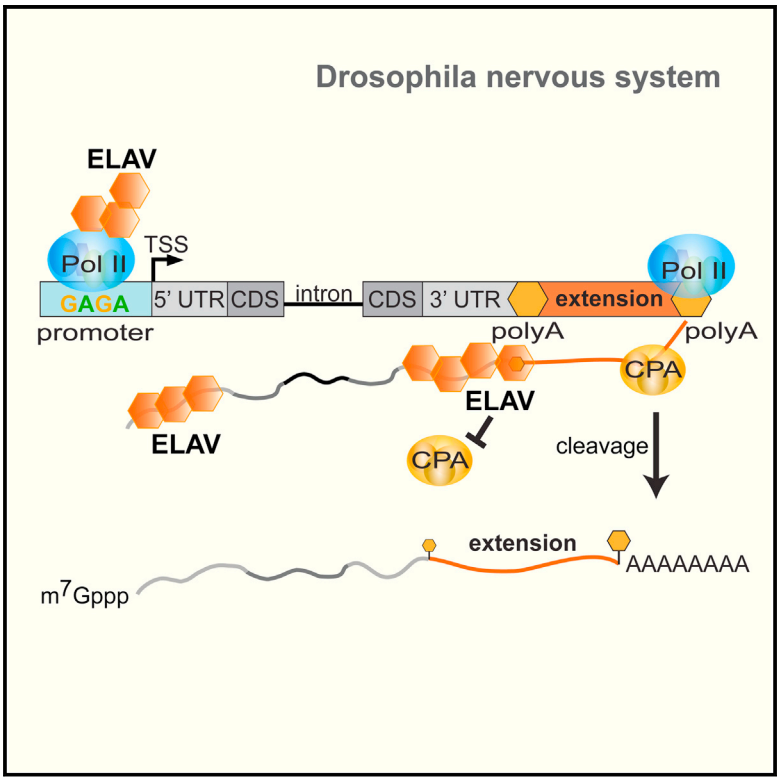
Short Article

Molecular Cell

ELAV Links Paused Pol II to Alternative Polyadenylation in the *Drosophila* Nervous System

Graphical Abstract



Authors

Katarzyna Oktaba, Wei Zhang, ..., Michael Levine, Valérie Hilgers

Correspondence

mlevine@berkeley.edu (M.L.), valerie.hilgers@berkeley.edu (V.H.)

In Brief

Oktaba and Zhang et al. provide evidence for a regulatory link between promoter-proximal pausing and alternative polyadenylation. Promoters of genes that undergo neural-specific 3′ UTR extension contain paused RNA Pol II, GAGA, and ELAV.

Highlights

- Heterologous promoters block ELAV-mediated 3′ UTR extension in the nervous system
- Extension-associated promoters can induce 3′ UTR extension in ectopic tissues
- Promoters of extended genes contain paused Pol II and GAGA
- ELAV binds to the promoter of extended genes

Accession Numbers

GSE63323

Oktaba et al., 2015, Molecular Cell *57*, 341–348
January 22, 2015
http://dx.doi.org/10.1016/j.molcel.2014.11.024

CellPress

Molecular Cell
Short Article

ELAV Links Paused Pol II to Alternative Polyadenylation in the *Drosophila* Nervous System

Katarzyna Oktaba,[1,2] Wei Zhang,[1,2] Thea Sabrina Lotz,[1] David Jayhyun Jun,[1] Sandra Beatrice Lemke,[1,3] Samuel Pak Ng,[1] Emilia Esposito,[1] Michael Levine,[1,*] and Valérie Hilgers[1,*]
[1]Division of Genetics, Genomics and Development, Department of Molecular and Cell Biology, Center for Integrative Genomics, University of California, Berkeley, Berkeley, CA 94720-3200, USA
[2]Co-first author
[3]Present address: Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany
*Correspondence: mlevine@berkeley.edu (M.L.), valerie.hilgers@berkeley.edu (V.H.)
http://dx.doi.org/10.1016/j.molcel.2014.11.024

SUMMARY

Alternative polyadenylation (APA) has been implicated in a variety of developmental and disease processes. A particularly dramatic form of APA occurs in the developing nervous system of flies and mammals, whereby various developmental genes undergo coordinate 3′ UTR extension. In *Drosophila*, the RNA-binding protein ELAV inhibits RNA processing at proximal polyadenylation sites, thereby fostering the formation of exceptionally long 3′ UTRs. Here, we present evidence that paused Pol II promotes recruitment of ELAV to extended genes. Replacing promoters of extended genes with heterologous promoters blocks normal 3′ extension in the nervous system, while extension-associated promoters can induce 3′ extension in ectopic tissues expressing ELAV. Computational analyses suggest that promoter regions of extended genes tend to contain paused Pol II and associated *cis*-regulatory elements such as GAGA. ChIP-seq assays identify ELAV in the promoter regions of extended genes. Our study provides evidence for a regulatory link between promoter-proximal pausing and APA.

INTRODUCTION

Nascent transcripts undergo 3′ cleavage and polyadenylation (CPA) prior to transcription termination to produce mature mRNAs. The C-terminal domain (CTD) of the large subunit of RNA Polymerase II (Pol II) serves as an interaction platform for multiple factors that control transcription initiation, elongation, and termination (Hsin and Manley, 2012). CPA factors have been detected in promoter regions (Glover-Cutter et al., 2008), where they interact with general transcription factors (Dantonel et al., 1997), transcriptional activators (Calvo and Manley, 2001; Nagaike et al., 2011), and the Pol II CTD (McCracken et al., 1997). Functional interactions between transcriptional initiation and termination have been documented (Andersen et al., 2013), for example, impaired 3′ processing can diminish initiation rates at yeast and human promoters (Mapendano et al., 2010; Zhang et al., 2012).

Transcriptome-wide studies have revealed that most genes contain multiple polyadenylation (poly(A)) signals and are subject to alternative polyadenylation (APA) (Brown et al., 2014; Elkon et al., 2013; Pelechano et al., 2013; Shi, 2012; Tian et al., 2005; Wang et al., 2008). The most common form of APA, tandem 3′ UTR APA, generates different mRNA isoforms possessing identical protein-coding sequences, but distinct 3′ UTRs. APA-mediated alterations of 3′ UTRs have been implicated in a variety of processes, including animal development and human disease. For example, global 3′ UTR shortening accompanies cell proliferation (Elkon et al., 2012; Ji and Tian, 2009) and can cause oncogenic transformation in cultured mammalian cells (Mayr and Bartel, 2009). Abnormal APA has been linked to oculopharyngeal muscular dystrophy (Jenal et al., 2012).

A particularly dramatic example of tissue-specific APA is seen in the developing nervous system of flies and vertebrates, whereby hundreds of genes exhibit 3′ UTR extension. Neural-specific 3′ UTR extensions have been documented in *Drosophila* (Hilgers et al., 2011; Smibert et al., 2012), zebrafish (Ulitsky et al., 2012), mouse, and human (Miura et al., 2013; Zhang et al., 2005), and seem to be a conserved feature of animal neurogenesis. The extended 3′ UTR sequences, which can reach tens of kilobases (kb) in length, are thought to confer post transcriptional regulation underlying specific neuronal functions, such as axonal transport.

In *Drosophila*, the nuclear RNA-binding protein embryonic lethal abnormal visual system (ELAV) was shown to be a key regulator of 3′ UTR extension. ELAV is expressed in the nuclei of neurons. It inhibits CPA by binding in the vicinity of proximal poly(A) sites of nascent transcripts, thereby promoting Pol II read-through and 3′ extension. Ectopic expression of ELAV was shown to be sufficient to induce ectopic extension of endogenous genes in nonneural tissues (Hilgers et al., 2012). Studies using cultured cells suggest that ELAV homologs perform similar functions in mammals (Mansfield and Keene, 2012). We hereafter refer to genes with extended 3′ UTRs in the nervous system as "extended genes."

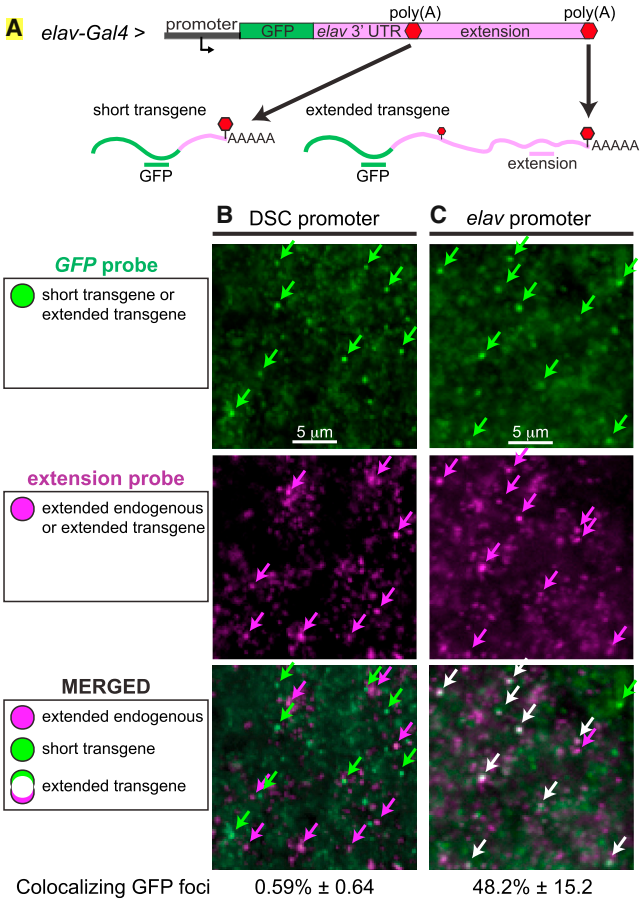


Figure 1. Native Promoters Are Required for Expression of 3′ Extensions

(A) *elav-Gal4* drives expression of a *GFP* transgene in the nervous system. There are two different promoter regions that were used, DSCP, or the native *elav* promoter. The *GFP* coding sequence was placed upstream of the entire extended 7.2 kb *elav* 3′ UTR. CPA at the proximal poly(A) produces the short 3′ UTR form of the mRNA, whereas CPA at the distal-most poly(A) produces the fully extended transcript. RNA probes directed against different regions of the transcripts were used to detect mRNAs.

