© 2019. Published by The Company of Biologists Ltd | Development (2019) 146, dev175265. doi:10.1242/dev.175265

TECHNIQUES AND RESOURCES

RESEARCH ARTICLE

A gene expression atlas of embryonic neurogenesis in *Drosophila* reveals complex spatiotemporal regulation of lncRNAs

Alexandra L. McCorkindale[1,2,*], Philipp Wahle[1], Sascha Werner[1], Irwin Jungreis[3,4], Peter Menzel[5], Chinmay J. Shukla[6,7], Rúben Lopes Pereira Abreu[1], Rafael A. Irizarry[7], Irmtraud M. Meyer[5,8], Manolis Kellis[4,5] and Robert P. Zinzen[1,*]

ABSTRACT

Cell type specification during early nervous system development in *Drosophila melanogaster* requires precise regulation of gene expression in time and space. Resolving the programs driving neurogenesis has been a major challenge owing to the complexity and rapidity with which distinct cell populations arise. To resolve the cell type-specific gene expression dynamics in early nervous system development, we have sequenced the transcriptomes of purified neurogenic cell types across consecutive time points covering crucial events in neurogenesis. The resulting gene expression atlas comprises a detailed resource of global transcriptome dynamics that permits systematic analysis of how cells in the nervous system acquire distinct fates. We resolve known gene expression dynamics and uncover novel expression signatures for hundreds of genes among diverse neurogenic cell types, most of which remain unstudied. We also identified a set of conserved long noncoding RNAs (lncRNAs) that are regulated in a tissue-specific manner and exhibit spatiotemporal expression during neurogenesis with exquisite specificity. lncRNA expression is highly dynamic and demarcates specific subpopulations within neurogenic cell types. Our spatiotemporal transcriptome atlas provides a comprehensive resource for investigating the function of coding genes and noncoding RNAs during crucial stages of early neurogenesis.

KEY WORDS: *Drosophila melanogaster*, Embryogenesis, lncRNA, Spatiotemporal transcriptome, Neurogenesis

INTRODUCTION

Development of complex tissues from naïve primordia requires the precise spatiotemporal deployment of transcriptional programs as cells subdivide, specify and differentiate. Owing to the availability of tissue- and cell type-specific markers characteristic for neurogenic cell types in the fruit fly embryo (Heckscher et al., 2014), *Drosophila* neurogenesis is highly tractable and several crucial regulators of neurogenesis have been identified over the past several decades (Skeath and Thor, 2003; Beckervordersandforth et al., 2008; Broadus et al., 1995; Landgraf et al., 1997; Rickert et al., 2011; Wheeler et al., 2006; Doe, 2017; Heckscher et al., 2014; Skeath et al., 1994; Weiss et al., 1998; Wheeler et al., 2009). Among the earliest events in embryonic neurogenesis is the subdivision of the lateral neurogenic ectoderm into columnar domains along the dorsoventral axis (Von Ohlen and Doe, 2000; Cowden and Levine, 2003). This is followed by the formation of proneural clusters and consecutive phases of delamination, whereby neuroblasts cease contact with surrounding cells of the neuroectodermal columns and ingress into the embryo (Campos-Ortega, 1995). Embryonic neuroblasts – *Drosophila* neural stem cells – undergo a series of self-renewing asymmetric divisions that produce ganglion mother cells, which give rise to glia and neurons (Broadus et al., 1995; Sousa-Nunes et al., 2010; Homem and Knoblich, 2012; Heckscher et al., 2014). Importantly, each of the three neurogenic columns gives rise to molecularly and functionally distinct sets of neuroblasts (Doe, 1992), but the molecular mechanisms that link spatial origin to the ensuing distinct fates remain poorly understood. To date, a small set of marker genes specifically expressed in individual columnar domains and in emerging cell types has been identified, but it remains unclear how these cell populations differ with respect to the global gene expression programs that shape their identities.

Although expression dynamics of protein-coding transcripts have given important insights into the mechanisms that drive cellular differentiation, it should be noted that an emerging class of noncoding transcripts – the long noncoding RNAs (lncRNAs) – may well emerge as pivotal regulators of neurogenesis. In mammals, lncRNAs have been shown to be especially abundant in differentiated neuronal cells (Briggs et al., 2015), are expressed often with exquisite spatiotemporal specificity in the nervous system (Sauvageau et al., 2013; Goff et al., 2015), and some lncRNA species even exhibit neuronal subtype specificity (Molyneaux et al., 2015; Liu et al., 2016). Though the functional importance of some lncRNAs for development and cellular identity has been demonstrated in *Drosophila* (Wen et al., 2016), including in the nervous system (Li and Liu, 2015; Landskron et al., 2018), very little is known about the cell type-specific expression and function of lncRNAs over the course of early neurogenesis.

Large-scale efforts have characterized spatial gene expression in RNA *in situ* hybridization screens (Tomancak et al., 2002; Inagaki et al., 2005; Tomancak et al., 2007; Lécuyer et al., 2007; Wilk et al., 2016), but such efforts are qualitative rather than quantitative and largely exclude lncRNAs. In contrast, efforts to determine global

[1]Laboratory for Systems Biology of Neural Tissue Differentiation, Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrueck Centre for Molecular Medicine (MDC) in the Helmholtz Association, Robert-Roessle-Strasse 12, 13125 Berlin, Germany. [2]Biofrontiers Institute, University of Colorado, Boulder, CO 80303, USA. [3]MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA 02139, USA. [4]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [5]Laboratory for Bioinformatics of RNA Structure and Transcriptome Regulation, Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrueck Centre for Molecular Medicine (MDC) in the Helmholtz Association, Robert-Roessle-Strasse 12, 13125 Berlin, Germany. [6]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA. [7]Dana Farber Cancer Institute, Boston, MA 02215, USA. [8]Freie Universität, Institute of Biochemistry, Department of Biology, Chemistry, Pharmacy, Thielallee 63, Berlin 14195, Germany.

*Authors for correspondence (ali.mccorkindale@inventia.life; robert.zinzen@mdc-berlin.de)

A.L.M., 0000-0002-0840-0784; R.P.Z., 0000-0002-8638-5102

Received 27 December 2018; Accepted 5 February 2019

Development (2019) 146, dev175265. doi:10.1242/dev.175265

transcriptome dynamics in the developing *Drosophila* embryo (Graveley et al., 2011; Brown et al., 2014; Young et al., 2012; Chen et al., 2016) may detect the expression of lncRNAs, but lack cell type resolution. As for most complex tissues, recapitulating early neurogenesis in cell culture is unfortunately not an option, because accurate specification and differentiation of cells depends on embryonic context, intricate interactions among cells within the neuroectoderm (Kunisch et al., 1994; Lai, 2004) and signaling gradients involving surrounding tissues (Bier and De Robertis, 2015; Rogers et al., 2017).

To overcome these limitations and to dissect stage- and cell type-specific transcriptomes in early neurogenesis, we adapted MARIS (Hrvatin et al., 2014) for use in developing *Drosophila* embryos. DIV-MARIS (*Drosophila in vivo* method for analyzing RNA following intracellular sorting) allows purification of chemically cross-linked cell types from staged developing embryos based on marker gene expression, followed by RNA extraction and next-generation sequencing. Here, we employ DIV-MARIS to determine the transcriptome dynamics in distinct neurogenic cell populations. We assess the gene expression programs of two principal neurogenic domains (the ventral and the intermediate columns) and of three differentiating cell types (neuroblasts, neurons and glia) at consecutive time points from primordial specification and subdivision to terminal differentiation.

DIV-MARIS reveals an extensive network of dynamic spatiotemporal gene expression during embryonic nervous system development. Our method reliably identifies known cell type-specific markers, but also reveals novel expression features. Furthermore, we uncover many genes – most of which have conserved homologs in human – that are expressed in distinct cell types throughout early neurogenesis and for which the functions remain to be elucidated. Hence, DIV-MARIS provides an accurate expression map of spatiotemporal transcriptional programs driving early nervous system development. Moreover, our analyses identified many lncRNAs expressed in cell type-specific patterns and for which no functional roles are yet known. Applying stringent criteria for selection, we characterize 13 neural cell type-enriched lncRNAs with varied temporal expression, abundance and subcellular localization. *In situ* visualization of lncRNA expression exposes an additional layer of specificity as neuroglial lncRNAs tend to be expressed highly, but only in extremely distinct subpopulations.

This study delivers a genome-wide, yet cell type-specific, view of gene expression during *Drosophila* neurogenesis from neurogenic columns to differentiated neurons and glia, provides insights into the expression properties of the coding and noncoding transcriptomes and will serve as a valuable tool for understanding how regulated coding and noncoding gene expression drives cell fate determination in early neurogenesis.

RESULTS

Isolation of neuroglial cell types with spatiotemporal resolution

Early *Drosophila* neurogenesis starts with the specification of the lateral neurogenic ectoderm at the onset of zygotic transcription. The neurogenic ectoderm is quickly subdivided into distinct neurogenic columns (Von Ohlen and Doe, 2000; Cowden and Levine, 2003), from which neuroblasts delaminate and undergo asymmetric division giving rise to ganglion mother cells (GMCs), followed by differentiation of GMCs into neurons and glia (Fig. 1A). To dissect the genome-wide transcriptional programs driving early neurogenesis, we purified specific cell populations comprising the neuroglial lineages using fluorescence-activated cell sorting (FACS) of chemically fixed cells. We isolated cells of the intermediate column (IC) and the ventral column (VC) using transgenic constructs by fusing IC- or VC-specific enhancers to reporter genes (Fig. S1A). Neuroblasts/GMCs, neurons and glia cells were purified using antibodies directed against the endogenous markers *prospero* (*pros*), *embryonic lethal abnormal vision* (*elav*) and *reversed polarity* (*repo*), respectively (Fig. 1D, Fig. S1B). Early neurogenesis is a rapidly unfolding process, with naïve primordia developing into differentiated cell types in a matter of hours (Fig. S2A). To assess the temporal dynamics of early neurogenesis, we collected these cell populations at developmental stages (bins) that encompass crucial events along the neurogenic lineages from early specification to terminal differentiation (Fig. 1B, Fig. S2A). Timed embryo collections were manually staged to confirm which neurogenic events were captured within the collection bins (Fig. S2B). The earliest collection bin (4-6 h after egg laying, AEL) primarily contains embryos immediately after specification and subdivision of the neurogenic ectoderm and encompasses the first rounds of neuroblast delamination. The second bin (6-8 h AEL) includes all waves of neuroblast delamination, proliferation and diversification, followed by early differentiation in the third bin (8-10 h AEL). A later collection towards the end of embryogenesis (18-22 h AEL) serves as a reference point for fully differentiated neurons and glia.

To isolate cell type-specific RNA from specific neurogenic cell types, we adapted the MARIS protocol (Hrvatin et al., 2014), but had to introduce several modifications to temporally resolve cell types from complex and quickly developing *Drosophila* embryos *in vivo*. DIV-MARIS (outlined in Fig. 1C) is a flexible method for the isolation of high-quality RNA from specific fixed cell types within complex and rapidly developing embryos. Briefly, staged embryos are collected, dissociated into single-cell suspensions, and immediately cross-linked with formaldehyde. Neurogenic cell types were stained using antibodies, either against transgenic reporters (for the ventral and intermediate columns; Fig. S1A), or against endogenous markers (for neuroblasts/GMCs, neurons and glia; Fig. 1D, Fig. S2B). Positively marked and unmarked populations were purified by FACS (Fig. 1C). We used microscopy (e.g. Fig. S1C) and analytical cytometry (e.g. Fig. S3) to confirm that the sorting strategy reliably isolated marked cells of interest; samples generally had purities >95% and samples below 90% purity were discarded. Furthermore, we evaluated the enrichment of DIV-MARIS-sorted cell types by quantitative RT-PCR against several marker genes associated with the cell types of interest (i.e. *pros* and *worniu* in neuroblasts/GMCs, *elav* and *Lim3* in neurons, *repo* and *gcm* in glia) as independent measures of cell-type enrichment (Fig. 1E). We confirmed specific enrichment of the expected markers in sorted cells compared with whole embryos, as well as their depletion in sorted marker-negative cells.

As DIV-MARIS robustly isolates neurogenic cell populations of interest, we extracted RNA from sorted populations at four developmental time points for whole-transcriptome sequencing. Principal component analysis demonstrated that variance between samples is primarily due to developmental time and cell type of origin (Fig. S4). The resulting cell type-specific gene expression atlas quantitatively assesses neurogenic transcription in five distinct neurogenic cell populations (enriched and depleted) across four developmental time points covering major neurogenic events (Fig. 1A,B).

