, A and B). About 15% of *C. elegans* genes belong to transcriptional units called operons, each containing two to eight genes that can be cotranscribed, cleaved into separate transcripts, polyadenylated, and trans-spliced with specific leaders (Fig. 2, A and B). The first gene in an operon is trans-spliced only to SL1; downstream genes are usually trans-spliced to 1 of 11 other SLs (SL2 to SL12), although we observed that two-thirds of these genes occasionally become trans-spliced to SL1. The processing of adjacent operon transcript ends (cleavage, polyA addition to the upstream transcript, and SL addition to the downstream transcript) is coupled mechanically by machinery resembling the cis-splicing apparatus (17). Comparing 3'UTRs within operons, we observe that the "first" (SL1-spliced), "middle" (any gene between first and last), and "last" genes progressively decrease in average length (from 266 to 213 nt), number of 3'UTR isoforms per gene (from 2.64 to 2.51), and frequency of 3'UTRs with no PAS (from 23 to 18% in  $\sim 1400$  sites) (Fig. 2B).

However, only a small fraction (13%) of the 7026 mainly SL1-spliced genes clearly belongs to an operon, and these genes differ notably from non-operon SL1-spliced genes in their usage of the canonical AAUAAA hexamer (22% of 1409 sites versus 32% of 10,879 sites, respectively). Furthermore, we observed the canonical PAS motif much more frequently in non-trans-spliced than in SL-containing transcripts (43% of 5131 sites versus 30% of 14,873 sites) (Fig. 2A). Whereas "standard" non-trans-spliced genes have  $\sim 30\%$  more 3'UTR isoforms per gene than "isolated" ones having no neighbor within 2 kb (2.4 versus 1.7), these non-trans-spliced genes are more similar to each other than to trans-spliced genes, because they have shorter and fewer 3'UTR isoforms and higher canonical PAS usage. Thus, trans-splicing within operons appears to enhance (directly or indirectly) the activity of noncanonical PAS sequences upstream, and trans-splicing at the 5' end correlates with distinct properties at the 3' end of the same transcript, independent of 5'-end processing downstream.

Unexpectedly, the 3'UTRome reveals polyadenylated transcripts for nearly all histone genes (fig. S12 and table S6). The major class of replication-dependent histones (H2a, H2b, H3, and H4) is not thought to be polyadenylated in



**Fig. 1.** The 3'UTRome and 3'UTR PAS. **(A)** The number of genes and isoforms detected in, or specific to, each data set and cumulative totals in WS190 and 3'UTRome annotations. **(B)** PAS motif frequencies: AAUAAA (39%), variant PAS (1 to 9%), and no PAS (13%). **(C)** PAS usage in genes with one or two (short and long) 3'UTR isoforms. **(D)** Nucleotide distribution spanning  $\pm 60$  nt around the polyA addition site, in 3'UTRs with: AAUAAA (top), 10 most common variant PAS (middle), and no PAS (bottom). Alignments, centered at  $-19$  nt, show a T-spike at 5 nt downstream of PAS (asterisk), polyA addition site (red arrow), and T-rich region downstream of cleavage site. The A-rich peak downstream of "no PAS" is not enriched for AAAAAA, suggesting an A-rich motif at that location rather than artifactual A-rich ends.

metazoans; instead, their 3' ends form a stem-loop structure that is recognized and cleaved several nucleotides downstream by U7 small nuclear ribonucleoprotein and factors such as stem-loop binding protein (18, 19). *C. elegans* has 61 cDNA-supported histone genes (9) that all harbor conserved sequences with 3' stem-loop potential; however, they also contain conserved PAS elements downstream of the hairpin sequence (20). Because *C. elegans* histone transcripts have also been shown to terminate in the typical stem-loop structure and to be depleted in successive rounds of polyA selection (20), we were surprised to recover polyadenylated transcripts for 57 histone genes in multiple, independent data sets (fig. S12 and table S6). This finding suggests that, at least in *C. elegans* (and perhaps also in higher metazoans), the usual route for histone mRNA 3'-end processing may include initial cleavage and polyA addition at conserved PAS sites, followed by further processing to remove sequences downstream of the stem-loop.