(B and C) Double fluorescent in situ hybridization assays using probes indicated in (A). Single confocal sections of a portion of the developing CNS in stage 13 embryos. Note that the extension probe detects not only the transgene, but also the endogenous *elav* transcript, which is expressed in the nervous system. Colocalization of the *GFP* and extension probes indicates expression of extended transcripts from the transgene. (B) The reporter transgene carrying the DSCP does not exhibit colocalization of *GFP* and extension probes. Extension signals (magenta arrows) do not colocalize with the green *GFP* signal, indicating that they correspond to endogenous *elav* mRNAs. (C) Replacing the DSCP with the native *elav* promoter region induces 3′ extension of the *GFP* transgene. There is extensive colocalization of the *GFP* (green arrows) and extension probes (magenta arrows), indicating expression of extended 3′ UTR sequences from the transgene (white arrows in merged image). The percentages of GFP foci that colocalized with extension foci are indicated. Numbers represent mean ± SD of six embryos for each promoter. See also Figures S1 and S2.

Here, we show that ELAV-mediated 3′ UTR extension is dependent on transcription initiation. Promoters of extended genes generate 3′ UTR extension from reporter transgenes in the *Drosophila* nervous system. These promoters can also induce 3′ extension in nonneural tissues upon ectopic expression of ELAV. Computational analyses reveal that promoters of extended genes typically contain paused Pol II and are enriched in "pausing elements" such as the GAGA motif (Li and Gilmour, 2013). Moreover, ELAV ChIP-seq assays suggest that ELAV associates with the promoter regions of extended genes, but is present at significantly lower levels at nonextended genes. We propose that ELAV is recruited to the promoter regions of extended genes via paused Pol II and inhibits CPA at proximal poly(A) sites during transcription elongation.

RESULTS AND DISCUSSION

The Native Promoter Is Necessary for 3′ UTR Extension

ELAV is an RNA-binding protein that has been shown to bind to U-rich regions in target mRNAs, including *neuroglian* (Lisbin et al., 2001) and *erect wings* (Soller and White, 2003). Recently, the Hox gene *Ultrabithorax* (*Ubx*) was shown to be bound by ELAV through similar elements to regulate alternative splicing, but ELAV was not found to bind to predicted binding sites in the *Ubx* 3′ UTR (Rogulja-Ortmann et al., 2014). Similarly, we also failed to identify specific ELAV recognition sequences within extended 3′ UTRs. In the present study, we investigate how ELAV is selectively recruited to appropriate targets during neurogenesis.

We examined the activities of synthetic reporter genes in transgenic embryos to determine whether extended 3′ UTRs are sufficient for the selective recruitment of ELAV in vivo. Transgenes contain the *Drosophila* synthetic core promoter (DSCP, see Pfeiffer et al., 2008) attached to a *GFP* coding sequence followed by the entire extended 3′ UTR of *elav*, one of the targets of ELAV (Figure 1A). If *elav* 3′ UTR sequences are sufficient to recruit ELAV, then this transgene should produce mRNAs containing extended 3′ UTRs.

Expression of 3′ UTR sequences was monitored via double labeling assays with *GFP* coding sequences to distinguish transgene mRNAs from endogenous *elav* transcripts (Figure 1, schematics). Expression of the transgene was confirmed by colocalization of *GFP* with a probe directed against the short 3′ UTR (Figure S1A available online). However, we did not observe colocalization of *GFP* with extended sequences, indicating that mRNAs produced from the transgene lack 3′ extensions (Figure 1B). The only signals containing 3′ extensions corresponded to endogenous *elav* mRNAs (Figure S1B).

Additional experiments were done to determine why the transgene fails to produce extended transcripts. We excluded the possibility that the *GFP* coding sequence somehow inhibits expression of extended sequences by creating *GFP* transgenes lacking proximal poly(A) signals (Figures S1C and S1D). Such constructs no longer depended on ELAV for 3′ extension and were found to produce mRNAs containing extended 3′ UTR sequences when expressed in ectopic tissues lacking ELAV (Figures S1C and S1D).

To test whether promoter sequences play a role in ELAV recruitment, we swapped the DSCP with a 333 base pairs (bp) genomic DNA fragment encompassing the native *elav* promoter region, consisting of 92 bp upstream and 241 bp downstream of

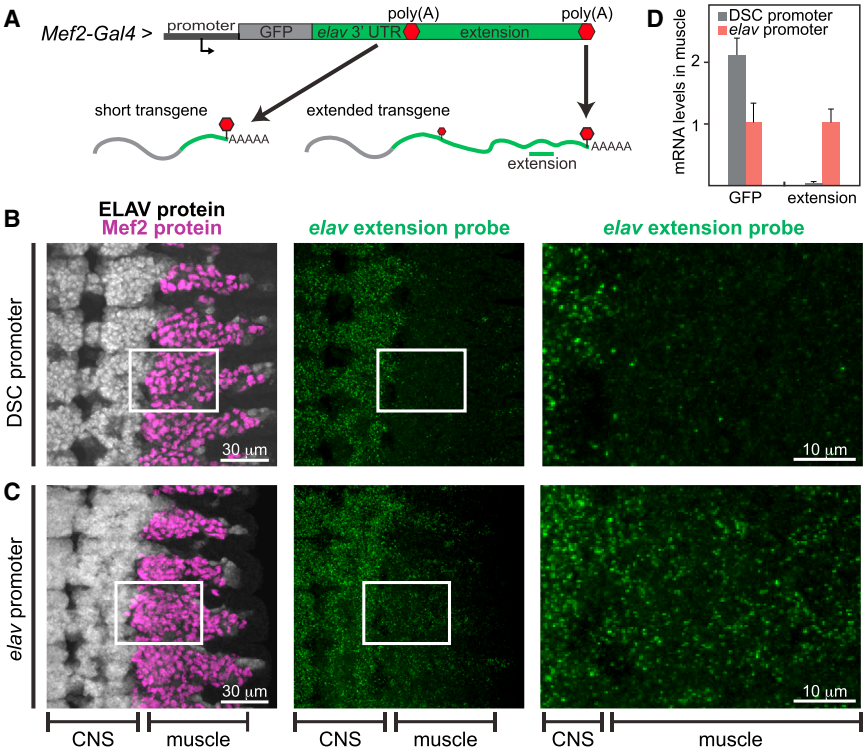


Figure 2. The Native *elav* Promoter Mediates 3′ Extension in Muscle

(A) *Mef2-Gal4* drives expression of a *GFP* transgene in muscle cells. The promoter used for expression was either the DSCP or the native *elav* promoter. The *GFP* coding sequence was placed upstream of the entire extended 7.2 kb *elav* 3′ UTR. CPA at the proximal poly(A) produces the short 3′ UTR form of the mRNA, whereas CPA at the distal-most poly(A) produces the fully extended transcript. An RNA probe directed against the *elav* extension was used to detect the extended transcript.

(B and C) Left panels show projections of consecutive confocal sections of stage 13 embryos stained with antibodies against ELAV (white, in the nervous system) and Mef2 (magenta, in muscle). Ventral views, anterior is up. Middle panels, hybridization signals with the *elav* extension probe (green). Signal in the CNS corresponds to the endogenous *elav* mRNA. Panels on the right show enlarged views of the boxed regions in the left and middle panels. Background staining in muscle tissue is observed with the DSCP transgene (B, right), indicating little or no expression of the extended 3′ UTR. In contrast, there is significant expression of extended transcripts from the transgene containing the *elav* promoter (C, right).

(D) mRNA quantification by qPCR using primer combinations detecting all transgene mRNAs (GFP) or specific to the extended transcript (extension). RNA was extracted from dissected muscle tissue in first instar larvae expressing the transgene depicted in (A), carrying either the DSCP or the *elav* promoter. Levels were normalized to rp49 RNA. Both promoters foster robust transgene expression as indicated by GFP levels, but expression of extension sequences is only detected with the *elav* promoter. Error bars represent mean ± SD of six samples for each promoter.

See also Figure S3.

the (TSS) (Yao and White, 1994). Strikingly, we observed colocalization of *GFP* and extension sequences (Figure 1C), indicating expression of the *elav* 3′ UTR extension, as seen for the endogenous locus.

To confirm that 3′ extension depends on native promoter regions of extended genes, we also tested a construct bearing the fully extended *brat* 3′ UTR downstream of *GFP*, using three different promoters: the DSCP, the native promoter producing the short form of *brat*, and the native promoter producing the extended form of *brat* (Figure S2A, A'). Only the *brat* promoter associated with endogenous extension mediated expression of transgenic transcripts containing 3′ UTR extensions (Figures S2B–S2D). These observations suggest that the promoter regions of extended genes are essential for the ELAV-mediated expression of 3′ UTR extensions.

Native Promoters Mediate 3′ Extension in Ectopic Tissues

The preceding results suggest that promoter sequences are important for the synthesis of 3′ extensions in the developing nervous system. We further explored their importance by examining nonneural tissues. Ectopic ELAV can drive 3′ UTR extension in ectopic tissues from endogenous loci (Hilgers et al., 2012). We sought to determine whether ectopic ELAV could also induce ectopic 3′ extensions from transgenic DNAs.

We expressed both the *GFP-elav* transgene and ELAV protein in muscle cells using a *Mef2-Gal4* driver (Figure 2A). In this context, mRNA expression from the reporter is easily distinguished from endogenous *elav* expression, which occurs only in the nervous system. The DSCP fails to generate 3′ UTR extensions (Figure 2B, muscle), and only endogenous *elav* transcripts in the CNS were detected (Figure 2B, CNS). In contrast, the *GFP-elav* transgene containing the native *elav* promoter produced transcripts with extended 3′ UTRs in muscle tissue (Figure 2C, muscle). Quantification of transgene expression in dissected muscle tissue using quantitative PCR (qPCR) shows that both promoters drive robust transgene expression (GFP signal), but only the native promoter drives expression of extension sequences (Figure 2D). Similarly, the second *brat* promoter (see above), but not the DSCP, was also able to drive expression of an extended *brat* 3′ UTR in muscle cells (Figures S3A–S3C).