Development (2019) 146, dev175265. doi:10.1242/dev.175265

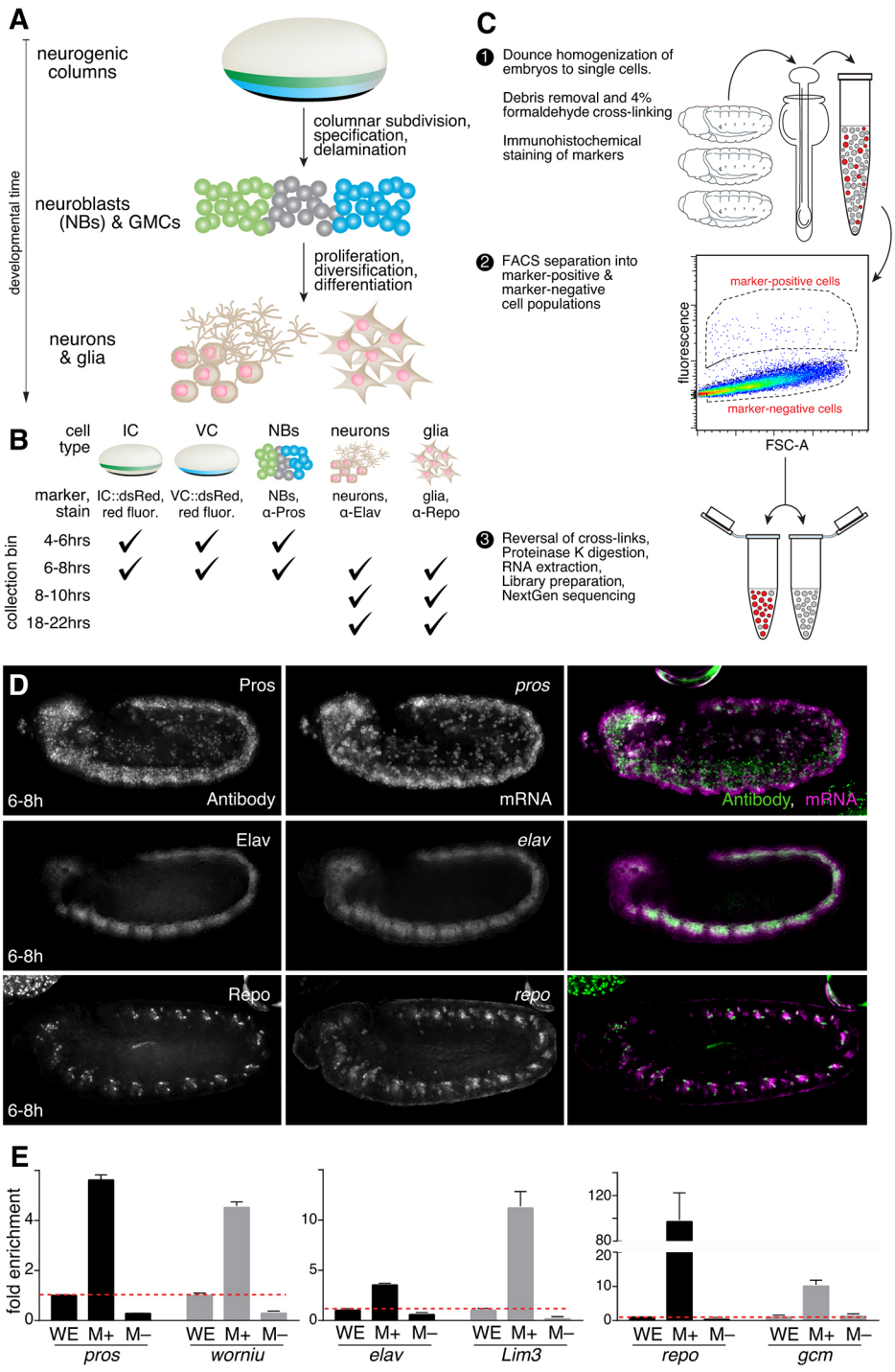


Fig. 1. DIV-MARIS for the enrichment of staged neurogenic cell types. (A) Biological materials studied over the course of neurogenesis: intermediate column (IC); ventral column (VC); neuroblasts (NBs); neurons; glia. (B) Time windows (collection bins) of sorted materials. (C) Overview of the DIV-MARIS protocol. (D) Merge (right) of antibody (left) and RNA-FISH (middle) shows that the sorting strategy faithfully marks cell types of interest (NBs, Pros; neurons, Elav; glia, Repo); embryos at 6-8 h. (E) Expression of marker genes specific to the cell type of interest measured by qPCR in marker-enriched (M+), and marker-depleted (M–) populations, calculated relative to whole embryo (WE, dashed red line); embryos collected at 4-10 h AEL. Error bars represent s.e.m. Embryos are ∼500 µm in length, shown anterior leftwards and ventral downwards; stage 10/11.

Cell type-specific expression of protein-coding genes during neurogenesis

In addition to purity, we evaluated sorting specificity by assessing gene expression of the five cell type marker genes (*ind*, *vnd*, *pros*, *elav* and *repo*) across the sorted populations in terms of normalized counts (Table S1). In all cases, strong enrichment of marker gene expression levels in the marker-enriched compared with the depleted samples was observed, as expected (Fig. 2A, Fig. S5). For example, the high and near-exclusive enrichment of *repo* transcript in purified glia demonstrates sorting effectiveness of DIV-MARIS when using a highly specific and exclusive marker (Fig. 2A, Fig. S5E). Similarly, *elav* transcript levels were highly enriched in purified neurons compared with glia (Fig. 2A, Fig. S5D), whereas lower levels were detected in early neuroblasts and columnar material, which is in line with observations that the common neuronal marker *elav* is transiently expressed pan-neurogenically at the onset of differentiation (Berger et al., 2007). The columnar markers *vnd* and *ind* mark distinct columnar neurogenic territories that each give rise to neuroblasts, neurons and glia. Accordingly, although *vnd* and *ind* transcripts were largely exclusive to their respective neurogenic columns, each was detectable to some degree in neuroblasts, most likely because

Development (2019) 146, dev175265. doi:10.1242/dev.175265

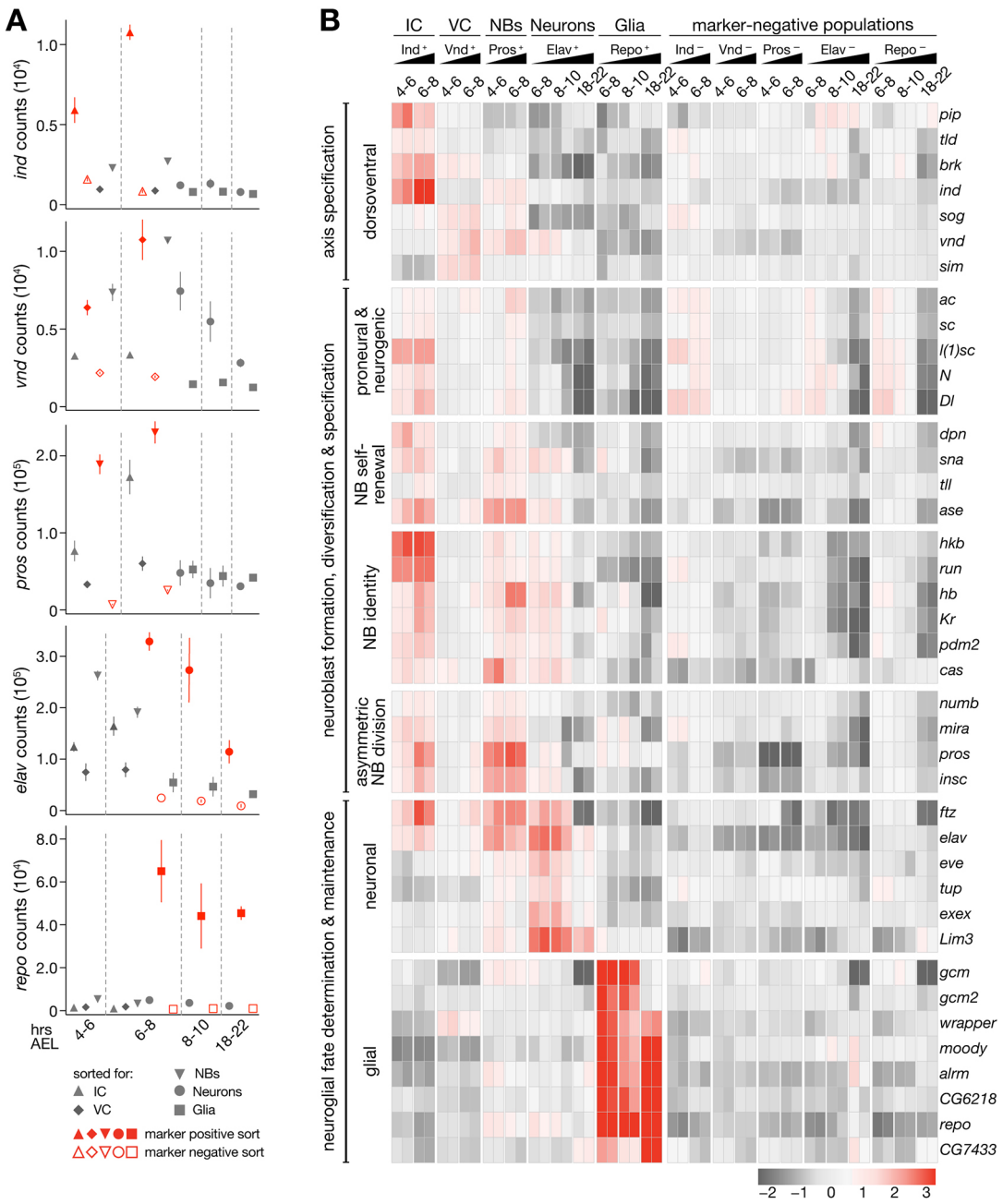


Fig. 2. Defining mRNA signatures of neuroglial cell types. (A) Normalized expression values for each marker gene used for FACS (*ind*, *vnd*, *pros*, *elav* and *repo*) across sorted samples. Error bars represent s.e.m. (B) Heat map of expression profiles of *Drosophila* nervous system genes. Row mean-centered expression values calculated by variance-stabilizing transformation (VST) of gene-level RNA-seq counts (scale=$\log_2$ ratio of row mean).

early neuroblasts stem from one of the respective neurogenic columns co-purified by FACS (Fig. 2A, Fig. S5A-C). Interestingly, normalized counts for *vnd* were higher than those for *ind*, which likely reflects the fact that the ventral column generates more neuroblasts in the first waves of delamination compared with the intermediate column (Doe, 1992).

To validate cell type-specific gene expression, we examined genes with known neurogenic roles (Von Ohlen and Doe, 2000; Skeath and Thor, 2003; Sousa-Nunes et al., 2010; Crews, 2010; Sandler and Stathopoulos, 2016) and confirmed specificity of mRNA expression in cell types previously associated with gene function (Fig. 2B). *exex*, for example, is a homeodomain transcription factor required in motor neurons that project to ventral somatic muscles (Santiago et al., 2014) and we found it exclusively in young neurons (Fig. 2B). Whereas markers of neuroblast identity were not only enriched in neuroblasts, but also depleted in the differentiated cell types neuroblasts give rise to (neuronal and especially glial), neurogenic column marker expression was often maintained in neuroblasts, highlighting that neuroblasts retain columnar identity after delamination, as they adopt column-specific fates (Doe, 1992).

To systematically uncover protein-coding genes that demarcate columnar and cell-type identities in nervous system development, we looked for genes expressed in a similar pattern to known

Development (2019) 146, dev175265. doi:10.1242/dev.175265

neurogenic genes by Pearson correlation (r>0.9). We uncovered 753 additional genes (summarized in Table S2) and though many have no known association with embryonic neurogenesis, *in situ* screens annotating expression using controlled anatomical imaging vocabulary (ImaGO; Hammonds et al., 2013; Tomancak et al., 2002, 2007) indicate that this gene set is indeed specifically expressed in components of the developing nervous system. For example, the most enriched ImaGO terms for this gene set include 'ventral nerve cord primordium', 'brain primordium' and 'ventral nerve cord' (Fig. S6A). GO analysis revealed the most enriched molecular function for this gene set to be 'DNA binding', and the most enriched biological processes were 'chromosome organization' and 'nucleic acid metabolic process' (Fig. S6B,C). Furthermore, protein domains enriched among the proteins specifically expressed in compartments of the developing nervous system included histone folds, chromatin interaction domains and sequence-specific DNA-interaction domains, such as zinc fingers and homeobox domains (Fig. S6D).

We were surprised that one-quarter of the genes deployed similarly to known neurogenic marker genes remain largely unstudied (199 'computed genes') and though many of these candidates lack any described function, more than 62% can be directly mapped to human homologs and some have even been linked to nervous system function.

We focused on a subset (40) of these genes, which were predicted to be expressed in neuroglial cell types with clear spatiotemporal specificity (Fig. S7A). In concordance with DIV-MARIS predictions, RNA *in situ* hybridization data (Hammonds et al., 2013; Tomancak et al., 2002, 2007) confirmed that a selection of these candidate genes mark specific subsets of cells in the developing nervous system (Fig. S7B).

Thus, DIV-MARIS reliably captures and uncovers cell type-specific gene expression dynamics during embryogenesis. As many of the specifically expressed genes encode known and predicted transcription factors and signaling pathway components (Table S2), this cell type-specific expression map identifies new regulatory nodes that likely play central roles in the specification and differentiation of neuroglial cell types.

Specific expression and properties of long noncoding RNAs along the neuroglial lineage

To explore lncRNA expression during early neurogenesis, we first identified nervous system-specific lncRNAs by calculating enrichment of expression in marker-positive versus marker-depleted samples at each time point using DESeq2 (Love et al., 2014; $\log_2$FC>1.0, *Padj*<0.05). We found 325 such lncRNA candidates (Table S3) and evaluated them according to several criteria, including spatiotemporal regulation through neurogenesis, expression above an abundance threshold [transcript counts per million reads (TPM)>300] in at least one cell type, absence of sense overlap with a protein-coding gene, and transcript boundaries consistent with lncRNA annotations. Applying these stringent criteria, we selected 13 high-confidence lncRNA candidates that are strongly and specifically expressed in a variety of cell types of the *Drosophila* nervous system (Fig. 3, Fig. S8).

To assess spatiotemporal expression of these lncRNAs, we calculated their relative abundance among all cell types and collection bins. The lncRNAs were depleted in marker-negative non-neurogenic cells and exhibited dynamic spatiotemporal enrichment in specific marker-positive cell types (Fig. 3A). Strikingly, although we found very few lncRNAs with distinct expression in the earlier and more naïve intermediate and ventral columns, specific lncRNA deployment could be readily observed in more mature and differentiated cell types, such as neuroblasts, neurons and/or glia, indicating that lncRNA expression is a hallmark of differentiated cells more so than of primordia.