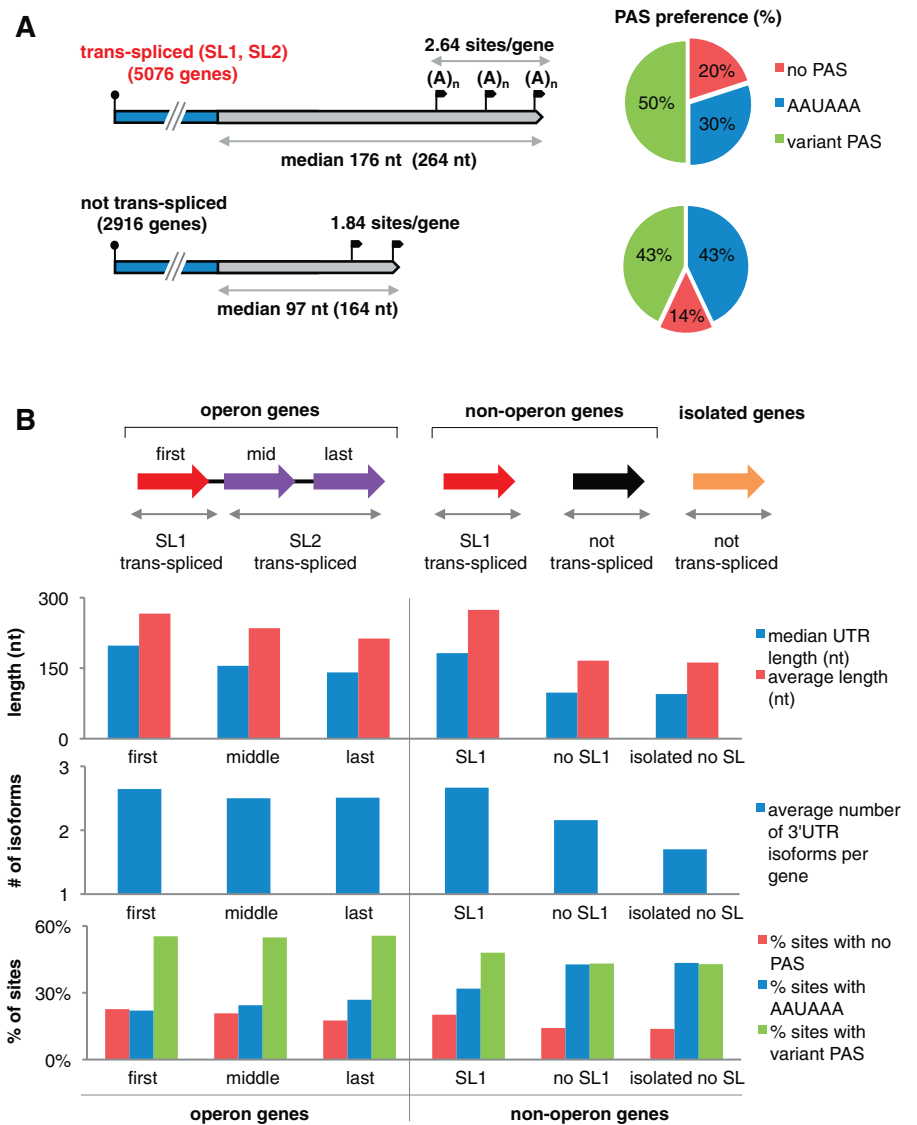
We searched 3'UTRs for conserved sequence motifs and other potential functional elements. We updated our atlas of predicted conserved miRNA targets for the 3'UTRome, using the PicTar algorithm with new 3- and 5-way multispecies alignments (Fig. 3, fig. S13, and table S7). Roughly half of the newly predicted sites match our previous predictions (21), but many sites are gained or lost (fig. S13A and table S7). These differences reflect improvements in both 3'UTR annotations and multispecies alignments, which increase the accuracy of conserved-seed site identification and signal-to-noise ratios (8). More than 3000 PAS motifs are positionally conserved among *Caenorhabditis* species, including within alternative 3'UTRs (fig. S13B). Thus, maintenance of multiple specific 3' termini may be functionally important for some genes. Thousands of unexplained conserved sequence blocks of varying lengths within 3'UTRs (Fig. 3B and table S7) may represent previously unrecognized functional elements that await further characterization. In vivo Argonaute (ALG-1) binding sites (15) overlap significantly with predicted miRNA target sites but not with other conserved blocks (table S7), indicating that the latter are, overall, not directly related to microRNA function (8). For 1876 convergently transcribed neighboring genes, overlapping 3' regions could pair as double-stranded RNA if coexpressed, potentially triggering endogenous small interfering RNA production (22) that could down-regulate cognate mRNAs (fig. S14 and data set S4).