We also tested whether the promoter sequence from one extended gene could promote extension of the 3′ UTR of another such gene. Indeed, a *GFP* transgene containing the *elav* promoter and *brat* extended 3′ UTR exhibited ELAV-mediated APA (Figures S3D and S3E). These observations suggest a link between transcription initiation and ELAV-mediated APA.

Promoters of Extended Genes Contain GAGA and Paused Pol II

To determine whether the promoter regions of extended genes share common sequence motifs, we examined 252 neural-specific transcripts produced by 219 different genes exhibiting 3′ UTR extensions (Smibert et al., 2012). The most significantly

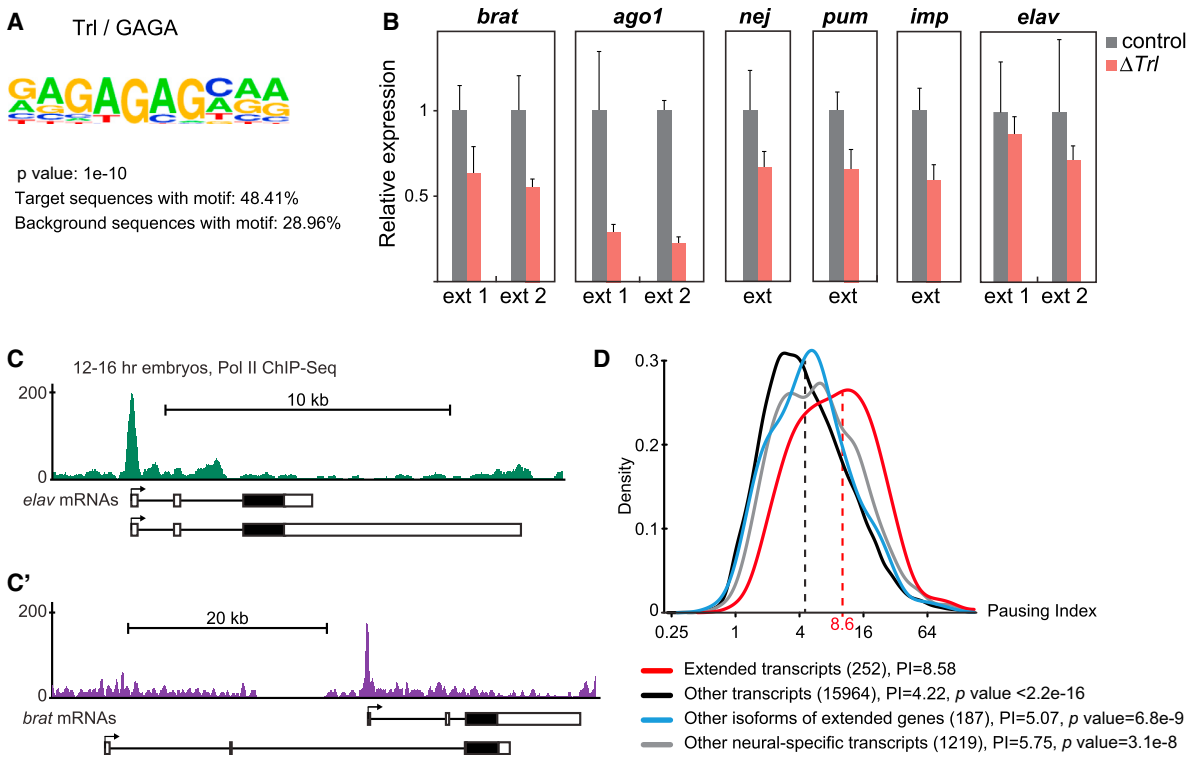


Figure 3. Extended Genes Contain the GAGA Motif and Paused Pol II

(A) A motif search among 252 neural-specific transcripts exhibiting 3′ UTR extensions yielded the GAGA motif as the most significantly enriched motif compared to background sequences (all other annotated gene promoters).

(B) Quantification of indicated transcripts by qPCR using primer combinations specific to the partially extended (ext 1) or fully extended (ext 2) 3′ UTR forms of each gene. RNA was extracted from brains of *yw* (control) or *Trl* mutant (Δ*Trl*) third instar larvae. Extension levels were normalized to coding regions of each gene to reflect levels relative to the short isoforms. For each primer pair, expression in control larvae was set to the value 1. In *Trl* mutants, 3′ UTR extension of each of the six analyzed genes is significantly reduced (p values < 0.01, unpaired Student's t test) compared to control larvae. Error bars represent mean ± SD of three samples for each genotype.

(C and C') Normalized Pol II ChIP-seq reads at the *elav* (C) and *brat* (C') loci in 12–16 hr embryos (Nègre et al., 2011). Short and extended isoforms are represented below the tracks. Arrows denote the start site and directionality of transcription. Pol II peaks indicate promoter-proximal pausing at the *elav* locus (C) and at the promoter expressing the extended form of *brat*, but not the short form (C').

(D) PI distribution and median PI values of the promoters of the indicated groups of transcripts in whole embryos. The numbers in parentheses denote the number of transcripts in each group. Promoters of extended transcripts are significantly more paused than promoters of any control group. Wilcoxon rank-sum test p values were calculated by comparing the PI of extended transcripts with each group of controls.

See also Figure S4.

enriched motif is the GAGA element (p value = 1×10^{-10}), which occurs in nearly half of all extended genes (Figures 3A and S4A). To investigate the functional significance of the GAGA element in promoters of extended genes, we tested whether 3′ UTR extension is diminished in animals lacking the GAGA-binding protein, Trithorax-like (Trl). For all six genes we examined, the ratio between extension sequences and coding sequences was reduced between 15% and 75% in *Trl* mutant flies (Figure 3B). These observations suggest that the GAGA motifs in the promoters of extended genes are important for proper 3′ UTR extension.

The GAGA element is a motif commonly found in the promoter regions of genes containing paused Pol II. Paused Pol II is a pervasive feature of gene regulation in metazoan development, and at least 10%–30% of all genes in *Drosophila* contain paused Pol II. It is thought that paused promoters are poised for rapid activation and thereby exhibit synchronous induction in the different cells of a tissue (e.g., see, Boettiger and Levine, 2009). Another function of promoter pausing might be to ensure proper recruitment of essential factors for RNA elongation and processing (Adelman and Lis, 2012).

We found that most extended genes contain paused Pol II, based on whole genome Pol II ChIP-seq assays (Nègre et al., 2011). Some extended genes express both short and long isoforms from the same promoter (for example *elav*, Figure 3C), while others (e.g., *brat*) employ different promoters for the different isoforms. In the latter case, only the promoter driving the extended isoform contains paused Pol II (Figure 3C').

To determine whether paused Pol II might be associated with the formation of 3′ UTR extensions, we compared the overall Pol II pausing index (PI) of extended genes and various control genes. We found that extended transcripts are derived from

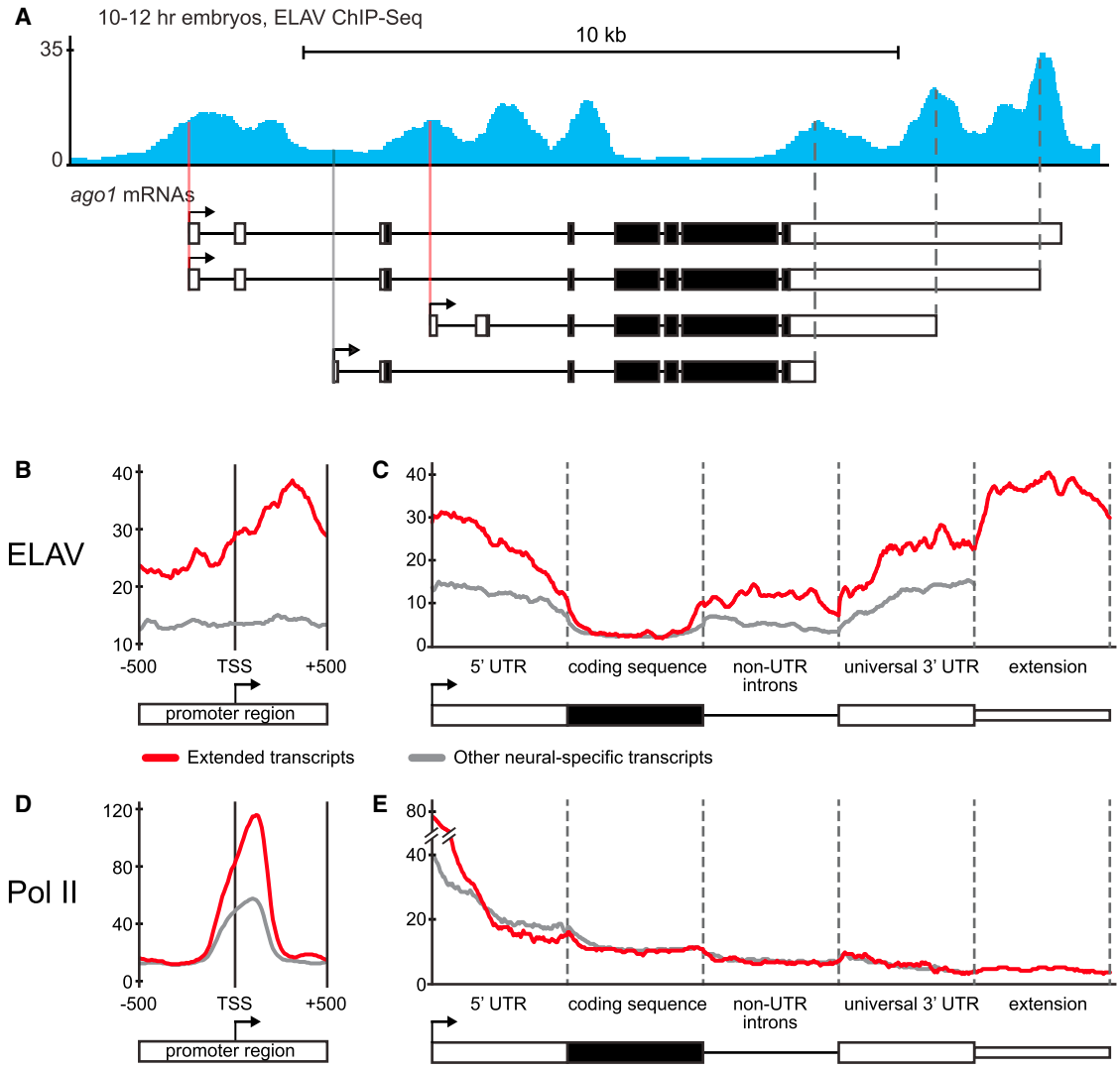


Figure 4. ELAV Binds to Promoter Regions of Extended Genes

(A) Normalized ELAV ChIP-seq reads at the *ago1* locus in 10–12 hr embryos. Shown is a merged track of duplicate experiments. ELAV is found at the promoters of the extended *ago1* isoforms (red lines), but not the shortest 3′ UTR form (gray line). There are peaks of ELAV binding at each proximal poly(A) site (dotted lines) where it suppresses CPA. The coding region is notably devoid of ELAV binding.