To confirm that these transcripts are bona fide lncRNAs, we evaluated the coding potential of each by phylogenetic codon substitution frequency (PhyloCSF) (Lin et al., 2011). Each lncRNA locus exhibited a total PhyloCSF score below zero across all frames, consistent with a complete lack of coding potential (Fig. 3B). Given that some lncRNAs have been shown to exhibit variable subcellular localization with localized functions (Chen, 2016), we assessed the general subcellular expression of these transcripts by Fractionation-Seq. Briefly, we generated a subcellular reference transcriptome of the cytoplasmic and nuclear compartments of 6-8 h and 18-22 h embryos and examined abundance of each of these lncRNA transcripts between these fractions. Intriguingly, the 13 lncRNAs exhibited distinct subcellular localization patterns with varying degrees of nuclear/cytoplasmic restriction (Fig. 3C), ranging from almost exclusively cytoplasmic (e.g. *CR30009* and *cherub*) to almost exclusively nuclear (e.g. *CR45312*) detection, including instances in which location appeared to be temporally regulated (e.g. *CR44978*).

To assess lncRNA abundance relative to other transcripts (noncoding and protein-coding) in the neurogenic cell types, we normalized read counts for each transcript in each sample (TPM, Table S4). The maximum expression score for lncRNAs across cell types (maxTPM) showed that although expression varies among lncRNAs, they are generally not expressed at low levels; rather, lncRNA expression was well within the range of what may be expected for protein-coding genes significantly regulated during neurogenesis (Fig. 3D). That these lncRNAs are bona fide regulated transcripts is further supported by specific splicing, which was observed for several of the neurogenic lncRNAs (Fig. 3E, Fig. S8). Thus, these lncRNAs are unlikely to be merely by-products of spurious transcription; rather, they are subject to regulated expression, RNA processing, and controlled export, which supports a potential role in neurogenesis.

One intriguing example of a lncRNA demonstrating specific expression over the course of early neurogenesis is *CR30009*. This lncRNA showed increased expression in the early intermediate column and in neuroblasts, but was most highly enriched in glial cells during all assayed time windows (Fig. 3A). Furthermore, *CR30009* is spliced and primarily exported to the cytoplasm (Fig. 3C,E), features indicative of specific co- and post-transcriptional regulation. However, *CR30009* has the lowest coding potential out of all tested lncRNAs: its PhyloCSF score per codon (−42.647) was more than three standard deviations below the mean for noncoding regions in *Drosophila* (−18.7±7.2, Fig. 3B). Furthermore, *CR30009* is one of the most highly abundant transcripts in glia – noncoding or protein-coding ($\log_2$ maxTPM=15.75, Fig. 3D, Table S3) – which underscores the potential functional importance of *CR30009* in gliogenesis. Notably, this lncRNA appears to exist predominantly as an unannotated short isoform and exhibits regions of high noncoding sequence conservation among drosophilids within the first exon and at the 3′ end of the transcript (Fig. 3E).

A second example, *CR43283* (also known as *cherub*), exhibited dynamic temporal regulation. Expression of *cherub* was strongly enriched in the earliest neuroblasts at 4-6 h, but enrichment quickly waned in later neuroblasts (6-8 h); however, over time *cherub* became specifically expressed being strongly enriched in differentiated neurons and glia by the end of neurogenesis at 18-

Development (2019) 146, dev175265. doi:10.1242/dev.175265

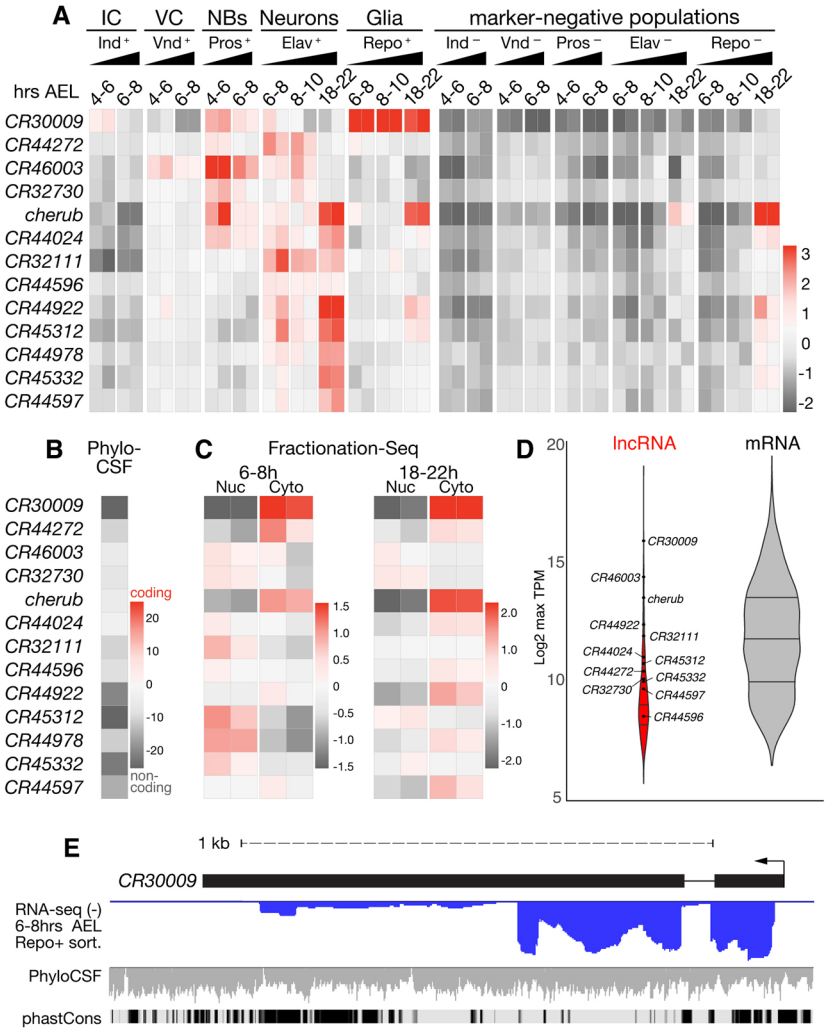


Fig. 3. Neuroglial lncRNAs are highly regulated transcripts. (A) Row mean-centered expression values of lncRNAs in marker-enriched and -depleted samples (scale=log2 ratio of row mean, gene level VST). (B) PhyloCSF scores (ScorePerCodon) for the putative ORF with the highest coding potential within each transcript. Scale is from ~1 s.d. above the mean score of coding regions (very high coding potential) down to ~1 s.d. below the mean of noncoding regions (very low coding potential). (C) Row mean-centered expression profiles in 6-8 h and 18-22 h embryo nuclear and cytoplasmic fractions generated by Fractionation-Seq; values as in A. (D) Violin plot showing distribution of maximum TPM (maxTPM) values for all lncRNAs (red; *n*=325) and mRNAs (gray; *n*=3835) differentially expressed (log_2FC>1.0, *Padj*<0.05) between any marker-positive and marker-negative cell type; lncRNAs presented in A are highlighted. (E) *CR30009* genomic locus showing stranded RNA-seq data from sorted glia at 6-8 h AEL (negative strand; blue), overlay of smoothed PhyloCSF scores of individual codons in each of three frames (horizontal line is 0), and conservation among drosophilids (phastCons).

22 h AEL (Fig. 3A). We note that enriched expression of the lncRNA in Elav- and Repo-negative samples may be caused by *cherub*-positive glia in the neuron-depleted fraction and *cherub*-positive neuroblasts/neurons in the glia-depleted fraction. *cherub* was also specifically localized to the cytoplasm throughout embryogenesis and is clearly spliced, but harbors no coding potential (Fig. 3B,C, Fig. S8D).

CR32730 was first detected in 4-6 h neuroblasts and was moderately enriched at 8-10 h in the neuronal, but not in the glial, population (Fig. 3A). *CR32730* is transcribed antisense to the intron of *CG9650* (Fig. S8C), a putative neurogenic transcription factor that has been implicated in CNS development (McGovern et al., 2003). However, *CR32730* appears to be transcribed independently of *CG9650*, which was expressed at low levels in early neuroblasts according to DIV-MARIS (Fig. S8C), suggesting that their roles could be independent. Fractionation-Seq predicts that *CR32730* is moderately enriched in the nuclear fraction in early and late embryos (Fig. 3C).

Expression of another lncRNA, *CR46003*, was first detected in the ventral column and was most highly enriched in early neuroblasts, but expression persisted in neuroblasts and early neurons (Fig. 3A). *CR46003* was one of the most abundant lncRNAs in our dataset and did not exhibit clear subcellular enrichment in either early or late embryos (Fig. 3C,D). Intriguingly, the transcription start site of *CR46003* is antisense to *CR46004*, which contains a miRNA implicated in behavior (Picao-Osorio et al., 2017) (Fig. S8B).

CR44024 expression was first enriched in early neuroblasts and persisted through neuronal differentiation, and is predicted to be excluded from the intermediate and ventral columns and glia (Fig. 3A). This lncRNA is not predicted to exhibit distinct subcellular localization in early (6-8 h) embryos, but was moderately enriched in the cytoplasm at the end of embryogenesis (18-22 h, Fig. 3C). *CR44024* was also one of the highly expressed lncRNAs in our dataset with expression on a par with protein-coding genes (Fig. 3D). The transcript is intergenic, and appears to be spliced, although not in accordance with its annotated transcript model (Fig. S8E).

In summary, DIV-MARIS predicted spatiotemporal expression of a number of lncRNAs during neurogenesis. Through the application of stringent criteria, we refined this list to a high-confidence selection of noncoding transcripts with diverse predicted expression patterns and properties. To confirm these predictions for several lncRNA candidates, we first visualized their expression in the context of a whole developing embryo.

Neurogenic lncRNAs mark specific neuroglial subsets

To visualize lncRNA expression, we performed multiplex RNA fluorescent *in situ* hybridization (RNA-FISH; Kosman et al., 2004) against the five examples discussed above (*CR30009*,

Development (2019) 146, dev175265. doi:10.1242/dev.175265

cherub, *CR46003*, *CR32730* and *CR44024*) together with neurogenic marker genes. Remarkably, RNA-FISH revealed exquisite spatiotemporal specificity of lncRNA expression for each of the lncRNAs tested.

CR30009 – predicted by DIV-MARIS to be highly enriched in glia – was indeed co-expressed with *repo* as expected in clusters of glial cells as early as stage 9/10 (Fig. 4A,B, Fig. S9). *CR30009* remained co-expressed with most *repo*-expressing cells through stage 13/14 (Fig. 4C,D). However, timing of *CR30009* expression suggests it to be independent of *repo*, indicating that this lncRNA constitutes an earlier marker of the glial lineage than currently known. Although most *repo*-positive cells also expressed *CR30009* in stage 9-12 embryos, the lncRNA was largely expressed in small puncta within other cells in the ventral nerve cord and brain that are likely to be neuroblasts (Fig. 4A,B, Fig. S9). Accordingly, DIV-MARIS predicts *CR30009* expression in 4-6 h and 6-8 h *pros*-positive cells (Fig. 3A, Fig. S10). It is feasible, therefore, that the lncRNA *CR30009* constitutes the earliest neuroblast marker of the glial lineage identified to date, which accumulates into larger, brighter foci during early phases of glial differentiation (Figs S9, S10).

RNA-FISH against the lncRNA *cherub* revealed strong spatiotemporal regulation of *cherub* broadly in accordance with DIV-MARIS, which predicted *cherub* to be strongly and specifically enriched in early (4-6 h) neuroblasts, and late (18-22 h) neurons and glia (Fig. 3A). We observed clear induction of *cherub* expression within six small clusters of cells in the ventral nerve cord during stage 12, each of which also expressed *pros* (Fig. 5A) and to a lesser degree, *elav* (Fig. 5B); both of these observations are in line with *cherub* constituting a neuroblast marker. During stage 13, *cherub* was seen in several additional clusters in the brain (Fig. 5B, Fig. S11A,B). By stage 14-15, *cherub* was very strongly expressed in multiple defined *pros* neuroblast clusters, but appeared to be excluded from mature neurons and glia (Fig. 5C,D, Figs S11C,D, S12), and remained strongly expressed through the remainder of embryogenesis (stage 16/17, Fig. S12B), in line with DIV-MARIS predictions (Fig. 3A).