We examined alternative 3'UTR isoforms in different developmental stages (Fig. 4) and found a downward trend in average length and number of 3'UTRs per gene from the embryonic through the adult stage (Fig. 4, A and B). Among genes expressed in more than one developmental stage, embryos display the largest proportion of stage-specific 3'UTR isoforms, and these tend toward longer isoforms (Fig. 4, B and C, tables S8 and S9, and data set S5). Some genes switch 3'UTR length coincident with developmental transitions, most notably from embryo to L1, L1 to dauer entry, dauer exit to

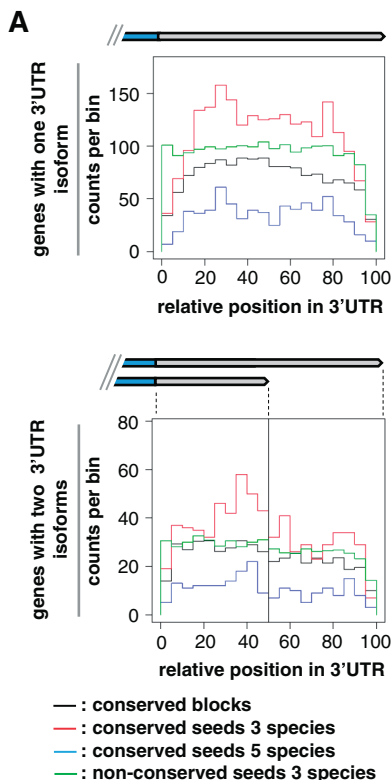
L4, and in adult hermaphrodites versus males (Fig. 4D, table S9, and data sets S5 and S6). Thus, 3' UTR-mediated gene regulation may be widespread in the *C. elegans* embryo, and differential expression of alternative isoforms may represent a mechanism to engage or bypass 3'UTR-mediated regulatory controls in specific developmental contexts (23, 24).

The 3'UTRome compendium evidences support for multiple mechanisms of transcript 3'-end formation in *C. elegans*, including standard PAS-directed 3'-end formation from a large collection of PAS variants, regularly spaced "shadow" polyA addition sites devoid of recognizable signals, and both operon-dependent and -independent correlations between features at the 5' and 3' ends of

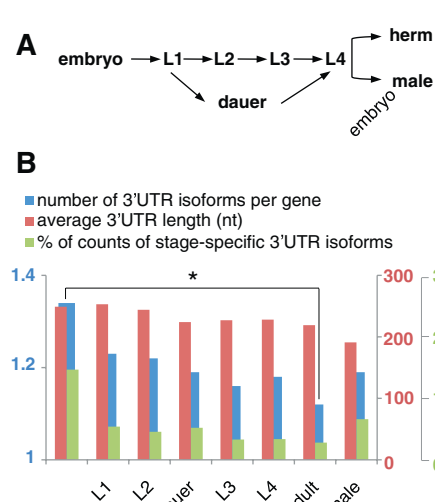
the same or of consecutive transcripts that are consistent with the possibility that trans-splicing and 3'-end processing within a gene could occur by functionally linked mechanisms. We characterize thousands of previously unknown and alternative 3'UTR isoforms throughout development, define a comprehensive catalog of PAS elements, discover a surprising number of polyadenylated transcripts with no discernable PAS, and definitively document polyadenylation of histone transcripts. We also identify conserved sequence elements in 3'UTRs that may interact with trans-acting factors such as miRNAs and RNA-binding proteins, some of which occur within variable regions of alternative 3'UTRs. A collection of cloned 3'UTRs



**Fig. 2.** 3'UTRs in operons and trans-spliced versus non-trans-spliced mRNAs. **(A)** Trans-spliced (top) and non-trans-spliced (bottom) mRNAs: 3'UTR median (and average) lengths, number of 3'UTR isoforms per gene (polyA sites, black flags), and PAS preference (pie charts: % 3'UTRs with AAUAAA, variant PAS, and no PAS). **(B)** (Top) Schematic of operon (left,  $n = 574$  operons), non-operon (center,  $n = 4348$  genes), and isolated (right,  $n = 2098$ ) genes. Initial operon genes (red) are SL1-trans-spliced; downstream genes (purple) usually acquire one of the other SLs (SL2 to SL12). Non-operon genes are either SL1-trans-spliced (red,  $n = 3530$ ) or not trans-spliced (black,  $n = 818$ ). Isolated genes (having no neighbors within 2 kb) are not trans-spliced (orange,  $n = 2098$ ). **(Bottom)** 3'UTR lengths, number of isoforms, and PAS sites for operon and non-operon genes.