(B) Meta-gene plots of ELAV ChIP-seq data sets at the promoter region (±500 bp relative to the start site) in 10–12 hr embryos. Promoter regions of extended transcripts show significantly higher ELAV binding than other neural-specific transcripts (Wilcoxon test p value = 1.3×10^{-9}).

(C) Meta-gene analysis of ELAV binding across the entire transcription unit in 10–12 hr embryos. ELAV binding is higher in extended transcripts at the 5′ UTR, introns, and the 3′ UTR as compared with other neural-specific transcripts. In all genes, ELAV binding is excluded from the coding sequence.

(D and E) Meta-gene analysis of Pol II binding at the promoter region (D) or across the entire transcription unit (E) in 12–16 hr embryos. Promoter regions of extended transcripts show significantly higher Pol II binding than other neural-specific transcripts. Other regions do not differ in their Pol II binding profile between the two groups.

See also Tables S1 and S2 and Figure S4.

significantly more paused (PI = 8.58) promoters than any of the control groups, including neural-specific (but nonextended) genes (PI = 5.75) (Figures 3D and 4D). Thus, there is a clear association between Pol II pausing and 3′ UTR extension, which transcends the general pausing seen for neural-specific gene expression. Extended transcripts are also strongly paused in muscle cells (Figure S4B; PI = 7.97), where they are not actively transcribed and where ELAV is not expressed (Gaertner et al., 2012). Thus, Pol II pausing at extended genes occurs independently of ELAV.

ELAV Binds to the Promoter Regions of Extended Genes

The preceding analyses raise the possibility that ELAV is selectively recruited to the promoter regions of extended genes. To test this hypothesis, we performed ChIP-seq assays using anti-ELAV antibodies. ELAV is an RNA-binding protein that

directly binds and inhibits proximal poly(A) elements of target transcripts (Hilgers et al., 2012). We therefore reasoned that it should be possible to identify the genome-wide distribution of ELAV by crosslinking ELAV/RNA complexes to associated DNA templates. ELAV ChIP-seq assays were conducted with nuclei obtained from 6–8 hr and 10–12 hr embryos. These stages were selected based on our previous observations regarding the timing of 3′ extensions in the nervous system (Hilgers et al., 2011).

We identified 6,879 genomic regions bound by ELAV in 6–8 hr embryos (Table S1) and 8,076 regions in 10–12 hr embryos (Table S2). There is a striking enrichment of ELAV in the promoter regions of extended genes. For example, *argonaute1* (*ago1*) produces multiple APA isoforms driven from three different promoters. The two promoters that produce extended transcripts display ELAV peaks, whereas the promoter that expresses the short (ubiquitous) isoform does not (Figures 4A and S4C, filled lines). High levels of ELAV are also found at 3′ poly(A) sites (Figures 4A and S4C, dotted lines), consistent with previous RNA immunoprecipitation assays (Hilgers et al., 2012).

We combined the ChIP-seq data into a "meta-gene" plot that provides simple visualization of key sites of ELAV binding (Figures 4B, 4C, S4D, and S4E, see Experimental Procedures). There is a significant enrichment of ELAV at the promoter regions of extended genes as compared with neural-specific nonextended genes (Figure 4B, Wilcoxon test p value = 1.3×10^{-9}). A distinct ELAV peak is seen near the TSS, although ELAV binding continually increases across the 5′ UTR and peaks at ~300 bp downstream of the start site.

ELAV not only binds to promoter regions, but also to 3′ UTRs and introns of extended genes. ELAV is strikingly depleted from coding sequences. As expected, binding markedly increases in the vicinity of proximal poly(A) sites and remains high across extended regions where there are additional poly(A) elements (Figures 4C and S4E).

We also performed a meta-gene analysis of previously published Pol II ChIP-seq data (Nègre et al., 2011). Pol II binding is highly enriched in the promoter regions of extended genes, which is consistent with our earlier evidence that such genes tend to contain paused Pol II (Figures 4D and S4F). The Pol II binding profile did not otherwise differ from nonextended neural-specific genes (Figures 4E and S4G). It is possible that ELAV binds to both nascent transcripts and associated DNA templates, since ELAV is usually detected at distal poly(A) sites of extended genes prior to full transcriptional extension (e.g., Figure S4C).

We have presented evidence that paused Pol II fosters selective recruitment of ELAV and coordinates expression of extended 3′ UTR sequences during neurogenesis. The basis for selective recruitment of ELAV is a bit of a mystery since it has been shown to interact with broadly distributed low-complexity RNA sequences (e.g., U-rich). Increased interaction between paused promoters and termination regions might help promote 3′ extension, for example, by bringing ELAV to the promoter via gene looping (Henriques et al., 2012; O'Sullivan et al., 2004; Tan-Wong et al., 2012). The observed association of ELAV with the paused promoter regions of extended genes provides a foundation for selectivity and also strengthens the link between transcription initiation and 3′ cleavage (Hsin and Manley, 2012). It is improbable that paused Pol II is sufficient for recruitment of ELAV, since not all paused genes exhibit APA. It is therefore likely that additional sequence elements, for example, in extended 3′ UTRs, are essential for recruitment. ELAV proteins are highly conserved, and it is easy to imagine that the regulation of 3′ extension in the vertebrate CNS depends on selective promoter recruitment as seen in *Drosophila*.

EXPERIMENTAL PROCEDURES

Plasmids and Fly Strains

Flies were cultured on standard medium, and crosses were performed at 25°C. *Trl* mutants had the genotype Trl^{62}/Trl^{67}. *Mef2-Gal4* and *elav-Gal4* strains were obtained from the Bloomington Stock Center. Trl^{62} and Trl^{67} flies were provided by Paul Schedl. *GFP* reporter plasmids were constructed by inserting the *eGFP* coding sequence BglII/NotI into pBID-upstream activation sequence UASc (Wang et al., 2012). Native promoter sequences (300–350 bp surrounding the TSS) were amplified from fly genomic DNA and cloned into pBID-UASc-eGFP SacI/BglII, thus removing the DSC promoter and maintaining the UAS repeats. Extended 3′ UTR sequences were amplified from genomic DNA and cloned into the modified pBID-UASc-eGFP NotI/XbaI. Extension sequences lacking the short 3′ UTR including the proximal poly(A) were cloned in the same way. In those constructs, additional proximal poly(A) signals present in the extension sequences were mutated from AATAAA into AACAAA. Constructs were injected, and transgenic flies were generated using targeted integration. Primer sequences are available in the Supplemental Experimental Procedures.

In Situ Hybridization and Immunocytochemistry

Embryos were collected, fixed, and hybridized with riboprobes according to standard protocols. Detection of RNA probes was carried out with anti-digoxigenin and anti-biotin primary antibodies (Roche) and fluorescent secondary antibodies (Molecular Probes). Rat anti-ELAV-7E8A10 was obtained from the Developmental Studies Hybridoma Bank, and rabbit anti-DMef2 was a gift from Bruce Paterson. Confocal imaging was performed on a Zeiss LSM 700 microscope. Colocalizing GFP foci were manually counted in confocal images. Approximately 150 GFP foci were assessed per embryo for at least six embryos per experiment.

RNA Quantification

Total RNA was extracted from dissected first instar larval muscle tissue (Figure 2D) or dissected third instar larval brains (Figure 3B) using TRIzol (Invitrogen). DNase treatment and reverse transcription used the QuantiTect Reverse Transcription Kit (QIAGEN). qPCR was performed on a 1:20 dilution of the samples and monitored in a Viia7 real-time PCR system using SYBR Green reagents (Applied Biosystems). Primer sequences are available on request.

ChIP-Seq

ChIP-seq from *Drosophila* embryos was performed essentially as described in Oktaba et al. (2008), with modifications as described in the Supplemental Experimental Procedures. ChIP-seq libraries were constructed with the NEBNext ChIP-seq Library Prep Master Mix Set for Illumina (NEB) using NEBNext Multiplex Oligos for Illumina (NEB). ChIP and input DNA libraries were single-end sequenced with 50 bp reads using an Illumina HiSeq2000 instrument by the Functional Genomics Laboratory at the University of California at Berkeley. Data were processed as described in the Supplemental Experimental Procedures.

Computational Analysis of Promoters of Extended Transcripts

Known nervous system specific extended transcripts and control groups of transcripts were defined and filtered as described in the Supplemental Experimental Procedures. Enriched sequence motifs in the promoters of 3′ extended genes were identified using HOMER software (Heinz et al., 2010).

A region ±200 bp relative to the TSS was searched, and all other annotated gene promoters were used as the background set. PIs were determined as described in the Supplemental Experimental Procedures.

Meta-Gene Analysis

ELAV ChIP-seq data from two biological replicates in 10–12 hr embryos and Pol II ChIP-seq data in 12–16 hr embryos (Nègre et al., 2011) were used for this analysis. Enriched ELAV and Pol II binding regions were identified as described in the Supplemental Experimental Procedures. All the reads outside the ELAV or Pol II binding regions were filtered out, respectively. Each of the following six gene body regions was divided into 100 windows: (1) promoter (TSS ±500 bp), (2) 5′ UTR, (3) coding sequence, (4) non-UTR introns, (5) universal 3′ UTR, and (6) 3′ UTR extension. The filtered reads were mapped to these regions and reads per kb per million mapped reads were calculated for each window. Meta-gene plots were smoothened by using the moving average of seven windows.