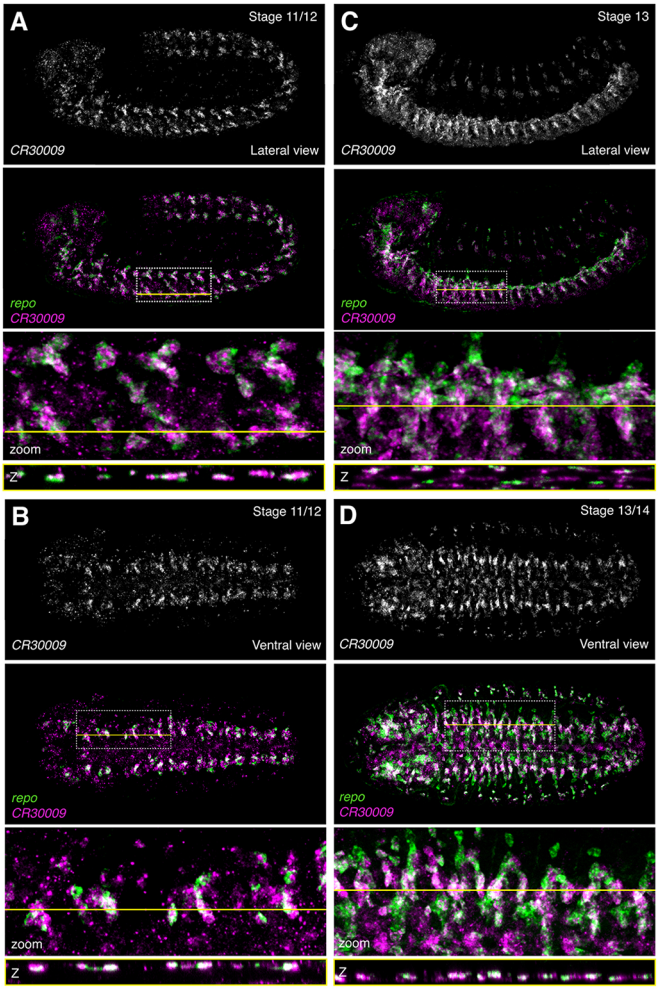


Fig. 4. The lncRNA *CR30009* is expressed in glial subsets. RNA-FISH against *CR30009* and the glial marker *repo*. (A) Lateral view, stage 11/12. (B) Ventral view, stage 11/12. (C) Lateral view, stage 13. (D) Ventral view; stage 13/14. Top: *CR30009* alone; below: merge of *CR30009* (magenta) with *repo* (green). Second from bottom: enlargement of the region of interest (ROI) indicated by the dotted white box. Bottom: Slice through *z*-stack at the level indicated by the yellow line. Embryos are ∼500 µm along the long axis, oriented anterior leftwards and ventral downwards in lateral views.

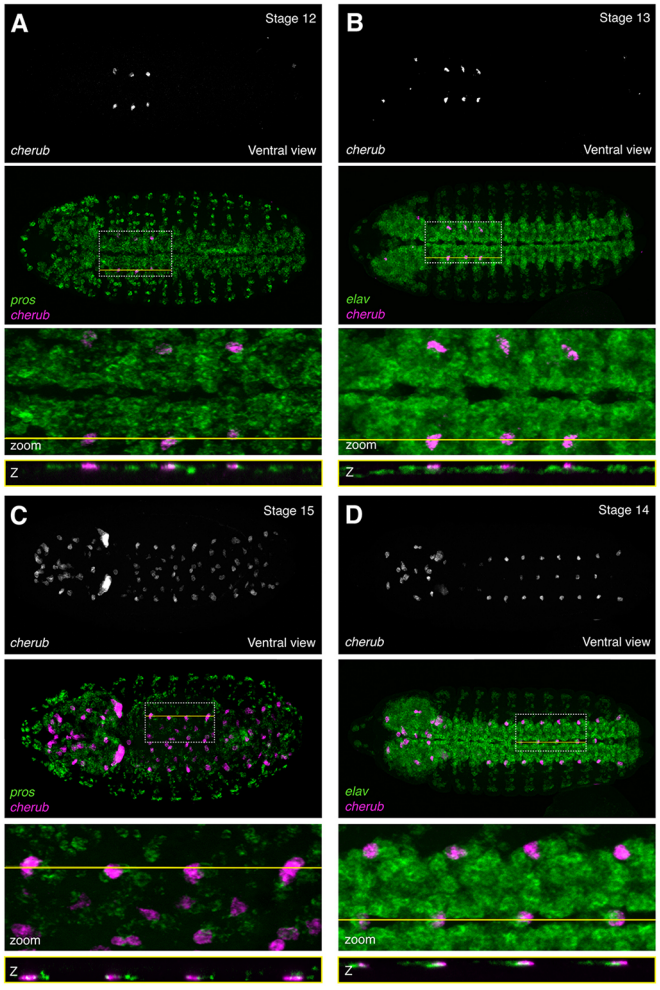


Fig. 5. The lncRNA *cherub* is expressed with strict spatiotemporal specificity primarily in a subset of neuroblasts. RNA-FISH against *cherub*, the neuroblast marker *pros* and the neuronal marker *elav*. Ventral views. (A) *cherub* with *pros*; stage 12. (B) *cherub* with *elav*; stage 13. (C) *cherub* with *pros*; stage 15. (D) *cherub* with *elav*; stage 14. Top: *cherub* alone. Second from top: *cherub* (magenta) overlaid with marker (green). Second from bottom: enlargement of the region of interest (ROI) indicated by the dotted white box. Bottom: Slice through *z*-stack at the level indicated by the yellow line. Embryos are ∼500 µm along the long-axis, oriented anterior leftwards.

Development (2019) 146, dev175265. doi:10.1242/dev.175265

DIV-MARIS predicts similar spatiotemporal expression of *CR46003* and *CR32730* in neuroblasts and neurons (Fig. 3A). Indeed, RNA-FISH revealed very similar patterns of expression of the two lncRNAs. *CR46003* exhibited the earliest expression of all lncRNAs tested here and is was in a small cell cluster already at stage 5-6 (Fig. S13). By stage 9-10, punctate expression of *CR46003* appeared in defined *pros*-expressing clusters along the embryonic ventral midline (Fig. 6A), in agreement with the DIV-MARIS-predicted enrichment in cells of the ventral column and neuroblasts at 4-6 and 6-8 h AEL (Fig. 3A). *CR46003* expression expanded to a greater number of cells within and beyond the ventral nerve cord and brain from stage 11-13, many of which also expressed *pros* (Fig. 6B, Fig. S14A-C) and some expressed *elav* as well (Fig. S15). As predicted by DIV-MARIS, RNA-FISH demonstrated that *CR32730* follows a very similar pattern of expression to *CR46003* from stage 9-10 to stage 13 (Fig. 6, Fig. S14).

Although we were not able to detect the transient *CR44024* expression in early (stage 9-10) *pros*-positive neuroblasts as predicted by DIV-MARIS, we did observe that this lncRNA exhibits highly dynamic temporal regulation. At stage 12, *CR44024* was induced within small *elav*-positive clusters flanking the midline (Fig. S16). Starting at stage 13, *CR44024* was expressed much more broadly, yet was still restricted to subsets of *elav*- and *pros*-expressing cells within the ventral nerve cord and central brain (Fig. 7).

Lastly, we assessed the subcellular localization of individual lncRNAs. For example, Fractionation-Seq (Fig. 3C) predicted *CR30009* and *cherub* to be predominantly cytoplasmic. This is supported for both lncRNAs by high-resolution confocal microscopy, as both transcripts were primarily detected in the cytoplasm (Fig. S17A,B). *CR46003* and *CR32730* both showed a slight bias for nuclear localization by Fractionation-Seq, which was confirmed by microscopy as both lncRNAs were clearly stained within the nucleus, though it should be noted that subnuclear puncta were observed (Fig. S17C,D). Similarly, *CR44024* appeared to be restricted primarily to the nucleus in the ventral nerve cord at stage 14 (Fig. S17E), matching the prediction.

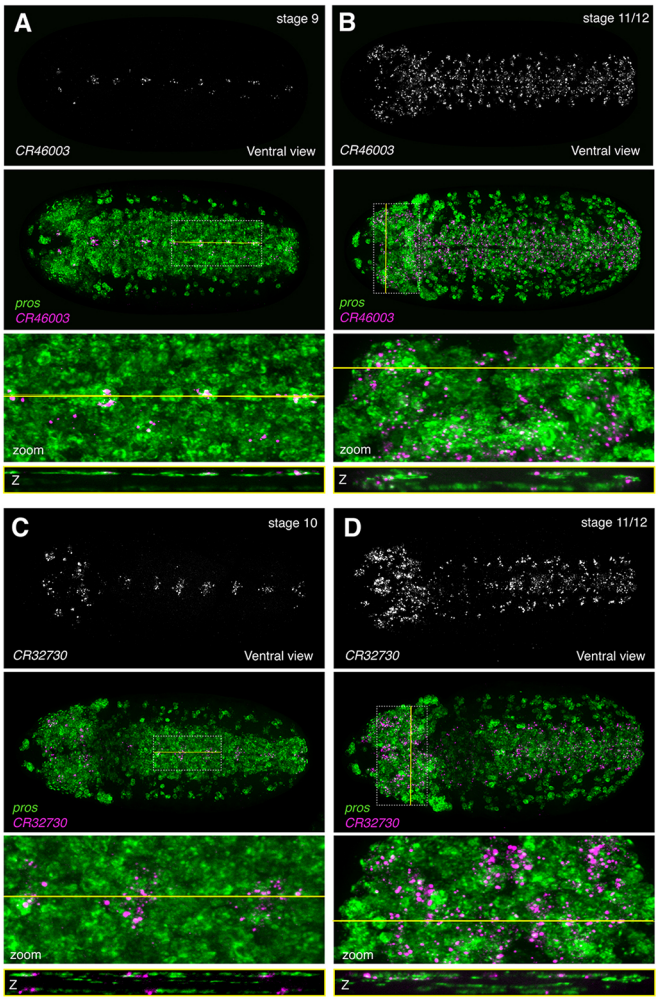


Fig. 6. The lncRNAs *CR46003* and *CR32730* are expressed with similar spatiotemporal specificity in a subset of neuroblasts. RNA-FISH against *CR46003* and *CR32730* together with the neuroblast marker *pros*. Ventral views. (A) *CR46003*; stage 9. (B) *CR46003*; stage 11/12. (C) *CR32730*; stage 10. (D) *CR32730*; stage 11/12. Top: lncRNA alone. Second from top: lncRNA (magenta) overlaid with *pros* (green). Second from bottom: enlargement of the region of interest (ROI) indicated by the dotted white box. Bottom: Slice through *z*-stack at the level indicated by the yellow line. Embryos are ∼500 µm along the long axis, oriented anterior leftwards.

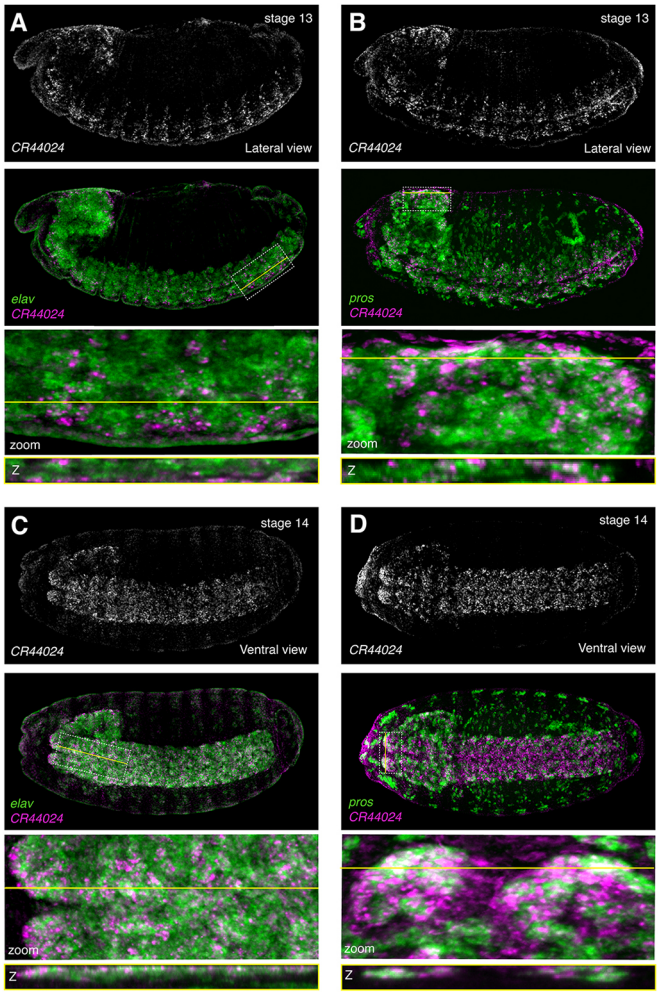


Fig. 7. The lncRNA *CR44024* is expressed later in embryogenesis in neuronal subsets. RNA-FISH against *CR44024*, the neuroblast marker *pros* and the neuronal marker *elav*. (A) *CR44024* and *elav*; lateral view; stage 13. (B) *CR44024* and *pros*; lateral view; stage 13. (C) CR44024 and *elav*; ventral view; stage 14. (D) *CR44024* and *pros*; ventral view; stage 14. Top: *CR44024* alone. Second from top: *CR44024* (magenta) overlaid with marker (green). Second from bottom: enlargement of the region of interest (ROI) indicated by the dotted white box. Bottom: Slice through *z*-stack at the level indicated by the yellow line. Embryos are ∼500 µm along the long axis, oriented anterior leftwards and ventral downwards in lateral views.

Development (2019) 146, dev175265. doi:10.1242/dev.175265

The identification of such complex, yet specific, expression patterns highlights the importance of tissue- and cell type-specific expression analysis. Whole embryo studies, for example, not only lack spatial resolution, but expression signatures – even of highly expressed genes – may be lost if their expression is specific to a small-enough subset of cells. Here, we provide a map for the cell type-specific expression of coding, as well as noncoding, RNAs over the course of embryonic neurogenesis in the developing *Drosophila* embryo. Although hundreds of coding and dozens of lncRNAs are deployed with specific spatial and temporal dynamics, it should be noted that direct imaging of expression within a spatial context can reveal nuances of expression that is beyond the resolution of many cell type-specific genomic approaches.

DISCUSSION

Complex tissues are defined by the intricate interplay of individual cell types that differ in their gene expression programs. Tissue culture has long been an important tool for the genome-wide investigation of cellular responses as it avoids much of the heterogeneity inherent to living tissues. Unfortunately, it is often precisely this heterogeneity and the dynamic contacts between cells and tissues that shape cellular identities and transcriptomic responses. Hence, to determine the gene regulatory programs that drive complex organismal development, it is crucial to (1) preserve the cellular interactions *in vivo*, (2) acquire genome-wide transcriptomic data with spatial and/or cell-type resolution, and (3) assure temporal resolution.