**Fig. 3.** Conserved sequence elements in 3'UTRs. **(A)** Histogram distributions of conserved sequence blocks (black, counts shown at 1/5th scale), conserved miRNA seeds in three (red; *C. elegans*, *C. remanei*, *C. briggsae*) and five (blue; *C. elegans*, *C. remanei*, *C. briggsae*, *C. brenneri*, *C. japonica*) species, and nonconserved miRNA seeds (green, 1/25th scale) along the normalized length of 3'UTRs, in genes with one isoform (top) or exactly two isoforms (bottom). For genes with one isoform, the length scale is 100%; for two isoforms, 0 to 50% represents the short-isoform span, and 51 to 100% indicates the span exclusive to the long isoform. Counts were binned by fraction of total length and, thus, varied in absolute length. **(B)** Length distribution (up to 20 nt) of conserved sequence blocks in 3'UTRs (excluding miRNA target and PAS sites), in three (blue;  $n = 16,204$  conserved blocks) and five (red;  $n = 4758$ ) species. See also table S7.



**Fig. 4.** 3'UTRs during development. **(A)** *C. elegans* developmental transitions: embryogenesis, four larval stages, and adults. In unfavorable environments, L1 larvae arrest in dauer stage and can re-enter the life cycle as L4 larvae. herm, hermaphrodites. **(B)** The number of 3'UTR isoforms per gene decreases significantly during development (blue) (\* $p \sim 0.004$ , permutation test). The average length of 3'UTRs decreases during development (red). Adult males have shorter average 3'UTRs than hermaphrodites. Embryos show more stage-specific 3'UTR isoforms for genes expressed during multiple developmental stages (green) (see table S8). **(C)** Proportion of genes showing stage-specific expression of alternative 3'UTR isoforms (see table S9). Embryos and dauers favor longer 3'UTR isoforms. **(D)** Differential 3'UTR-isoform expression during development (*ubc-18* shown; see data sets S5 and S6 for details). The bar chart illustrates the relative abundance of short versus long 3'UTR isoforms for *ubc-18* in each stage (sum per stage = 100%, left y axis). The line graph shows the relative abundance across all stages (sum per gene across all stages = 100%, right y axis). Green bars highlight differences in 3'UTR isoform usage in the embryo-to-L1 transition and between adult hermaphrodite and male stages. Green arrows indicate dauer entry and exit transitions.

for several thousand *C. elegans* genes is available to the research community for high-throughput downstream analyses and in vivo studies (table S10 and data set S7) (8).