Analysis of ELAV Binding at Promoter Regions

ELAV binding at promoter regions was calculated as the ELAV enrichment over background (input DNA), averaged between two biological replicates, within ±500 bp relative to the TSS. ELAV binding at 252 promoters of 3′ UTR extended transcripts was compared to 1,219 promoters of nonextended neural-specific transcripts using the Wilcoxon rank-sum test.

ACCESSION NUMBERS

ChIP-seq data have been deposited in the Gene Expression Omnibus (GEO) database under GSE63323.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2014.11.024.

AUTHOR CONTRIBUTIONS

V.H. designed, performed, and analyzed experiments and wrote the paper. K.O. performed and analyzed experiments. W.Z. analyzed data. T.S.L., D.J.J., S.B.L., S.P.N., and E.E. performed experiments. M.L. designed experiments and wrote the paper.

ACKNOWLEDGMENTS

We thank James Manley for critical reading of the manuscript. V.H. is supported by a fellowship from the German Research Foundation (DFG HI 1552/3-1). K.O. is supported by a fellowship from the European Molecular Biology Organization (EMBO ALTF 492-2011). This study was funded by a grant from the NIH (GM34431). This work used the Vincent J. Coates Genomics Sequencing Laboratory at the University of California at Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303.

Received: July 23, 2014
Revised: November 3, 2014
Accepted: November 19, 2014
Published: December 24, 2014

REFERENCES

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat. Rev. Genet. *13*, 720–731.

Andersen, P.K., Jensen, T.H., and Lykke-Andersen, S. (2013). Making ends meet: coordination between RNA 3′-end processing and transcription initiation. Wiley Interdiscip. Rev. RNA *4*, 233–246.

Boettiger, A.N., and Levine, M. (2009). Synchronous and stochastic patterns of gene activation in the Drosophila embryo. Science *325*, 471–473.

Brown, J.B., Boley, N., Eisman, R., May, G.E., Stoiber, M.H., Duff, M.O., Booth, B.W., Wen, J., Park, S., Suzuki, A.M., et al. (2014). Diversity and dynamics of the Drosophila transcriptome. Nature *512*, 393–399.

Calvo, O., and Manley, J.L. (2001). Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation, and termination. Mol. Cell *7*, 1013–1023.

Dantonel, J.C., Murthy, K.G., Manley, J.L., and Tora, L. (1997). Transcription factor TFIID recruits factor CPSF for formation of 3′ end of mRNA. Nature *389*, 399–402.

Elkon, R., Drost, J., van Haaften, G., Jenal, M., Schrier, M., Oude Vrielink, J.A., and Agami, R. (2012). E2F mediates enhanced alternative polyadenylation in proliferation. Genome Biol. *13*, R59.

Elkon, R., Ugalde, A.P., and Agami, R. (2013). Alternative cleavage and polyadenylation: extent, regulation and function. Nat. Rev. Genet. *14*, 496–506.

Gaertner, B., Johnston, J., Chen, K., Wallaschek, N., Paulson, A., Garruss, A.S., Gaudenz, K., De Kumar, B., Krumlauf, R., and Zeitlinger, J. (2012). Poised RNA polymerase II changes over developmental time and prepares genes for future expression. Cell Reports *2*, 1670–1683.

Glover-Cutter, K., Kim, S., Espinosa, J., and Bentley, D.L. (2008). RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. Nat. Struct. Mol. Biol. *15*, 71–78.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors primecis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Henriques, T., Ji, Z., Tan-Wong, S.M., Carmo, A.M., Tian, B., Proudfoot, N.J., and Moreira, A. (2012). Transcription termination between polo and snap, two closely spaced tandem genes of D. melanogaster. Transcription *3*, 198–212.

Hilgers, V., Perry, M.W., Hendrix, D., Stark, A., Levine, M., and Haley, B. (2011). Neural-specific elongation of 3′ UTRs during Drosophila development. Proc. Natl. Acad. Sci. USA *108*, 15864–15869.

Hilgers, V., Lemke, S.B., and Levine, M. (2012). ELAV mediates 3′ UTR extension in the Drosophila nervous system. Genes Dev. *26*, 2259–2264.

Hsin, J.P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev. *26*, 2119–2137.

Jenal, M., Elkon, R., Loayza-Puch, F., van Haaften, G., Kühn, U., Menzies, F.M., Oude Vrielink, J.A., Bos, A.J., Drost, J., Rooijers, K., et al. (2012). The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. Cell *149*, 538–553.

Ji, Z., and Tian, B. (2009). Reprogramming of 3′ untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. PLoS ONE *4*, e8419.

Li, J., and Gilmour, D.S. (2013). Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. EMBO J. *32*, 1829–1841.

Lisbin, M.J., Qiu, J., and White, K. (2001). The neuron-specific RNA-binding protein ELAV regulates neuroglian alternative splicing in neurons and binds directly to its pre-mRNA. Genes Dev. *15*, 2546–2561.

Mansfield, K.D., and Keene, J.D. (2012). Neuron-specific ELAV/Hu proteins suppress HuR mRNA during neuronal differentiation by alternative polyadenylation. Nucleic Acids Res. *40*, 2734–2746.

Mapendano, C.K., Lykke-Andersen, S., Kjems, J., Bertrand, E., and Jensen, T.H. (2010). Crosstalk between mRNA 3′ end processing and transcription initiation. Mol. Cell *40*, 410–422.

Mayr, C., and Bartel, D.P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell *138*, 673–684.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature *385*, 357–361.

Miura, P., Shenker, S., Andreu-Agullo, C., Westholm, J.O., and Lai, E.C. (2013). Widespread and extensive lengthening of 3′ UTRs in the mammalian brain. Genome Res. *23*, 812–825.

Nagaike, T., Logan, C., Hotta, I., Rozenblatt-Rosen, O., Meyerson, M., and Manley, J.L. (2011). Transcriptional activators enhance polyadenylation of mRNA precursors. Mol. Cell *41*, 409–418.

Nègre, N., Brown, C.D., Ma, L., Bristow, C.A., Miller, S.W., Wagner, U., Kheradpour, P., Eaton, M.L., Loriaux, P., Sealfon, R., et al. (2011). Acis-regulatory map of the Drosophila genome. Nature *471*, 527–531.

O'Sullivan, J.M., Tan-Wong, S.M., Morillon, A., Lee, B., Coles, J., Mellor, J., and Proudfoot, N.J. (2004). Gene loops juxtapose promoters and terminators in yeast. Nat. Genet. *36*, 1014–1018.

Oktaba, K., Gutiérrez, L., Gagneur, J., Girardot, C., Sengupta, A.K., Furlong, E.E., and Müller, J. (2008). Dynamic regulation by polycomb group protein complexes controls pattern formation and the cell cycle in Drosophila. Dev. Cell *15*, 877–889.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. Nature *497*, 127–131.

Pfeiffer, B.D., Jenett, A., Hammonds, A.S., Ngo, T.T., Misra, S., Murphy, C., Scully, A., Carlson, J.W., Wan, K.H., Laverty, T.R., et al. (2008). Tools for neuroanatomy and neurogenetics in Drosophila. Proc. Natl. Acad. Sci. USA *105*, 9715–9720.

Rogulja-Ortmann, A., Picao-Osorio, J., Villava, C., Patraquim, P., Lafuente, E., Aspden, J., Thomsen, S., Technau, G.M., and Alonso, C.R. (2014). The RNA-binding protein ELAV regulates Hox RNA processing, expression and function within the Drosophila nervous system. Development *141*, 2046–2056.

Shi, Y. (2012). Alternative polyadenylation: new insights from global analyses. RNA *18*, 2105–2117.

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 Reports *1*, 277–289.

Soller, M., and White, K. (2003). ELAV inhibits 3′-end processing to promote neural splicing of ewg pre-mRNA. Genes Dev. *17*, 2526–2538.

Tan-Wong, S.M., Zaugg, J.B., Camblong, J., Xu, Z., Zhang, D.W., Mischo, H.E., Ansari, A.Z., Luscombe, N.M., Steinmetz, L.M., and Proudfoot, N.J. (2012). Gene loops enhance transcriptional directionality. Science *338*, 671–675.

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.

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.

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.

Wang, J.W., Beck, E.S., and McCabe, B.D. (2012). A modular toolset for recombination transgenesis and neurogenetic analysis of Drosophila. PLoS ONE *7*, e42102.

Yao, K.M., and White, K. (1994). Neural specificity of elav expression: defining a Drosophila promoter for directing expression to the nervous system. J. Neurochem. *63*, 41–51.

Zhang, H., Lee, J.Y., and Tian, B. (2005). Biased alternative polyadenylation in human tissues. Genome Biol. *6*, R100.

Zhang, D.W., Mosley, A.L., Ramisetty, S.R., Rodríguez-Molina, J.B., Washburn, M.P., and Ansari, A.Z. (2012). Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. J. Biol. Chem. *287*, 8541–8551.