DIV-MARIS to investigate global cell type-specific gene expression dynamics

To investigate the transcriptome dynamics over the course of neurogenesis from primordial to neuronal and glial identities, we developed a method of isolating specific cell types from *Drosophila* embryos with resolution in time and space. DIV-MARIS is widely applicable and can be employed for spatiotemporal transcriptional profiling of basically any cell type of interest in the *Drosophila* embryo and other complex tissues, as long as markers allowing for sorting a cell type of interest are available (i.e. appropriate antibodies or transgenic markers e.g. enhancer-reporter constructs). DIV-MARIS employs chemical cross-linking of the cellular material, thus ensuring that the developmental status quo is preserved, and elaborate sorting strategies based on multiple markers, which could be devised to fine-tune the sub-population selection one wishes to purify (Molyneaux et al., 2015).

Here, we purified fixed cells based on markers of specific neurogenic cell populations in the early *Drosophila* embryo. DIV-MARIS faithfully resolved known expression patterns of neurogenic protein-coding genes, but also identified cell type-specific expression of additional genes with yet unknown neurogenic functions; neuroglial expression was confirmed by *in situ* hybridization for a few dozen mRNAs, but hundreds more are predicted to exhibit spatiotemporal expression over the course of early neurogenesis. This compendium lays the groundwork for a comprehensive understanding of the mechanisms driving early neurogenesis and, given that many of the spatiotemporally expressed genes encode regulatory factors such as transcription factors and signaling molecules, careful examination of their neurogenic roles will be required.

Identification of spatiotemporal lncRNA expression

This study has identified many cell type-specific lncRNAs with potential neurogenic function. We emphasize that this is not yet an exhaustive list of lncRNAs expressed in the nervous system, as our filtering criteria were conservative. Instead, we focused on a high-confidence set of 13 lncRNAs with a variety of expression and transcript characteristics. Given that these noncoding transcripts are (1) temporally expressed in specialized cell types and subtypes of the nervous system, (2) moderately-to-highly abundant and (3) often exhibit hallmarks of RNA processing (such as splicing and nuclear-cytoplasmic shuttling), these lncRNAs appear to be subject to regulated expression rather than being by-products of spurious transcription.

Notably, we did not identify any lncRNAs with expression restricted to the early neuroectodermal columns. According to DIV-MARIS, there is some enrichment of *CR30009* in the intermediate column and *CR46003* in the ventral column (Fig. 3A), but as the respective territorial markers of the ventral and intermediate columns are still detectable in the neuroblast progenitors, this expression may be specific to neuroblasts, in which higher enrichment is observed. High and specific lncRNA expression appears to be a feature of differentiating and differentiated cell types of the nervous system, rather than of primordial territories.

Multiplex RNA-FISH shows that lncRNAs often exhibit a high degree of cell-type specificity. Though co-expression was generally detected with cell type-specific markers, as predicted by DIV-MARIS, we could observe much more nuanced spatiotemporal lncRNA regulation than we could have predicted – the noncoding transcripts investigated here tended to be expressed in highly specific subsets of neurogenic cell types (Figs 4–7). It is therefore feasible that these lncRNAs perform highly specialized functions in subsets of cells contributing to discrete regions of the nervous system.

For example, *CR46003* and *CR32730* are the first lncRNAs that appear to specifically mark midline and midline-proximal structures (Fig. 6, Fig. S14). Given the midline's highly specialized role as a signaling and organizing center (Wheeler et al., 2006; Crews, 2010; Zhou et al., 1995), it is intriguing to speculate that such lncRNAs may help shape the midline fates. Although lncRNAs were enriched in a variety of neurogenic populations, *CR30009* was consistently and highly enriched in *repo*-positive glia and to some degree in *pros*-expressing neuroblasts (Fig. 4, Figs S9, S10). It is feasible that *CR30009* may play a role in the priming of glial fates from the earliest stages of differentiation, possibly mediating the transition from neuroblasts and GMCs to specifically the glial fate. As most glia in the *Drosophila* embryonic CNS originate from the lateral column, it will be of interest to determine whether *CR30009* expression and function are limited to glia of the lateral neurogenic ectoderm, or present in ventral column-derived glia as well.

Are these lncRNAs functional? *cherub* serves as a nice example arguing that several of them likely are. The lncRNA *cherub* was recently identified as a highly upregulated transcript in neuroblast-derived tumors in larvae (Landskron et al., 2018). In larvae, *cherub* is asymmetrically inherited by the self-renewing neuroblast to allow fate progression of the sibling cell and *cherub*'s specific predicted enrichment in embryonic neuroblasts (Fig. 3A) indicates that this lncRNA could exhibit a similar function in the early embryo. However, the precise temporal regulation of *cherub* was surprising, as RNA-FISH identified its presence not in early, but in differentiating and fully differentiated neurons and glia by the end of embryogenesis (Fig. 5, Figs S11, S12).

Intricate spatiotemporal expression regulation is a hallmark of many lncRNAs (Wilk et al., 2016; Karaiskos et al., 2017; Landskron et al., 2018). Various lncRNAs have been demonstrated to play diverse biological roles – nuclear and cytoplasmic – from integral parts of riboprotein complexes, to regulating dosage compensation,

Development (2019) 146, dev175265. doi:10.1242/dev.175265

to affecting genome topology. lncRNA complexity has been reported to be especially pronounced in the nervous system (Briggs et al., 2015; Molyneaux et al., 2015) and even early stages of embryonic neurogliogenesis appear to be no exception. However, the challenge clearly remains to unravel the neurogenic roles of these putative noncoding regulators, and the molecular mechanisms by which they act. This study represents a valuable resource for understanding transcriptome complexity in the emerging nervous system and it lays the basis for further studies into the mechanisms by which noncoding genes, but also hundreds of specifically deployed coding genes, shape nervous system development.

MATERIALS AND METHODS

Fly lines

For details of fly strains and husbandry, see supplementary Materials and Methods.

FACS purification and RNA isolation using DIV-MARIS

Briefly, embryos were dissociated into single-cell suspensions, and cells were fixed in 4% formaldehyde. Fixed cell suspensions were immunostained under RNase-free conditions and FACS-purified using a FACS-AriaII cell sorter (BD Biosciences). Marker-enriched and -depleted cell populations were collected in biological duplicates. FACS-purified cells were subject to cross-link reversal and proteinase K digestion prior to RNA isolation. Additional experimental details for DIV-MARIS are provided in the supplementary Materials and Methods; primary and secondary antibodies used in this study are listed in Table S6.

Nuclear-cytoplasmic fractionation

Cytoplasmic and nuclear extracts were isolated from whole *Drosophila* embryos by detergent-based hypotonic lysis for RNA isolation. Additional experimental details are provided in the supplementary Materials and Methods.

Quantitative RT-PCR (qPCR)

qPCR was performed using standard SYBR Green, with the Bio-Rad CFX96 Touch Real-Time PCR Detection System. Additional information is available in the supplementary Materials and Methods; qPCR primer sequences are listed in Table S7.

Library preparation and RNA-sequencing

All RNA-seq libraries were constructed using the NuGEN Ovation *Drosophila* RNA-Seq System with 10-100 ng total RNA input. Library concentration was quantified using the Qubit dsDNA HS Assay (Thermo Fisher Scientific, Q32854) and quality was determined on a BioAnalyzer using Agilent High Sensitivity DNA Kits (Agilent, 5067-4626). All libraries were sequenced on the Illumina HiSeq4000 at a mean depth of 62.5 million 75 bp paired-end reads per sample. RNA-seq datasets generated for this study are detailed in Tables S10 and S11.

Bioinformatic analysis of RNA-seq data

Sequencing files were demultiplexed using bcl2fastq (v2.19, Illumina), and quality determined using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). A genomic reference index for *Drosophila melanogaster* was constructed with RSEM using the most recent genome build (BDGP release 6) and transcriptome annotation (Release 6.15) obtained from Flybase (www.flybase.org). Annotations used for lncRNAs have been described by Young et al. (2012). Paired-end reads were pseudo-aligned to the RSEM reference index using Salmon (Release 0.8.1) using the following parameters: $ salmon quant –libType ISF –seqBias –gcBias –posBias -p 8 –numBootstraps 100.

Gene-level counts were prepared for differential expression analysis with tximport as part of the Bioconductor package (Release 3.5). Feature length-scaled TPM values were calculated with tximport using the following command: >tximport(files, type="salmon", countsFromAbundance= "lengthScaledTPM", tx2gene=tx2gene).

Given the cell-type heterogeneity between samples in this dataset, we used normalized counts instead of TPM or FPKM for more accurate inter-sample comparisons of gene abundance. We normalized gene-level counts via variance stabilizing transformation (Table S1). Variance-stabilized transformed counts, principal component analysis, and differential expression were calculated using DESeq2 (Love et al., 2014) as part of the Bioconductor package (Release 3.5), using default parameters.

PhyloCSF

PhyloCSF uses substitutions and codon frequencies in a genome alignment of 23 drosophilid species to distinguish the evolutionary signature of selection for protein-coding function (Lin et al., 2011). For each transcript, PhyloCSF generates a score for the putative open reading frame (ORF) with highest coding potential; transcripts with positive scores are more likely to be protein coding. The candidate ORFs, their PhyloCSF scores, and other related information are included in Table S5.

Briefly, local alignments used for PhyloCSF were extracted from the 23-drosophilid subset of the 27-way MULTIZ insect whole-genome alignments (Blanchette et al., 2004), downloaded from UCSC: http://hgdownload.soe.ucsc.edu/goldenPath/dm6/multiz27way/ (Tyner et al., 2017). PhyloCSF scores were computed using the 23flies parameters with the options '-f3 --orf=ATGStop --allScores --bls', which computes the score of every ORF within the transcript that begins with ATG, is followed by a stop codon, and is at least the default length of 25 codons. Because CR44272 has no putative ORFs that long, we used '--minCodons=19' to lower the threshold for that gene to the length of its longest putative ORF. We then selected the ORF in each transcript having the highest PhyloCSF score. The reported 'ScorePerCodon' is the PhyloCSF score divided by the number of codons in the putative ORF. To identify potential cases in which one of the transcripts under consideration contains part of a coding ORF but the complete ORF is in an unidentified overlapping transcript, we also ran PhyloCSF using the --orf=StopStop3 option, with --minCodons=10, which looks for ORF fragments ending in a stop codon. However, that did not identify any plausible partial coding ORFs. The PhyloCSF track images in Fig. 3 and Fig. S6 are overlays of the 'Smoothed PhyloCSF' tracks in all three frames on the appropriate strand, from the PhyloCSF track hub in the UCSC genome browser, documented at: https://data.broadinstitute.org/compbio1/PhyloCSFtracks/trackHub/hub.DOC.html.

Generation of coverage plots

The strand-specific and paired-end RNA-seq reads were mapped to the *Drosophila melanogaster* reference genome dm6 with the splicing-aware mapper STAR v2.5.3a (Dobin et al., 2013) using default parameters and a *Drosophila*-specific adjustment for maximum intron length and mate distance of 50 kb. The resulting BAM files were filtered to include only uniquely mapping read pairs and then converted into strand-specific genome coverage tracks in BigWig format for visualization in the UCSC genome browser (Kent et al., 2010; Raney et al., 2014) using the stranded-coverage (https://github.com/pmenzel/stranded-coverage) and wigToBigWig from the UCSC genome browser tools.

Immunohistochemistry and FISH

Immunohistochemistry and RNA-FISH were performed as previously described (Kosman et al., 2004; Karaiskos et al., 2017). Primary and secondary antibodies used in this study are listed in Table S6. The procedure for probe synthesis is detailed in supplementary Materials and Methods, and RNA probes are listed in Tables S8 and S9.

Microscopy

Confocal stacks were imaged at BIMSB/MDC using a Leica SP8 equipped with 405 nm laser diode, white light laser, and hybrid detectors, with a 20× glycerol objective. For each field of view, 65-85 slices were acquired using ∼AU=1 pinholes and taking care not to saturate signal. Appropriate slices were maximum intensity projected. Imaging work at the BioFrontiers Institute's Advanced Light Microscopy Core was carried out on either a Nikon A1R laser scanning confocal microscope (NIST-CU Cooperative Agreement 70NANB15H226) or on a Nikon Ti-E spinning

disc confocal microscope (BioFrontiers Institute, Howard Hughes Medical Institute).

Acknowledgements
We are grateful to John L. Rinn for insightful comments and extensive discussions, to Petar Glažar and Panagiotis Papavasileiou for considerable RNA-seq troubleshooting and preliminary data analysis, Andrew Woehler and Joe Dragavon for assistance with confocal imaging, David Schechner for providing the cell fractionation protocol, Sara Ugowski and Claudia Kipar for maintaining the fly facility at the BIMSB/MDC, and Sabrina Krüger and Agnieszka Klawiter for help generating transgenic fly lines. Moreover, we are grateful to all members of the Zinzen Laboratory for in-depth discussions and technical assistance on experiments presented in this study. We also acknowledge the BIMSB Genomics Platform and the Systems Biology Imaging Platform at the MDC.

Competing interests
The authors declare no competing or financial interests.

Author contributions
Conceptualization: A.L.M., R.P.Z.; Methodology: A.L.M., P.W., I.J., R.A.I., M.K., R.P.Z.; Validation: A.L.M., S.W., R.L.P.A., R.P.Z.; Formal analysis: A.L.M., I.J., P.M., C.J.S., M.K., R.P.Z.; Investigation: A.L.M., S.W., R.P.Z.; Data curation: A.L.M., R.A.I., I.M.M., R.P.Z.; Writing - original draft: A.L.M., R.P.Z.; Writing - review & editing: A.L.M., P.W., S.W., I.J., P.M., C.J.S., R.L.P.A., R.A.I., I.M.M., M.K., R.P.Z.; Visualization: A.L.M., P.W., S.W., R.L.P.A., R.P.Z.; Supervision: R.P.Z.; Project administration: R.P.Z.; Funding acquisition: R.P.Z.