**References and Notes**

1. C. H. de Moor, H. Meijer, S. Lissenden, *Semin. Cell Dev. Biol.* **16**, 49 (2005).
2. M. Wickens, D. S. Bernstein, J. Kimble, R. Parker, *Trends Genet.* **18**, 150 (2002).
3. D. P. Bartel, *Cell* **136**, 215 (2009).
4. L. He *et al.*, *Nature* **435**, 828 (2005).
5. S. Chatterjee, J. K. Pal, *Biol. Cell* **101**, 251 (2009).
6. L. Stein, P. Sternberg, R. Durbin, J. Thierry-Mieg, J. Spieth, *Nucleic Acids Res.* **29**, 82 (2001).
7. T. W. Harris *et al.*, *Nucleic Acids Res.* **38** (Database issue), D463 (2010).
8. See supporting online material for details.
9. D. Thierry-Mieg, J. Thierry-Mieg, *Genome Biol.* **7** (suppl. 1), S12.1 (2006).
10. D. Dupuy *et al.*, *Genome Res.* **14**, 2169 (2004).
11. J. Reboul *et al.*, *Nat. Genet.* **34**, 35 (2003).
12. M. Mangone, P. Macmenamin, C. Zegar, F. Piano, K. C. Gunsalus, *Nucleic Acids Res.* **36** (Database issue), D57 (2008).
13. L. W. Hillier *et al.*, *Genome Res.* **19**, 657 (2009).
14. H. Shin *et al.*, *BMC Biol.* **6**, 30 (2008).
15. D. G. Zisoulis *et al.*, *Nat. Struct. Mol. Biol.* **17**, 173 (2010).
16. T. Blumenthal *et al.*, *Nature* **417**, 851 (2002).
17. Y. Liu, T. Huang, M. MacMorris, T. Blumenthal, *RNA* **7**, 176 (2001).
18. Z. F. Wang, M. L. Whitfield, T. C. Ingledue, Z. Dominski, W. F. Marzluff, *Genes Dev.* **10**, 3028 (1996).
19. W. F. Marzluff, E. J. Wagner, R. J. Duronio, *Nat. Rev. Genet.* **9**, 843 (2008).
20. R. Keall, S. Whitelaw, J. Pettitt, B. Müller, *BMC Mol. Biol.* **8**, 51 (2007).
21. S. Lall *et al.*, *Curr. Biol.* **16**, 460 (2006).
22. K. Okamura, S. Balla, R. Martin, N. Liu, E. C. Lai, *Nat. Struct. Mol. Biol.* **15**, 581 (2008).
23. E. Lund, M. Liu, R. S. Hartley, M. D. Sheets, J. E. Dahlberg, *RNA* **15**, 2351 (2009).
24. A. J. Giraldez *et al.*, *Science* **312**, 75 (2006); published online 16 February 2006 (10.1126/science.1122689).
25. This work was supported in part by grants from NIH (U01-HG004276) to F.P., K.C.G., J.K.K., and N.R.; NIH grant (R00HG004515) to K.C.; Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan to Y.K., S.S., and Y.S.; NIH (R01GM088565), Muscular Dystrophy Association and the Pew Charitable Trusts to J.K.K.; a gift from the Ellison Foundation to M.V. and Institute Sponsored Research funds from the DFCI Strategic Initiative in support of the CCSB; the Helmholtz-Alliance on Systems Biology (Max Delbrück Centrum Systems Biology Network) to S.D.M.; and the Intramural Research Program of NIH, National Library of Medicine to J.T.-M. and D.T.-M. We thank J. V. Moran, T. Blumenthal, A. Billi, D. Mecnas, and B. Bargmann for discussions; T. Shin'i and Exelixis for *C. elegans* cDNA traces; and, for technical assistance, T. Nawy and B. Brown (statistical analysis); R. Sachidanandam, R. Lyons, and S. Genik (deep sequencing); P. MacMenamin and D. Schaub (3'UTRome database); M. Morris (data submission); and L. Huang (stage analysis). 3'UTRome data sets are available from NCBI Trace Archive, dbEST, Sequence Read Archive, Gene Expression Omnibus, and modENCODE (8). See supporting online materials and methods for details. Annotations are displayed at NCBI AceView (www.aceview.org) (9) and www.UTRome.org (12).

**Supporting Online Material**

www.sciencemag.org/cgi/content/full/science.1191244/DC1  
 Materials and Methods  
 Figs. S1 to S14  
 Tables S1 to S10  
 References  
 Data sets S1 to S7

21 April 2010; accepted 24 May 2010  
 Published online 3 June 2010;  
 10.1126/science.1191244

Include this information when citing this paper.



### The Landscape of *C. elegans* 3'UTRs

Marco Mangone, Arun Prasad Manoharan, Danielle Thierry-Mieg, Jean Thierry-Mieg, Ting Han, Sebastian D. Mackowiak, Emily Mis, Charles Zegar, Michelle R. Gutwein, Vishal Khivansara, Oliver Attie, Kevin Chen, Kourosh Salehi-Ashtiani, Marc Vidal, Timothy T. Harkins, Pascal Bouffard, Yutaka Suzuki, Sumio Sugano, Yuji Kohara, Nikolaus Rajewsky, Fabio Piano, Kristin C. Gunsalus and John K. Kim (June 3, 2010)

*Science* **329** (5990), 432-435. [doi: 10.1126/science.1191244]  
originally published online June 3, 2010

Editor's Summary

#### A New Look at Old Data

Although the full genome of *Caenorhabditis elegans* has been available for over 10 years, only a portion of the full-length messenger RNAs (mRNAs) and their expression through development has been analyzed. By combining multiple methods, involving a careful reannotation of existing data sets, sequencing of poly-A captured RNAs from multiple developmental stages, as well as cloning and deep sequencing of a subset of genes, **Mangone *et al.*** (p. 432, published online 3 June) defined ~26,000 distinct mRNA three-prime untranslated regions (3'UTRs) for ~85% of the protein-coding genes. Most *C. elegans* genes have more than one 3'UTR, arising through different mechanisms including trans-splicing-coupled cleavage and polyadenylation. The work highlights the importance of the 3'UTRs in the regulation of mRNA stability and translation.

---

This copy is for your personal, non-commercial use only.

---

- Article Tools** Visit the online version of this article to access the personalization and article tools:  
<http://science.sciencemag.org/content/329/5990/432>
- Permissions** Obtain information about reproducing this article:  
<http://www.sciencemag.org/about/permissions.dtl>

*Science* (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2016 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.