Molecular Cell, Volume *57*

Supplemental Information

ELAV Links Paused Pol II to Alternative Polyadenylation in the *Drosophila* Nervous System

Katarzyna Oktaba, Wei Zhang, Thea Sabrina Lotz, David Jayhyun Jun, Sandra Beatrice Lemke, Samuel Pak Ng, Emilia Esposito, Michael Levine, and Valérie Hilgers

SUPPLEMENTAL FIGURES AND FIGURE LEGENDS

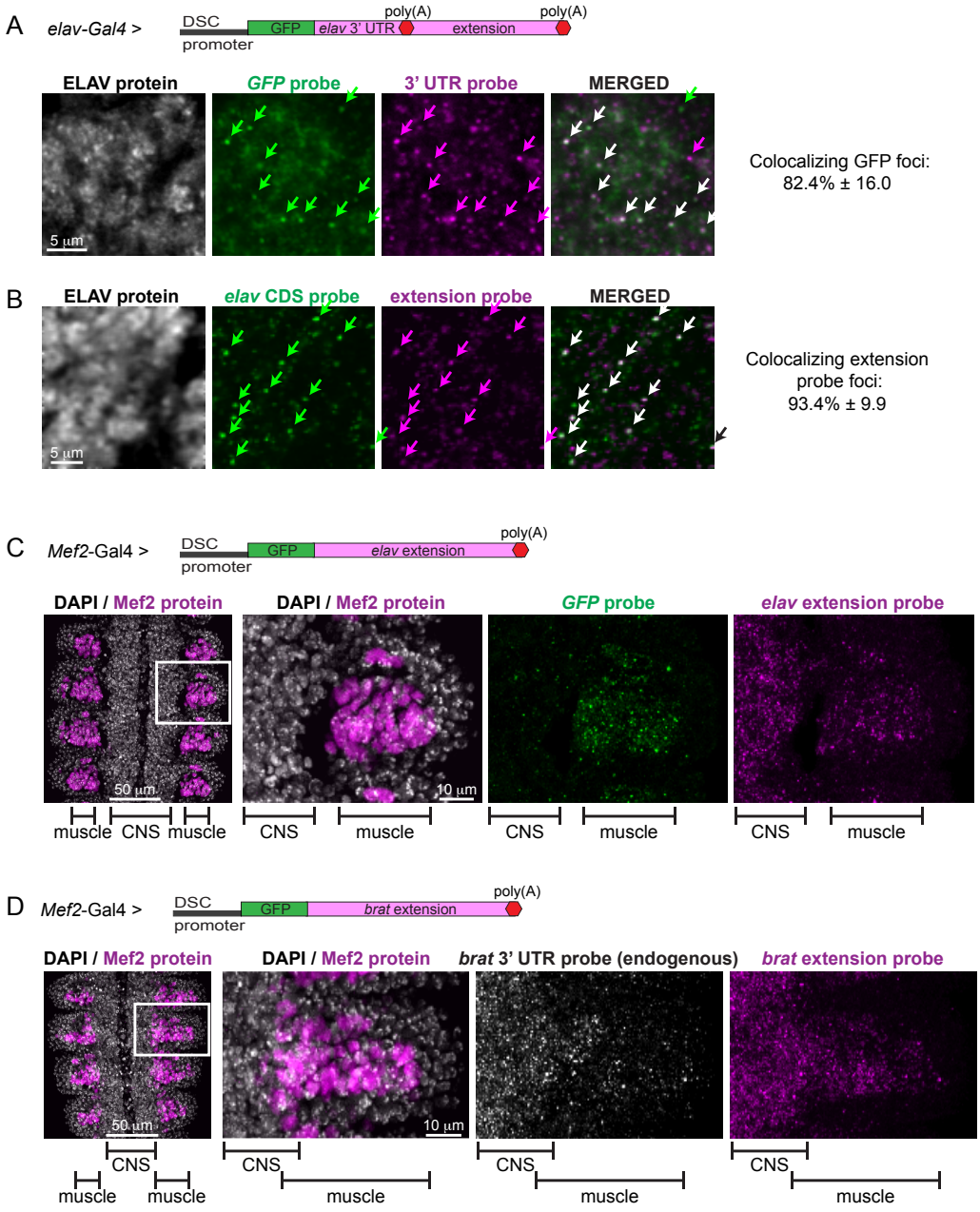


Figure S1. Transgene properties that promote expression of 3' extensions. Related to Figure 1.

A-B. Native promoters are required for expression of 3' extensions.

A. *elav-Gal4* drives expression of a *GFP* transgene in the nervous system. The promoter used for expression was the DSCP. The *GFP* coding sequence was

placed upstream of the entire extended 7.2 kb *elav* 3' UTR. CPA at the proximal poly(A) produces the short 3' UTR form of the mRNA, whereas CPA at the distal-most poly(A) produces the fully extended transcript. RNA probes directed against different regions of the transcripts were used to detect mRNAs. Shown are double fluorescent in situ hybridization assays. Single confocal sections of a portion of the developing CNS in stage 13 embryos. Colocalization of the *GFP* and 3' UTR probes indicates expression of the short transcript from the transgene.

B. Virtually all foci from the *elav* extension probe colocalize with the probe directed against the endogenous *elav* coding sequence, which indicates that the extension signal originates from the endogenous *elav* transcript. Numbers represent mean ± SD of six embryos for each sample.

C-D. Bypassing the requirement for ELAV recruitment allows for transcription of extension sequences from the DSCP.

Mef2-Gal4 drives expression of *GFP* transgenes in muscle cells. The promoter used for expression was the DSCP. The *GFP* coding sequence was placed upstream of the extended portion of the *elav* 3' UTR (C) or the extended portion of the *brat* 3' UTR (D), thereby excluding the respective short 3' UTRs and proximal poly(A) signals. CPA at the indicated poly(A) produces a transcript that was detected using RNA probes directed against the *GFP* coding sequence (C) as well as a distal region of the *elav* (C) or *brat* (D) 3' UTR extension. Shown are double fluorescent in situ hybridization assays combined with antibody staining against Mef2 protein as a muscle marker. Projections of consecutive confocal sections of stage 13 embryos. Ventral views; anterior is up.

C. The *GFP* mRNA signal in muscle cells shows muscle-specific expression of the transgene. Signal from the *elav* extension probe in the central nervous system (CNS) corresponds exclusively to the endogenous extended *elav* transcript. Detection of extension sequences in muscle cells indicates expression of extended transcripts from the transgene.

D. Signal from the *brat* 3' UTR (that is not present in the transgene mRNA) and *brat* extension probes in the CNS corresponds exclusively to endogenous *brat*

transcripts. Detection of extension sequences in muscle cells indicates expression of extended transcripts from the transgene.

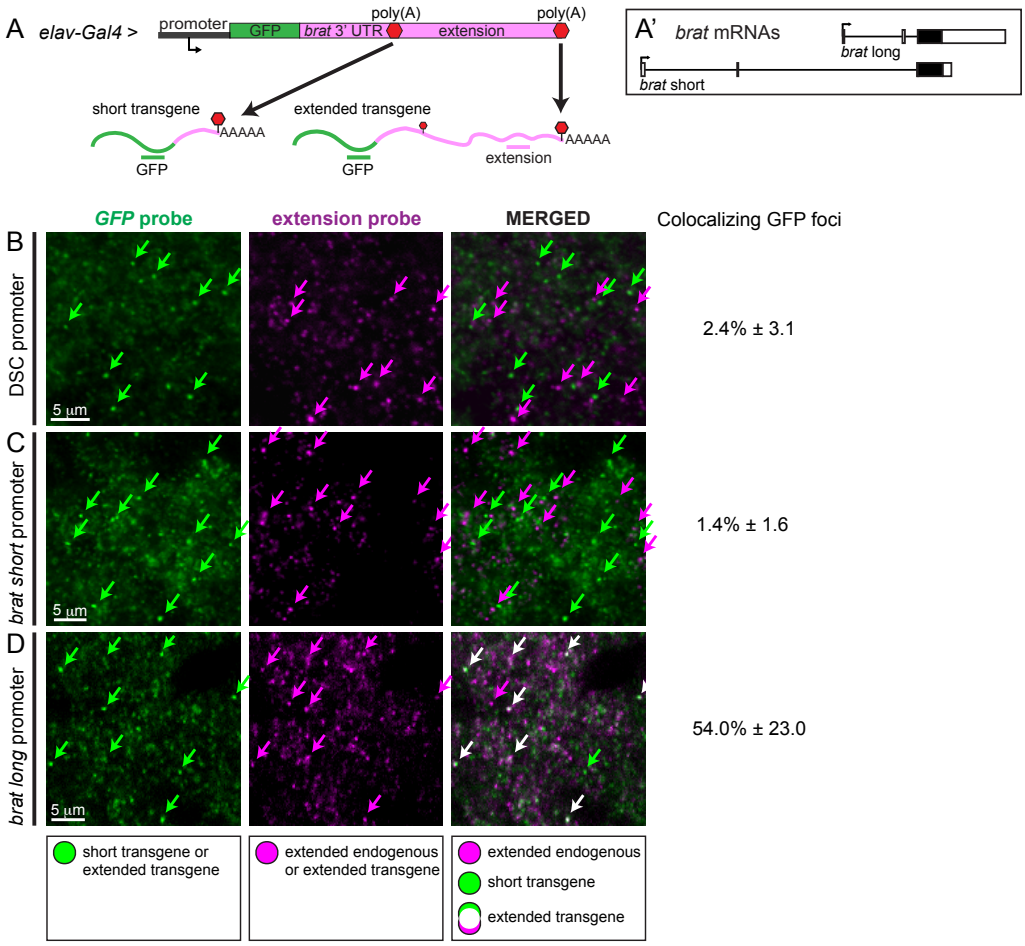


Figure S2. Native promoters are required for expression of 3' extensions. Related to Figure 1.

A. *elav-Gal4* drives expression of a *GFP* transgene in the nervous system. Three different promoter regions were used: DSCP (B), the native promoter producing the short form of *brat* (C), and the native promoter producing the extended form of *brat* (D). A' depicts the configuration of endogenous extended and short *brat* mRNAs with their respective promoters. The *GFP* coding sequence was placed upstream of the entire extended 8.5 kb *brat* 3' UTR. CPA at the proximal poly(A) produces the short 3' UTR form of the mRNA, whereas CPA at the distal-most

poly(A) produces the fully extended transcript. RNA probes directed against different regions of the transcripts were used to detect mRNAs.

B-D. Double fluorescent in situ hybridization assays using probes indicated in A. Single confocal sections of a portion of the developing CNS in stage 13 embryos. Note that the extension probe detects not only the transgene, but also the endogenous *brat* transcript, which is expressed in the nervous system. Colocalization of the *GFP* and extension probes indicates expression of extended transcripts from the transgene.

B,C. The reporter transgenes carrying the DSCP (B) or the native promoter of the short form of *brat* (C) do not exhibit colocalization of *GFP* and extension probes. Extension signals (magenta arrows in merged image) do not colocalize with the green *GFP* signals, indicating that they correspond to endogenous *brat* mRNAs.