Funding
A.L.M. and P.W. were supported by a generous grant from the Deutsche Forschungsgemeinschaft (DFG; Priority Program SPP1738). I.J. was supported by the National Institutes of Health [HG004037] and GENCODE Wellcome Trust [U41 HG007234]. Deposited in PMC for immediate release.

Data availability
RNA-seq data have been deposited in NCBI Gene Expression Omnibus under accession number GSE106095.

Supplementary information
Supplementary information available online at
http://dev.biologists.org/lookup/doi/10.1242/dev.175265.supplemental

References

Beckervordersandforth, R. M., Rickert, C., Altenhein, B. and Technau, G. M. (2008). Subtypes of glial cells in the Drosophila embryonic ventral nerve cord as related to lineage and gene expression. *Mech. Dev.* **125**, 542-557.
Berger, C., Renner, S., Lüer, K. and Technau, G. M. (2007). The commonly used marker ELAV is transiently expressed in neuroblasts and glial cells in the Drosophila embryonic CNS. *Dev. Dyn.* **236**, 3562-3568.
Bier, E. and De Robertis, E. M. (2015). Embryo development. BMP gradients: a paradigm for morphogen-mediated developmental patterning. *Science* **348**, aaa5838-aaa5838.
Blanchette, M., Kent, W. J., Riemer, C., Elnitski, L., Smit, A. F., Roskin, K. M., Baertsch, R., Rosenbloom, K. and Clawson, H. (2004). Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708-715.
Briggs, J. A., Wolvetang, E. J., Mattick, J. S., Rinn, J. L. and Barry, G. (2015). Mechanisms of long non-coding RNAs in mammalian nervous system development, plasticity, disease, and evolution. *Neuron* **88**, 861-877.
Broadus, J., Skeath, J. B., Spana, E. P., Bossing, T., Technau, G. and Doe, C. Q. (1995). New neuroblast markers and the origin of the aCC/pCC neurons in the Drosophila central nervous system. *Mech. Dev.* **53**, 393-402.
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.
Campos-Ortega, J. A. (1995). Genetic mechanisms of early neurogenesis in Drosophila melanogaster. *Mol. Neurobiol.* **10**, 75-89.
Chen, B., Zhang, Y., Zhang, X., Jia, S., Chen, S. and Kang, L. (2016). Genome-wide identification and developmental expression profiling of long noncoding RNAs during Drosophila metamorphosis. *Sci. Rep.* **6**, 23330.
Chen, L.-L. (2016). Linking long noncoding RNA localization and function. *Trends Biochem. Sci.* **41**, 761-772.
Cowden, J. and Levine, M. (2003). Ventral dominance governs sequential patterns of gene expression across the dorsal-ventral axis of the neuroectoderm in the Drosophila embryo. *Dev. Biol.* **262**, 335-349.
Crews, S. T. (2010). Axon-glial interactions at the Drosophila CNS midline. *Cell Adh. Migr.* **4**, 1-5.
Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21.
Doe, C. Q. (1992). Molecular markers for identified neuroblasts and ganglion mother cells in the Drosophila central nervous system. *Development* **116**, 855-863.
Doe, C. Q. (2017). Temporal patterning in the Drosophila CNS. *Annu. Rev. Cell Dev. Biol.* **33**, 219-240.
Goff, L. A., Groff, A. F., Sauvageau, M., Trayes-Gibson, Z., Sanchez-Gomez, D. B., Morse, M., Martin, R. D., Elcavage, L. E., Liapis, S. C., Gonzalez-Celeiro, M. et al. (2015). Spatiotemporal expression and transcriptional perturbations by long noncoding RNAs in the mouse brain. *Proc. Natl Acad. Sci. USA* **112**, 6855-6862.
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.
Hammonds, A. S., Bristow, C. A., Fisher, W. W., Weiszmann, R., Wu, S., Hartenstein, V., Kellis, M., Yu, B., Frise, E. and Celniker, S. E. (2013). Spatial expression of transcription factors in Drosophila embryonic organ development. *Genome Biol.* **14**, R140.
Heckscher, E. S., Long, F., Layden, M. J., Chuang, C.-H., Manning, L., Richart, J., Pearson, J. C., Crews, S. T., Peng, H., Myers, E. et al. (2014). Atlas-builder software and the eNeuro atlas: resources for developmental biology and neuroscience. *Development* **141**, 2524-2532.
Homem, C. C. F. and Knoblich, J. A. (2012). Drosophila neuroblasts: a model for stem cell biology. *Development* **139**, 4297-4310.
Hrvatin, S., Deng, F., O'Donnell, C. W., Gifford, D. K. and Melton, D. A. (2014). Maris: method for Analyzing RNA following intracellular sorting K. Aalto-Setala, ed. *PLoS ONE* **9**, e89459–6.
Inagaki, S., Numata, K., Kondo, T., Tomita, M., Yasuda, K., Kanai, A. and Kageyama, Y. (2005). Identification and expression analysis of putative mRNA-like non-coding RNA in Drosophila. *Genes Cells* **10**, 1163-1173.
Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N. and Zinzen, R. P. (2017). The Drosophila embryo at single-cell transcriptome resolution. *Science* **358**, 194-199.
Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204-2207.
Kosman, D., Mizutani, C. M., Lemons, D., Cox, W. G., McGinnis, W. and Bier, E. (2004). Multiplex detection of RNA expression in Drosophila embryos. *Science* **305**, 846-846.
Kunisch, M., Haenlin, M. and Campos-Ortega, J. A. (1994). Lateral inhibition mediated by the Drosophila neurogenic gene delta is enhanced by proneural proteins. *Proc. Natl Acad. Sci. USA* **91**, 10139-10143.
Lai, E. C. (2004). Notch signaling: control of cell communication and cell fate. *Development* **131**, 965-973.
Landgraf, M., Bossing, T., Technau, G. M. and Bate, M. (1997). The origin, location, and projections of the embryonic abdominal motorneurons of Drosophila. *J. Neurosci.* **17**, 9642-9655.
Landskron, L., Steinmann, V., Bonnay, F., Burkard, T. R., Steinmann, J., Reichardt, I., Harzer, H., Laurenson, A.-S., Reichert, H. and Knoblich, J. A. (2018). The asymmetrically segregating lncRNA cherub is required for transforming stem cells into malignant cells. *eLife* **7**, R106.
Lécuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T. R., Tomancak, P. and Krause, H. M. (2007). Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* **131**, 174-187.
Li, M. and Liu, L. (2015). Neural functions of long noncoding RNAs in Drosophila. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **201**, 921-926.
Lin, M. F., Jungreis, I. and Kellis, M. (2011). PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27**, i275-i282.
Liu, S. J., Nowakowski, T. J., Pollen, A. A., Lui, J. H., Horlbeck, M. A., Attenello, F. J., He, D., Weissman, J. S., Kriegstein, A. R., Diaz, A. A. et al. (2016). Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biol.* **17**, 67.
Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.
Mcgovern, V. L., Pacak, C. A., Sewell, S. T., Turski, M. L. and Seeger, M. A. (2003). A targeted gain of function screen in the embryonic CNS of Drosophila. *Mech. Dev.* **120**, 1193-1207.
Molyneaux, B. J., Goff, L. A., Brettler, A. C., Chen, H.-H., Brown, J. R., Hrvatin, S., Rinn, J. L. and Arlotta, P. (2015). DeCoN: genome-wide analysis of in vivo transcriptional dynamics during pyramidal neuron fate selection in neocortex. *Neuron* **85**, 275-288.
Picao-Osorio, J., Lago-Baldaia, I., Patraquim, P. and Alonso, C. R. (2017). Pervasive behavioral effects of microRNA regulation in Drosophila. *Genetics* **206**, 1535-1548.
Raney, B. J., Dreszer, T. R., Barber, G. P., Clawson, H., Fujita, P. A., Wang, T., Nguyen, N., Paten, B., Zweig, A. S., Karolchik, D. et al. (2014). Track data hubs

enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* **30**, 1003-1005.

Rickert, C., Kunz, T., Harris, K.-L., Whitington, P. M. and Technau, G. M. (2011). Morphological characterization of the entire interneuron population reveals principles of neuromere organization in the ventral nerve cord of Drosophila. *J. Neurosci.* **31**, 15870-15883.

Rogers, W. A., Goyal, Y., Yamaya, K., Shvartsman, S. Y. and Levine, M. S. (2017). Uncoupling neurogenic gene networks in the Drosophila embryo. *Genes Dev.* **31**, 634-638.

Sandler, J. E. and Stathopoulos, A. (2016). Stepwise progression of embryonic patterning. *Trends Genet.* **32**, 432-443.

Santiago, C., Labrador, J.-P. and Bashaw, G. J. (2014). The homeodomain transcription factor Hb9 controls axon guidance in Drosophila through the regulation of Robo receptors. *Cell Rep.* **7**, 153-165.

Sauvageau, M., Goff, L. A., Lodato, S., Bonev, B., Groff, A. F., Gerhardinger, C., Sanchez-Gomez, D. B., Hacisuleyman, E., Li, E., Spence, M. et al. (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife* **2**, e01749.

Skeath, J. B. and Thor, S. (2003). Genetic control of Drosophila nerve cord development. *Curr. Opin. Neurobiol.* **13**, 8-15.

Skeath, J. B., Panganiban, G. F. and Carroll, S. B. (1994). The ventral nervous system defective gene controls proneural gene expression at two distinct steps during neuroblast formation in Drosophila. *Development* **120**, 1517-1524.

Sousa-Nunes, R., Cheng, L. Y. and Gould, A. P. (2010). Regulating neural proliferation in the Drosophila CNS. *Curr. Opin. Neurobiol.* **20**, 50-57.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S. Q., Lewis, S. E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S. E. et al. (2002). Systematic determination of patterns of gene expression during Drosophila embryogenesis. *Genome Biol.* **3**, RESEARCH0088.

Tomancak, P., Berman, B. P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S. E. and Rubin, G. M. (2007). Global analysis of patterns of gene expression during Drosophila embryogenesis. *Genome Biol.* **8**, R145.

Tyner, C., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., Eisenhart, C., Fischer, C. M., Gibson, D., Gonzalez, J. N., Guruvadoo, L. et al. (2017). The UCSC genome browser database: 2017 update. *Nucleic Acids Res.* **45**, D626-D634.

Von Ohlen, T. and Doe, C. Q. (2000). Convergence of dorsal, dpp, and egfr signaling pathways subdivides the drosophila neuroectoderm into three dorsal-ventral columns. *Dev. Biol.* **224**, 362-372.

Weiss, J. B., Von Ohlen, T., Mellerick, D. M., Dressler, G., Doe, C. Q. and Scott, M. P. (1998). Dorsoventral patterning in the Drosophila central nervous system: the intermediate neuroblasts defective homeobox gene specifies intermediate column identity. *Genes Dev.* **12**, 3591-3602.

Wen, K., Yang, L., Xiong, T., Di, C., Ma, D., Wu, M., Xue, Z., Zhang, X., Long, L., Zhang, W. et al. (2016). Critical roles of long noncoding RNAs in Drosophila spermatogenesis. *Genome Res.* **26**, 1233-1244.

Wheeler, S. R., Kearney, J. B., Guardiola, A. R. and Crews, S. T. (2006). Single-cell mapping of neural and glial gene expression in the developing Drosophila CNS midline cells. *Dev. Biol.* **294**, 509-524.

Wheeler, S. R., Stagg, S. B. and Crews, S. T. (2009). MidExDB: a database of Drosophila CNS midline cell gene expression. *BMC Dev. Biol.* **9**, 56.

Wilk, R., Hu, J., Blotsky, D. and Krause, H. M. (2016). Diverse and pervasive subcellular distributions for both coding and long noncoding RNAs. *Genes Dev.* **30**, 594-609.

Young, R. S., Marques, A. C., Tibbit, C., Haerty, W., Bassett, A. R., Liu, J.-L. and Ponting, C. P. (2012). Identification and Properties of 1,119 Candidate LincRNA Loci in the Drosophila melanogaster Genome. *Genome Biol. Evol.* **4**, 427-442.

Zhou, L., Hashimi, H., Schwartz, L. M. and Nambu, J. R. (1995). Programmed cell death in the Drosophila central nervous system midline. *Curr. Biol.* **5**, 784-790.

Supplementary Materials and Methods

Fly strains and husbandry

Drosophila melanogaster lines used in this study include

yellow-white	$y^1\ w^{1118}$	(source: BSC# 6598)
IC-GFP	$y^1\ w^{1118}$; ; *P{3xind_1.4-GFP}*	(this study)
IC-dsRed	$y^1\ w^{1118}$; *P{3xind_1.4-dsRed}*	(this study)
VC-dsRed	$y^1\ w^{1118}$; ; *P{2xvnd_743-dsRED}*	(Karaiskos et al. 2017)
vnd-lexA/VC:FNLDD	$y^1\ w^{1118}$; *M{3xvnd-lexA:Cit}*$^{attP\_ZH51C}$; *P{vnd:FNLDD}*	(Krueger et al., unpublished)
rho-lexA/VC:FNLDD	$y^1\ w^{1118}$; *M{3xrho-lexA:Cit}*$^{attP\_ZH51C}$; *P{vnd:FNLDD}*	(Krueger et al., unpublished)

The *IC-GFP* and *IC-dsRed* lines were created by standard P-element transgenesis (Rubin & Spradling 1982) in a $y^1\ w^{1118}$ background. The *ind* enhancer (*ind_1.4*) (Markstein et al. 2004) was cloned using primers (ngctagcgtcgacGCTTCAAAGCTCCGGGAAACG & nctcgagTCTGGGCCTTCGGTCCGAAAATG) flanked with NheI and SalI restriction sites (F primer) and with XhoI (R primer). PCR product was T/A cloned into pCRII-Duo and concatemerized using the compatibly cohesive sites XhoI and SalI. 3x *ind_1.4* constructs were directionally subcloned into the P-element vectors pH-Stinger or pRed-HStinger (Barolo et al. 2004) using NheI and XhoI.