D. Replacing the DSCP with the native promoter producing the extended form of *brat* induces 3' extension of the *GFP* transgene. There is extensive colocalization of the *GFP* (green arrows) and extension probes (magenta arrows), indicating expression of extension sequences from the transgene (white arrows in merged image). Non-colocalizing *GFP* signal (e.g., green arrow in merged image) corresponds to the short transgene, and non-colocalizing signal from the extension probe (e.g., magenta arrow in merged image) corresponds to the endogenously expressed extended *brat* mRNA. Numbers represent mean ± SD of six embryos for each promoter (except C: three embryos).

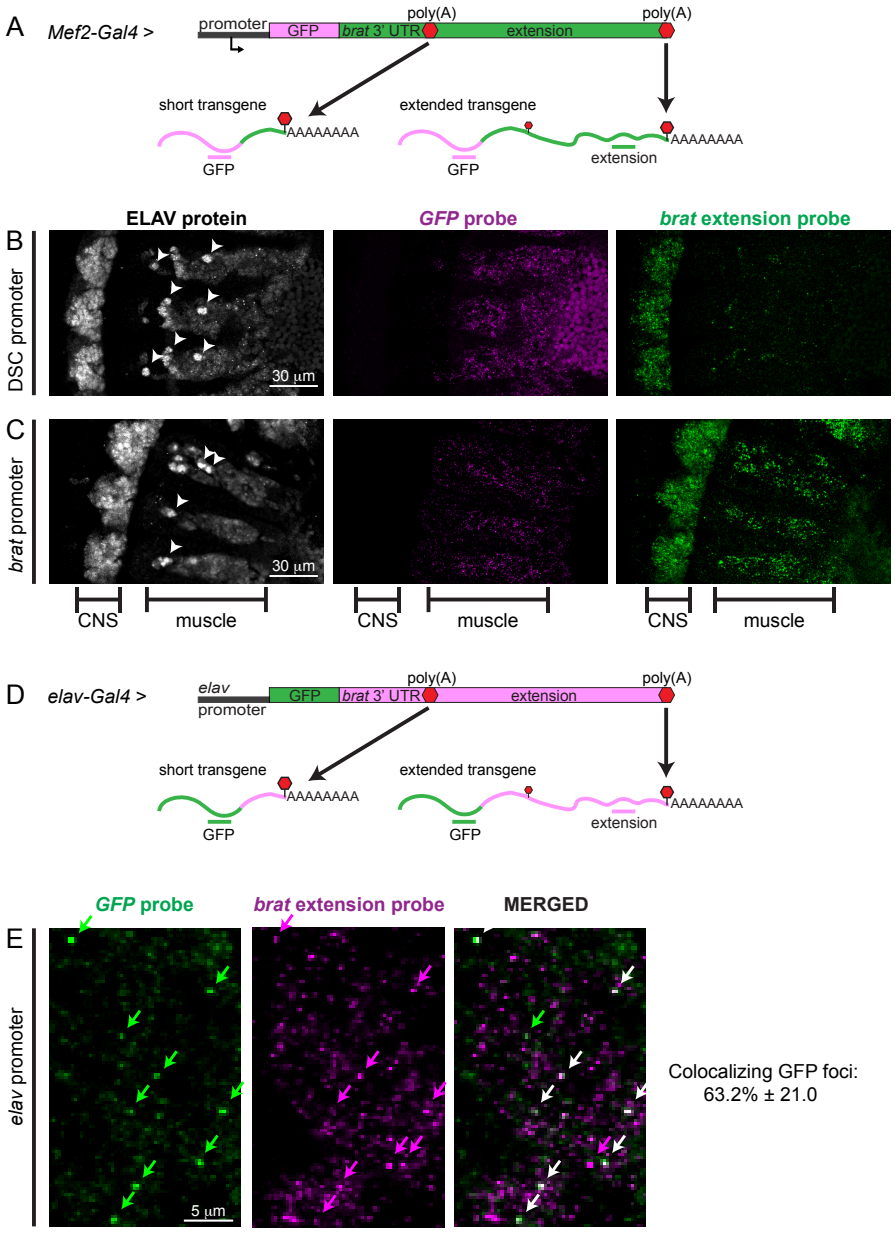


Figure S3. The native *brat* and *elav* promoters mediate *brat* 3' UTR extension. Related to Figure 2.

A-C: The native *brat* promoter mediates 3' UTR extension in ectopic tissues.

A: *Mef2-Gal4* drives expression of a *GFP* transgene in muscle cells. The promoter used for expression was either the DSCP, or the native *brat* promoter. The *GFP* coding sequence was placed upstream of the entire extended 8.5 kb *brat* 3' UTR. CPA at the proximal poly(A) produces the short 3' UTR form of the

mRNA, whereas CPA at the distal-most poly(A) produces the fully extended transcript. RNA probes directed against different regions of the transcripts were used to detect mRNAs.

B,C. Double fluorescent in situ hybridization assays using probes indicated in A, combined with antibody staining against ELAV protein. Projections of consecutive confocal sections of stage 13 embryos. Lateral views; anterior is up. The weak ELAV signal in muscle cells corresponds to ectopic expression driven by *Mef2-Gal4*, whereas the strong signal in the CNS corresponds to endogenous ELAV. Arrowheads indicate neurons of the PNS (strong ELAV signal). Ectopic expression of the short 3' UTR *GFP* transgene can be achieved from both DSC and *brat* promoters, as shown by detection of *GFP* probe signal in muscle cells in B and C (middle panels, magenta). Right panels exhibit hybridization signals with the *brat* extension probe (green). Signal in the CNS corresponds exclusively to the endogenous extended *brat* transcript, whereas expression in the muscle corresponds exclusively to reporter expression. Background staining in muscle is observed with the DSCP transgene (B), indicating little or no expression of the extended 3' UTR from the GFP transgene. In contrast, there is significant expression of extended transcripts from the transgene containing the *brat* promoter in muscle (C).

D,E. The native *elav* promoter mediates *brat* 3' UTR extension.

D. *elav-Gal4* drives expression of a *GFP* transgene in the nervous system. The promoter used for expression was the native *elav* promoter. The *GFP* coding sequence was placed upstream of the entire extended 8.5 kb *brat* 3' UTR. CPA at the proximal poly(A) produces the short 3' UTR form of the mRNA, whereas CPA at the distal-most poly(A) produces the fully extended transcript. RNA probes directed against different regions of the transcripts were used to detect mRNAs.

E. Double fluorescent in situ hybridization assays using probes indicated in D. Single confocal sections of a portion of the developing CNS in a stage 13 embryo. Note that the extension probe detects not only the transgene, but also the endogenous *brat* transcript, which is expressed in the nervous system. There is

extensive colocalization of the *GFP* (green arrows) and extension probes (magenta arrows), indicating expression of extended 3' UTR sequences from the transgene (white arrows in merged image). Non-colocalizing *GFP* signal (e.g., green arrow in merged image) corresponds to the short transgene, and non-colocalizing signal from the extension probe (e.g., magenta arrow in merged image) corresponds to the endogenously expressed extended *brat* mRNA. Numbers represent mean ± SD of six embryos.

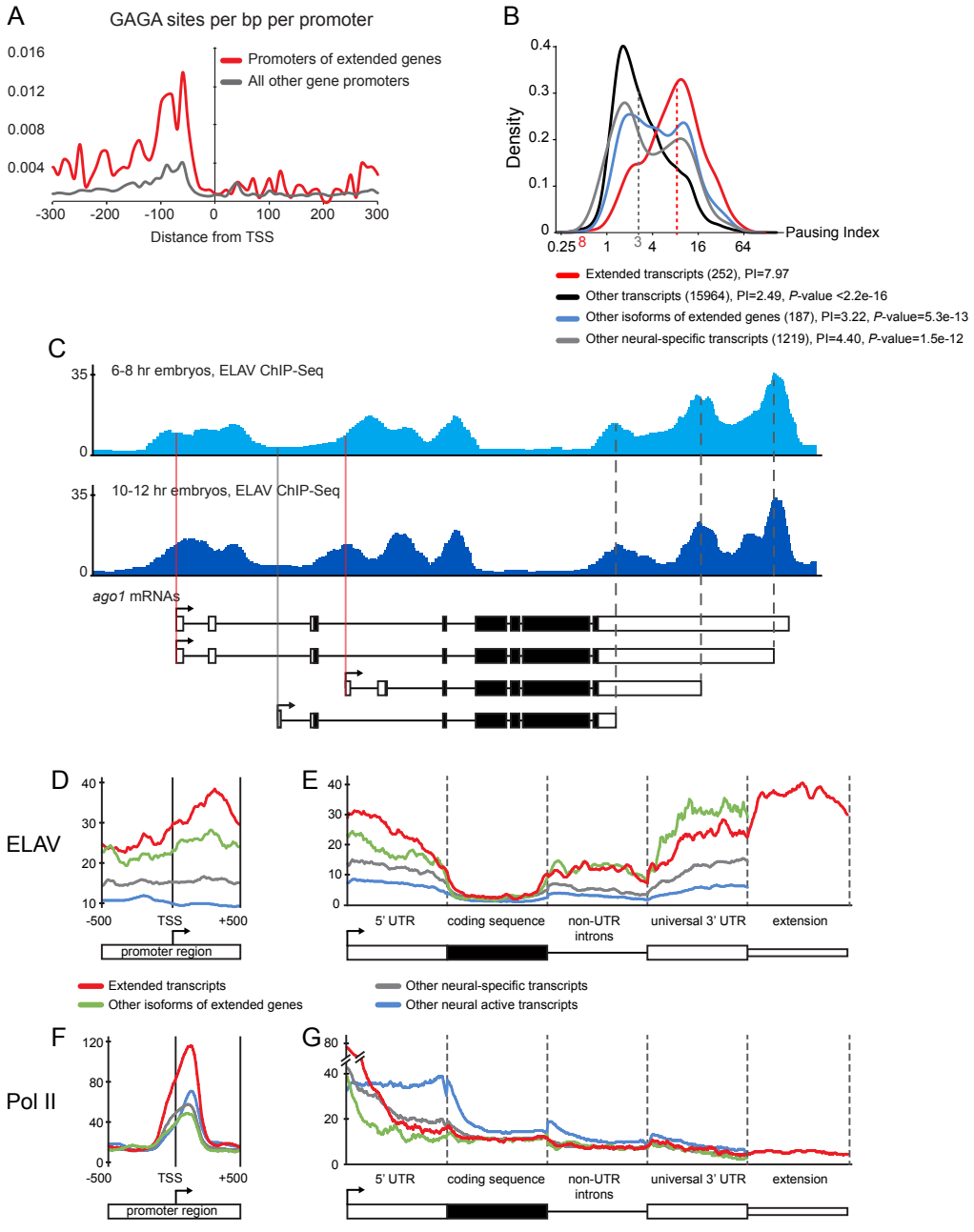


Figure S4. Promoters of extended genes contain the GAGA motif and paused Pol II and are bound by ELAV. Related to Figure 3 and Figure 4.