Fly stocks were maintained at 25°C, ~60% relative humidity on standard fly food with 12 hr light/dark cycles according to standard procedures. 2-hr embryo collections were done on apple juice agar plates with yeast paste after 3 x 1hr pre-lays in the morning, embryos were then aged for an appropriate amount of time at 25°C and ~60% relative humidity before dechorionation.

DIV-SortSeq

This protocol incorporates portions from MARIS (Hrvatin et al. 2014). All steps after dechorionation were performed on ice, using ice-cold DEPC-treated solutions. Primary and secondary antibodies used in this study are listed in Table S1. RNase-free BSA was obtained from Gemini Bioproducts and is critical for isolation of high-quality RNA.

Embryos were collected, dechorionated, and dissociated into single-cell suspension in 15ml PBS, pH 7.5 using 8-12 strokes with the loose pestle of a glass dounce homogenizer. Large debris was removed via filtration with 2 x 90°-rotated sheets of Miracloth into a 15ml conical vial, followed by centrifugation at 40x*g*, 4°C, 3 min. The supernatant was transferred to a 15ml conical vial, cells were pelleted at 1000x*g*, 4°C, 3 min, washed with 1ml PBS, and fixed with 4% formaldehyde at 4°C, 15min. Cross-linking was stopped with Quench Buffer (2.5M Glycine in PBS). Fixed cells were washed twice with 1ml ice-cold PBS, and stored at 4°C overnight in 1ml RNA*later*™ (Thermo, AM7020). Cells were rehydrated via dilution in 9ml ice-cold PBS, centrifugation at 1000 x*g*, 10min, and washed twice in 1ml ice-cold PBS. Fixed cell suspensions were immunostained under RNase-free conditions with primary antibodies with agitation at 4 °C, 1.5-2h in 250µl Stain Buffer (1% BSA (w/v), 0.1% saponin, 1:200 RNase inhibitor in RNase-free

PBS). After washing 3x 5 min in 1ml Wash Buffer (0.2% BSA, 0.1% saponin in PBS), cells were incubated with conjugated secondary antibodies with agitation at 4°C, 45 min-1h in 250µl Stain Buffer. After washing 3x 5 min in 1ml Wash Buffer, cells were resuspended in 1ml Sort Buffer (0.5% BSA, 2mM EDTA, 1:500 RNase inhibitor in PBS) and filtered with a 70µm cell strainer. Filtered cells were subjected to FACS purification on the FACSAria II (BD Biosciences) and sorted cells collected in Collection Buffer (2% BSA, 1:100 RNase inhibitor in PBS).

FACS-purified cells were pelleted and resuspended in 200µl Digestion Buffer (5M NaCl, 1M Tris-HCl pH 8.0, 200mM EDTA, 10% SDS, 3.2U Proteinase K, 1:100 RNase inhibitor) and incubated at 50°C, 15min (Proteinase K digestion) followed by 80°C, 15min (reversal of formaldehyde cross-links). Samples were transferred to ice and resuspended in 600µl TRIzol™ LS Reagent (Thermo, 10296028). RNA was isolated using the DirectZol™ RNA MicroPrep Kit (Zymo Research, R2060) according to the manufacturer's instructions. RNA concentration was measured using the Qubit™ RNA HS Assay kit (Thermo, Q32852) and RNA quality determined using the Agilent RNA 6000 Pico Kit (Agilent, 5067-1513). All RNA-seq libraries were constructed using the NuGEN Ovation *Drosophila* RNA-Seq System with 10 ng – 100 ng total RNA input. Library concentration was quantified using the Qubit™ dsDNA HS Assay (Thermo, Q32854) and quality was determined on a BioAnalyzer™ using Agilent High Sensitivity DNA Kits (Agilent, 5067-4626). All libraries were sequenced on the Illumina HiSeq4000 at a mean depth of 62.5 million 75bp paired-end reads per sample. RNA-seq datasets generated for this study are detailed in Tables S5 and S6. A detailed, step-wise protocol is available upon request.

Nuclear-cytoplasmic fractionation

The cell fractionation procedure incorporates portions from MARIS (Hrvatin et al. 2014). All steps were performed on ice, using ice-cold DEPC-treated solutions, and all centrifugation steps were performed at 4˚C. Embryos were processed to single-cell suspension as described above for *DIV*-SortSeq, then pelleted and resuspended in Cyto Extract Buffer (20mM Tris pH 7.6, 0.1mM EDTA, 2mM $MgCl_2$). After hypotonic swelling, cells were gently lysed by addition of 0.6% CHAPS for isolation of the cytoplasmic fraction. Nuclei were pelleted at 500x*g* for 5min, and the supernatant retained, and an appropriate volume of TRIzol™ LS Reagent was added (cytoplasmic fraction).

Nuclei were washed with Nuclei Wash Buffer (20mM Tris pH 7.6, 0.1mM EDTA, 2mM $MgCl_2$, 0.6% CHAPS), resuspended in Nuclei Resuspension Buffer (10mM Tris, pH 7.6, 150mM NaCl, 0.15% NP-40) and pelleted at 12,000x*g*, 10min in Sucrose Buffer (10mM Tris, pH 7.5, 150mM NaCl, 24% sucrose). After washing (1mM EDTA in PBS), nuclei pellet was resuspended in TRIzol™ Reagent. RNA was isolated and concentration and quality determined as described for above for *DIV*-SortSeq.

Quantitative RT-PCR (qPCR)

50ng of total RNA was reverse-transcribed with the QuantiTect Reverse Transcription kit (Qiagen, 205310) in a total volume of 20µl, according to the manufacturer's instructions. For each gene, 0.2µl

cDNA was used for input into qPCR using SensiFAST™ SYBR® No-ROX Kit (Bioline, 98020) and 5µM forward and reverse primers in a total volume of 20µl. qPCR primer sequences are listed in Table S2. qPCR thermal cycling and fluorescent data acquisition was performed using the BioRad CFX96 Touch™ Real-Time PCR Detection System. Expression fold changes were calculated via the $\Delta\Delta C_T$ method (Vandesompele et al. 2002; Schmittgen & Livak 2008), normalized to the mean C_T of two reference genes: *α-tubulin* and *actin 42A*.

RNA probe design and synthesis for RNA in situ hybridization.

Where available, cDNA clones were obtained from the *Drosophila* Gene Collection or *Drosophila* Genomics Resource Center (Stapleton et al. 2002), detailed in Table S3. Constructs were linearized via restriction digestion, and subjected to *in vitro* transcription using appropriate RNA Polymerases (Roche), using ribonucleotide mixtures containing dUTP-DIG, -FITC, or -Biotin (Roche).

Where cDNA clones were not available, PCR primers were designed to amplify a region within the transcribed locus from genomic DNA. Primer sequences are detailed in Table S4. A T7 promoter sequence appended to the reverse primer allowed *in vitro* transcription directly from the PCR product. After template digest by DNaseI, RNA probes were sheared at 65°C for 3-20 min in carbonation buffer (120mM Na_2CO_3, 80mM $NaHCO_3$, pH 10.2), length of carbonation depended on probe length. RNA probes were precipitated at -20°C overnight and resuspended in Hyb-A Buffer (50% formamide, 5X SSC, 0.1% Tween-20).

Supplementary Tables

Table S1. VST counts by gene across datasets

Click here to Download Table S1

Table S2. Neurogenic coding genes

Click here to Download Table S2

Table S3. Enriched lncRNAs

Click here to Download Table S3

Table S4. Length-scaled TPMs

Click here to Download Table S4

Table S5. PhyloCSF analysis

Click here to Download Table S5

Column interpretation for Table S5: PhyloCSF

Intervals	Intervals of the input transcript
Strand	Strand
TranscriptName	Name from the input bed file
TrUCSCview	Link to show the entire transcript in UCSC browser
ORFintervals	Intervals of the ORF (not including the stop codon)
ORFstart	0-based transcript coordinate of first base of ORF
ORFend	0-based transcript coordinate of last base of ORF
NumCodons	Number of codons in the ORF
PhyloCSF	Raw PhyloCSF score of the ORF
RelBL	Fraction of branch length of the phylogenetic tree spanned by species present in the alignment of this ORF
ScorePerCodon	PhyloCSF divided by NumCodons
PhyloCSFPsi	Length adjusted score, a log likelihood, in decibans
Pval	Probability a region of this length, none-of-which has ever been coding, has this score or higher
CorrectedPvalTr	p-val with Holm-Bonferroni correction for number of ORFs in this transcript
CorrectedPvalAll	p-val with Holm-Bonferroni correction for total number of ORFs
FDR	Benjamini & Hochberg false discovery rate.
LocalFDR	Local FDR (Efron et al. 2001).
AntiScorePerCodon	Score on opposite strand in frame that shares 3rd codon position
ScoreDiff	ScorePerCodon - AntiScorePerCodon
GC	GC content of ORF
CpGratio	Number of CpGs in ORF divided by the expected number based on C and G content
CodAlignView	Link to view ORF alignment in CodAlignView, the Codon Alignment Viewer, with 10-codon context on each side
OrfUCSCview	Link to show the ORF in UCSC browser

Table S6: Primary and secondary antibodies used in this study

Target	*Supplier*	*Catalog #*	*Application*	*Antibody Type*	*Conjugate*	*Dilution*
RFP	Thermo	710530	*Primary*	Rabbit polyclonal	N/A	1:500
GFP	Thermo	G10362	*Primary*	Rabbit polyclonal	N/A	1:500
Pros	DSHB	MR1A	*Primary*	Mouse monoclonal	N/A	1:20
Elav	DSHB	9F8A9	*Primary*	Mouse monoclonal	N/A	1:500
Repo	DSHB	8D12	*Primary*	Mouse monoclonal	N/A	1:20
Rabbit IgG	Thermo	A21428	*Secondary*	Goat polyclonal	*Alexa Fluor 555*	1:500
Mouse IgG	Thermo	A32727	*Secondary*	Goat polyclonal	*Alexa Fluor 555*	1:500

Table S7: qPCR primer sequences

Target	*Forward primer*	*Reverse primer*	*Size (bp)*
α-Tubulin	TGTCGCGTGTGAAACACTTC	AGCAGGCGTTTCCAATCTG	585
Actin 42A	GCGTCGGTCAATTCAATCTT	AAGCTGCAACCTCTTCGTCA	292
Prospero	CGGCATGGCTCCTACTTCTT	TAGCGCACCCAGAAGAACAT	78
Worniu	ATGGATAAACTCAAGTACAGCCG	AAGTCCACTGGTCCTTCATCA	107
Elav	ACGCTCCTGCCACAGAAAAA	CGTCGCCGTATTTCGCTC	211
Lim3	GATGGAGGATCGTAAGCTGATCT	GTAGGCCGTTTTCAGGGTCTC	154
Repo	CTCCGCCAAGTAGTTCCTCC	AGGCAGTAAAGGTGGTTCTCG	216
Gcm	ACAAGGCCAGAAGGAAGCAG	CAAGCCTGGATTTCCAAGCGA	76

Table S8: List of commercially-available cDNA clones for RNA probe synthesis

RNA target	*DGRC Reference #*
ind	RT01026
vnd	PCSP6029
pros	LD37627
elav	LD33076-IR
repo	GH05443-dg
CR30009	RE30084
CR32730	RE54940
CR32111	RE52337

Table S9: PCR primer sequences for RNA probe synthesis

Target	*Forward primer*	*T7 promoter + Reverse primer*	*Size (bp)*
CR46003	TGTGTCGCACAGGATGTGT	TAATACGACTCACTATAGGTGCTGGCGGGGAAATTATGT	908
cherub	CGAGGAACCTTCGGTGCATA	TAATACGACTCACTATAGGGCTTGGGTGATTTCGAGGGA	1511
CR44024	GTGTCGTGTCGGGTAAGTGT	TAATACGACTCACTATAGGAAGTGGCCTGTCTCAGAACG	1268
CR32111	GTATGCGCTCGAACTCGGTAA	TAATACGACTCACTATAGGGCCGGCATGAGCAAACACAAA	1232