A. Frequency of occurrence and distribution of identified GAGA motifs in promoters of extended or control genes relative to the TSS. GAGA motifs are most often located between -100 bp and the TSS in both groups of promoters and occur significantly more frequently in promoters of extended genes.

B. Pausing index (PI) distribution and median pausing index values of the promoters of the indicated groups of transcripts in muscle tissues (see Supplemental Experimental Procedures), where ELAV is absent. The numbers in parentheses denote the number of transcripts in each group. Promoters of extended transcripts are significantly more paused than promoters of any control group. Wilcoxon rank sum test *P*-values were calculated by comparing the pausing index of extended transcripts with each group of controls.

C. Normalized ELAV ChIP-Seq reads at the *ago1* locus in 6-8 hr and 10-12 hr embryos. Shown are merged tracks of duplicate experiments. ELAV peaks at each proximal poly(A) site (dotted lines) are found in both 6-8 hr and 10-12 hr embryos.

D,E. Meta-gene plots of ELAV ChIP-Seq datasets at the promoter region (D) (±500 bp relative to the TSS) or across the entire transcription unit (E) in 10-12 hr embryos. Each line (meta-gene) averages the ChIP-Seq data of all indicated transcripts. ELAV binding is higher in extended transcripts compared to other transcripts (see exception below) at the promoter region, 5' UTR, introns and the 3' UTR. In all genes, ELAV binding is excluded from the coding sequence. Differences in ELAV binding between extended transcripts and 'other isoforms of extended genes' are not significant. We think the reason is that transcripts from these two groups share many gene regions including sequences as close as ±100 bp relative to the TSS, introns and the universal 3' UTR. Moreover, both groups of transcripts are relatively small (252 and 187 transcripts, respectively).

F,G. Meta-gene analysis of Pol II binding at promoter region (F) or across the entire transcription unit (G) in 12-16 hr embryos. Promoter regions of extended transcripts show significantly higher Pol II binding than other control groups of transcripts. Other regions downstream of the TSS do not differ in their Pol II binding profile between the four groups, except the "other neuronal active transcripts" that show higher Pol II binding at the 5' UTR and the coding sequence due to their high level of expression. See also Table S1 and Table S2 for ELAV peak coordinates.

Table S1. Listing of chromosomal coordinates (UCSC dm3 release) of 6879 ELAV binding peaks in 6-8 hr embryos identified by ChIP-Seq.

Table S2. Listing of chromosomal coordinates (UCSC dm3 release) of 8076 ELAV binding peaks in 10-12 hr embryos identified by ChIP-Seq.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Primers

Native promoter sequences (300 bp surrounding the transcription start site) were amplified from genomic DNA and cloned into pBID-UASc-eGFP SacI/BglII, thus removing the DSC promoter and maintaining the UAS repeats.
Primers for *brat short*:
forward: 5'- TCCCATTTTGAATTTAAGTAAAACTTAGCC
reverse: 5'- AATTGGCCACAGAACAAAGCG
Primers for *brat long*:
forward: 5'- GTTGAGTGAGTTTTTTCGGCTG
reverse: 5'- ATAGGCTAGGTATGTGTATGTTTCTGTTG
Primers for *elav*:
forward: 5'-CTCGAGAGGCAACTATGAGATATGAG,
reverse: 5'- ATTTCGCTCGGTGTGAGATGA.

Extended 3' UTR sequences were amplified from genomic DNA and cloned into the modified pBID-UASc-eGFP NotI/XbaI.
Primers for *brat*:
forward: 5'-CACACAGACACACACACTCCATG
reverse: 5'- TGCCAAAACTCACTGATCGAATAA
Primers for *elav*:
forward: 5'-AGCGGCCCAAATGGAAG

reverse: 5'-TCGGTCATAGTGTCATTTATTCCAT

Extension sequences lacking the short 3' UTR and the proximal poly(A) were cloned in the same way.
Primers for *elav*:
forward: 5'-CCAATTTACCTATTAAGTAAGCAAAGAGC
reverse: 5'- CTTCGTAATTAAAACAACCCTTTCAGT
Primers for *brat*:
forward: 5'- AAAACAAGGCGATATTTATGTGCA
reverse: 5'- AGTCATTTAAGTCATTTATGCTTGCC

Computational filtering of the extended transcripts and the control groups

We focused our analysis on known neural-specific extended genes. We used the RefSeq (release 65) annotation and published RNA-seq data (Gaertner et al., 2012; Graveley et al., 2011) to filter the 401 described genes that undergo 3' UTR extension (Hilgers et al., 2011; Smibert et al., 2012). The transcripts of these 401 genes which satisfied all the following 3 criteria were included in this study: (1) had a 3' UTR extension of at least 200 bp, (2) had at least 1 read per kilobase of exon model per million mapped reads (RPKM) in either 10-12 hr or 14-17 hr embryo neurons, and (3) had significantly more expression in neurons than in muscle cells in 10-12 hr and 14-17 hr embryos (*P*-value ≤ 0.05 by RankProd (Hong et al., 2006)). After these filters, if several extended isoforms shared the same transcription start site (TSS), the longest extended isoform was used. If several isoforms shared the same transcription termination site (TTS), the isoform with the highest expression in neuron cells was used. 252 transcripts of 219 genes with 3' UTR extensions were included in the analysis.

Four control groups of transcripts were used in this study: (A) all transcripts except the known 3' UTR extended transcripts, (B) all non-extended isoforms of the 219 extended genes, (C) all non-extended transcripts that are active in

neurons (RPKM > 10 in neurons) and (D) all non-extended transcripts that are specifically expressed in neurons (RPKM > 1 in neurons and RankProd *P*-value ≤ 0.05 when comparing their expression in neurons vs. muscle cells). All isoforms containing an annotated 3' UTR extension of at least 200 bp were excluded from all control sets. If several isoforms shared the same TSS, only the one with the highest expression in whole embryos (Negre et al., 2011) was included. The four control sets consist of (A) 15964, (B) 187, (C) 5841 and (D) 1219 transcripts.

Determination of pausing indexes

We defined promoter regions as 200 bp surrounding +30 bp from the annotated TSS in RefSeq (release 65), and defined gene body regions as TSS +400 bp to the 3' end of the genes. Genes whose size did not exceed 400 bp were excluded. Pol II enrichment at promoters and gene bodies was calculated as the fold enrichment of the normalized Pol II reads over input (Negre et al., 2011). Pausing index was defined as the Pol II enrichment at promoter regions divided by the Pol II enrichment within gene bodies. For each promoter, the maximum pausing index in 4-24 hr embryos was used. For analysis of pausing indexes in muscle tissue, the maximum pausing index in the following samples was used: 2-4h Toll10b mutant embryos, 6-8h, 8-10h, 10-12h, and 14-17h Mef2-sorted muscle cells (Gaertner et al., 2012).

Chromatin Immunoprecipitation and sequencing

Chromatin was prepared from 0.5 g of dechorionated 6-8 hr and 10-12 hr wild-type (*yw*) embryos. Two independent chromatin preparations (biological replicates) were done for each time point. Sonication of chromatin was performed with a Bioruptor (Diagenode) yielding genomic DNA fragments with an average size of ~200 bp. ChIP assays were done using a mix of two antibodies (5 ug each) raised against the entire ELAV protein (483 aa): mouse anti-ELAV-9F8A9

and rat anti-ELAV-7E8A10 (DSHB). ChIP DNA was resuspended in 40 ul, 8 ul of which were used to evaluate specific enrichment by qPCR. The remaining 32 ul were used to construct ChIP-Seq libraries.

ELAV and Pol II ChIP-Seq data processing

All sequencing reads were aligned to the *Drosophila melanogaster* reference genome (UCSC dm3 release) using Bowtie version 0.12.7 (Langmead et al., 2009). The following Bowtie parameters were used to select only uniquely aligning reads with a maximum of two mismatches:

-k 1 -m 1 -n 2 --best --strata

We used the model-based analysis of ChIP-Seq (MACS) peak-finding algorithm version 1.4.1 to identify the regions of ELAV and Pol II ChIP-Seq enrichment over input DNA. The default MACS parameters were applied. For each time point, only the overlapped ELAV-binding regions between two biological replicates were used for further analysis. In total, 6879 and 8076 ELAV-binding regions were identified in 6-8 hr and 10-12 hr embryos, respectively.

SUPPLEMENTAL REFERENCES

Gaertner, B., Johnston, J., Chen, K., Wallaschek, N., Paulson, A., Garruss, A.S., Gaudenz, K., De Kumar, B., Krumlauf, R., and Zeitlinger, J. (2012). Poised RNA polymerase II changes over developmental time and prepares genes for future expression. Cell reports *2*, 1670-1683.

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.

Hilgers, V., Perry, M.W., Hendrix, D., Stark, A., Levine, M., and Haley, B. (2011). Neural-specific elongation of 3′ UTRs during Drosophila development. Proceedings of the National Academy of Sciences *108*, 15864-15869.

Hong, F., Breitling, R., McEntee, C.W., Wittner, B.S., Nemhauser, J.L., and Chory, J. (2006). RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. Bioinformatics (Oxford, England) *22*, 2825-2827.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Negre, N., Brown, C.D., Ma, L., Bristow, C.A., Miller, S.W., Wagner, U., Kheradpour, P., Eaton, M.L., Loriaux, P., Sealfon, R.*, et al.* (2011). A cis-regulatory map of the Drosophila genome. Nature *471*, 527-531.

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 reports *1*, 277-289.