Table S10: Summary of RNA-seq datasets – DIV-SortSeq

Sample name	Fly line	Target protein	Cell Type	Enriched/depleted	Time point	Replicate	# reads (M)
4-6h_Ind-neg_1	*IC*-dsRed	dsRed	Intermediate column	Depleted	4-6h	1	89.9
4-6h_Ind-neg_2	*IC*-dsRed	dsRed	Intermediate column	Depleted	4-6h	2	140.1
4-6h_Ind-pos_1	*IC*-dsRed	dsRed	Intermediate column	Enriched	4-6h	1	76.2
4-6h_Ind-pos_2	*IC*-dsRed	dsRed	Intermediate column	Enriched	4-6h	2	81.6
4-6h_Vnd-neg_1	*VC*-dsRed	dsRed	Ventral column	Depleted	4-6h	1	62.6
4-6h_Vnd-neg_2	*VC*-dsRed	dsRed	Ventral column	Depleted	4-6h	2	64.3
4-6h_Vnd-pos_1	*VC*-dsRed	dsRed	Ventral column	Enriched	4-6h	1	61.4
4-6h_Vnd-pos_2	*VC*-dsRed	dsRed	Ventral column	Enriched	4-6h	2	65.6
4-6h_Pros-neg_1	*vnd*-lexA/VC:FNLDD	Prospero	Neuroblasts	Depleted	4-6h	1	59.8
4-6h_Pros-neg_2	*vnd*-lexA/VC:FNLDD	Prospero	Neuroblasts	Depleted	4-6h	2	79.1
4-6h_Pros-pos_1	*vnd*-lexA/VC:FNLDD	Prospero	Neuroblasts	Enriched	4-6h	1	61.9
4-6h_Pros-pos_2	*vnd*-lexA/VC:FNLDD	Prospero	Neuroblasts	Enriched	4-6h	2	59.1
6-8h_Ind-neg_1	*IC*-GFP	GFP	Intermediate column	Depleted	6-8h	1	61
6-8h_Ind-neg_2	*IC*-GFP	GFP	Intermediate column	Depleted	6-8h	2	59.9
6-8h_Ind-pos_1	*IC*-GFP	GFP	Intermediate column	Enriched	6-8h	1	57
6-8h_Ind-pos_2	*IC*-GFP	GFP	Intermediate column	Enriched	6-8h	2	70.9
6-8h_Vnd-neg_1	*VC*-dsRed	dsRed	Ventral column	Depleted	6-8h	1	73.2
6-8h_Vnd-neg_2	*VC*-dsRed	dsRed	Ventral column	Depleted	6-8h	2	61
6-8h_Vnd-pos_1	*VC*-dsRed	dsRed	Ventral column	Enriched	6-8h	1	61.6
6-8h_Vnd-pos_2	*VC*-dsRed	dsRed	Ventral column	Enriched	6-8h	2	73.9
6-8h_Pros-neg_1	*vnd*-lexA/VC:FNLDD	Prospero	Neuroblasts	Depleted	6-8h	1	64.5
6-8h_Pros-neg_2	*vnd*-lexA/VC:FNLDD	Prospero	Neuroblasts	Depleted	6-8h	2	76.6
6-8h_Pros-pos_1	*vnd*-lexA/VC:FNLDD	Prospero	Neuroblasts	Enriched	6-8h	1	54.3
6-8h_Pros-pos_2	*vnd*-lexA/VC:FNLDD	Prospero	Neuroblasts	Enriched	6-8h	2	58.8
6-8h_Elav-neg_1	*IC*-GFP	Elav	Neurons	Depleted	6-8h	1	62.2
6-8h_Elav-neg_2	*yw*	Elav	Neurons	Depleted	6-8h	2	57.2
6-8h_Elav-pos_1	*IC*-GFP	Elav	Neurons	Enriched	6-8h	1	53.8
6-8h_Elav-pos_2	*yw*	Elav	Neurons	Enriched	6-8h	2	70.4
6-8h_Repo-neg_1	*IC*-GFP	Repo	Glia	Depleted	6-8h	1	59.6
6-8h_Repo-neg_2	*IC*-GFP	Repo	Glia	Depleted	6-8h	2	71.4
6-8h_Repo_pos_1	*IC*-GFP	Repo	Glia	Enriched	6-8h	1	66.6
6-8h_Repo_pos_2	*IC*-GFP	Repo	Glia	Enriched	6-8h	2	56.8
8-10h_Elav-neg_1	*y w*	Elav	Neurons	Depleted	8-10h	1	74.5
8-10h_Elav-neg_2	*y w*	Elav	Neurons	Depleted	8-10h	2	44.1
8-10h_Elav-pos_1	*y w*	Elav	Neurons	Enriched	8-10h	1	69.9
8-10h_Elav-pos_2	*y w*	Elav	Neurons	Enriched	8-10h	2	62.9
8-10h_Repo-neg_1	*vnd*-lexA/VC:FNLDD	Repo	Glia	Depleted	8-10h	1	61.5
8-10h_Repo-neg_2	*vnd*-lexA/VC:FNLDD	Repo	Glia	Depleted	8-10h	2	60.9
8-10h_Repo-pos_1	*vnd*-lexA/VC:FNLDD	Repo	Glia	Enriched	8-10h	1	61.4

8-10h_Repo-pos_2	*vnd*-lexA/VC:FNLDD	Repo	Glia	Enriched	8-10h	2	65.6
18-22h_Elav-neg_1	*vnd*-lexA/VC:FNLDD	Elav	Neurons	Depleted	18-22h	1	68.6
18-22h_Elav-neg_2	*vnd*-lexA/VC:FNLDD	Elav	Neurons	Depleted	18-22h	2	61.9
18-22h_Elav-pos_1	*vnd*-lexA/VC:FNLDD	Elav	Neurons	Enriched	18-22h	1	66
18-22h_Elav-pos_2	*vnd*-lexA/VC:FNLDD	Elav	Neurons	Enriched	18-22h	2	63.9
18-22h_Repo-neg_1	*vnd*-lexA/VC:FNLDD	Repo	Glia	Depleted	18-22h	1	58.3
18-22h_Repo-neg_2	*vnd*-lexA/VC:FNLDD	Repo	Glia	Depleted	18-22h	2	63.4
18-22h_Repo-pos_1	*vnd*-lexA/VC:FNLDD	Repo	Glia	Enriched	18-22h	1	70.5
18-22h_Repo-pos_2	*vnd*-lexA/VC:FNLDD	Repo	Glia	Enriched	18-22h	2	75.9

Table S11: Summary of RNA-seq datasets – Fractionation-Seq

Sample name	***Fly line***	***Fraction***	***Time point***	***Replicate***	***Number of reads (M)***
Cyto_18-22h_1	*rho*-lexA/VC:FNLDD	Cytoplasmic	18-22h	1	163.2
Cyto_18-22h_2	*rho*-lexA/VC:FNLDD	Cytoplasmic	18-22h	2	179.1
Cyto_6-8h_1	*rho*-lexA/VC:FNLDD	Cytoplasmic	6-8h	1	189.3
Cyto_6-8h_2	*rho*-lexA/VC:FNLDD	Cytoplasmic	6-8h	2	179.3
Nuc_18-22h_1	*rho*-lexA/VC:FNLDD	Nuclear	18-22h	1	203
Nuc_18-22h_2	*rho*-lexA/VC:FNLDD	Nuclear	18-22h	2	176.1
Nuc_6-8h_1	*rho*-lexA/VC:FNLDD	Nuclear	6-8h	1	193.9
Nuc_6-8h_2	*rho*-lexA/VC:FNLDD	Nuclear	6-8h	2	193.7
Whole_18-22h_1	*rho*-lexA/VC:FNLDD	Whole embryo	18-22h	1	183.2
Whole_18-22h_2	*rho*-lexA/VC:FNLDD	Whole embryo	18-22h	2	180.7
Whole_6-8h_1	*rho*-lexA/VC:FNLDD	Whole embryo	6-8h	1	197.4
Whole_6-8h_2	*rho*-lexA/VC:FNLDD	Whole embryo	6-8h	2	196.1

Table S12: All protein-coding 'computed genes' correlated (r > 0.9) with neurogenic marker genes in DIV-SortSeq expression data

Flybase ID	*Annotation ID*
FBgn0036725	CG18265
FBgn0034009	CG8155
FBgn0036008	CG3408
FBgn0030017	CG2278
FBgn0264449	CG43867
FBgn0051235	CG31235
FBgn0037166	CG11426
FBgn0032752	CG10702
FBgn0038114	CG11670
FBgn0037206	CG12768
FBgn0032485	CG9426
FBgn0038321	CG6218
FBgn0038720	CG6231
FBgn0033283	CG11635
FBgn0043806	CG32032
FBgn0033287	CG8701
FBgn0031540	CG3238
FBgn0031961	CG7102
FBgn0040984	CG4440
FBgn0026876	CG11403
FBgn0035689	CG7376
FBgn0025388	CG12179
FBgn0036503	CG13454
FBgn0040809	CG13465
FBgn0032489	CG15480
FBgn0041702	CG15107
FBgn0034403	CG18190
FBgn0031070	CG12702
FBgn0250754	CG42232
FBgn0261538	CG42662
FBgn0034514	CG13427
FBgn0036670	CG13029
FBgn0031252	CG13690
FBgn0040346	CG3704
FBgn0039013	CG4813
FBgn0037633	CG9839
FBgn0028506	CG4455
FBgn0037844	CG4570
FBgn0037924	CG14712
FBgn0035464	CG12006
FBgn0027602	CG8611
FBgn0037051	CG10565
FBgn0052756	CG32756
FBgn0036710	CG6479
FBgn0039735	CG7911
FBgn0023515	CG14814
FBgn0033160	CG11107
FBgn0033169	CG11123
FBgn0029672	CG2875
FBgn0025633	CG13366
FBgn0050183	CG30183
FBgn0030768	CG9723
FBgn0025626	CG4281
FBgn0035213	CG2199
FBgn0031403	CG15387
FBgn0043456	CG4747
FBgn0052428	CG32428
FBgn0031062	CG14230
FBgn0035315	CG8960
FBgn0032050	CG13096
FBgn0038552	CG18012
FBgn0031764	CG9107
FBgn0035878	CG7182
FBgn0034073	CG8414
FBgn0033766	CG8771
FBgn0037958	CG6962
FBgn0034447	CG7744
FBgn0035842	CG7504
FBgn0038272	CG7265
FBgn0039544	CG12877
FBgn0033615	CG7741
FBgn0030122	CG16892
FBgn0051365	CG31365
FBgn0052318	CG32318
FBgn0039566	CG4849
FBgn0037746	CG8478
FBgn0030813	CG4949
FBgn0029825	CG12728
FBgn0038768	CG4936
FBgn0037149	CG14561
FBgn0037372	CG2091
FBgn0030915	CG6179
FBgn0027514	CG1024
FBgn0031947	CG7154
FBgn0030660	CG8097
FBgn0036565	CG5235
FBgn0024364	CG11417
FBgn0036994	CG5199
FBgn0033527	CG11777
FBgn0039743	CG7946
FBgn0262719	CG43163
FBgn0036886	CG9300
FBgn0034264	CG10933
FBgn0025627	CG4194
FBgn0032682	CG10176
FBgn0031001	CG7884
FBgn0037213	CG12581
FBgn0038660	CG14291
FBgn0034933	CG3735
FBgn0030855	CG5800
FBgn0035414	CG14965
FBgn0036214	CG7264
FBgn0030317	CG1561
FBgn0036660	CG13025
FBgn0031597	CG17612
FBgn0036483	CG12316
FBgn0032512	CG9305
FBgn0037644	CG11964
FBgn0040385	CG12496
FBgn0033802	CG17724
FBgn0039733	CG11504
FBgn0037504	CG1142
FBgn0036522	CG7372
FBgn0035402	CG12082
FBgn0031769	CG9135
FBgn0038551	CG7357
FBgn0032348	CG4751
FBgn0035987	CG3689
FBgn0037918	CG6791
FBgn0035872	CG7185
FBgn0031529	CG9662
FBgn0031492	CG3542
FBgn0086855	CG17078
FBgn0033990	CG10265
FBgn0032454	CG5787
FBgn0037622	CG8202
FBgn0032751	CG17343
FBgn0266917	CG16941
FBgn0034114	CG4282
FBgn0034750	CG3732
FBgn0030738	CG9915
FBgn0035235	CG7879
FBgn0028474	CG4119
FBgn0085451	CG34422
FBgn0030293	CG1737
FBgn0050020	CG30020
FBgn0029941	CG1677
FBgn0033021	CG10417
FBgn0027503	CG11970
FBgn0032388	CG6686
FBgn0038546	CG7379
FBgn0035481	CG12605
FBgn0260451	CG14042
FBgn0035677	CG13293
FBgn0036202	CG6024
FBgn0035643	CG13287
FBgn0034184	CG9646
FBgn0052105	CG32105
FBgn0031257	CG4133
FBgn0037050	CG10566
FBgn0025712	CG13920
FBgn0027550	CG6495
FBgn0031762	CG9098
FBgn0035246	CG13928
FBgn0039808	CG12071
FBgn0030508	CG15760
FBgn0033960	CG10151
FBgn0030596	CG12398
FBgn0052085	CG32085
FBgn0030012	CG18262
FBgn0030223	CG2111
FBgn0051030	CG31030
FBgn0035903	CG6765
FBgn0030595	CG14406
FBgn0033983	CG10253
FBgn0031955	CG14535
FBgn0033497	CG12912
FBgn0034154	CG5267
FBgn0033872	CG6329
FBgn0030588	CG9521
FBgn0085382	CG34353
FBgn0085218	CG34189
FBgn0030594	CG9509
FBgn0034459	CG16716
FBgn0259823	CG42404
FBgn0032800	CG10137
FBgn0038926	CG13409
FBgn0039064	CG4467
FBgn0259163	CG42268
FBgn0259994	CG42492
FBgn0039030	CG6660
FBgn0036579	CG5027
FBgn0029708	CG3556
FBgn0030586	CG12539
FBgn0038596	CG14312
FBgn0263072	CG43347
FBgn0037525	CG17816
FBgn0030742	CG9919
FBgn0265084	CG44195
FBgn0030592	CG9514
FBgn0262476	CG43066
FBgn0035033	CG3548
FBgn0040351	CG11638
FBgn0028647	CG11902
FBgn0031313	CG5080
FBgn0039915	CG1732
FBgn0029896	CG3168
FBgn0039024	CG4721
FBgn0032897	CG9336
FBgn0034128	CG4409
FBgn0034417	CG15117
FBgn0031816	CG16947
FBgn0030261	CG15203
FBgn0033446	CG1648
FBgn0052354	CG32354
FBgn0031627	CG15630
FBgn0036927	CG7433
FBgn0034618	CG9485
FBgn0083972	CG34136
FBgn0031589	CG3714
FBgn0036760	CG5567
FBgn0032899	CG9338

Supplementary Figures

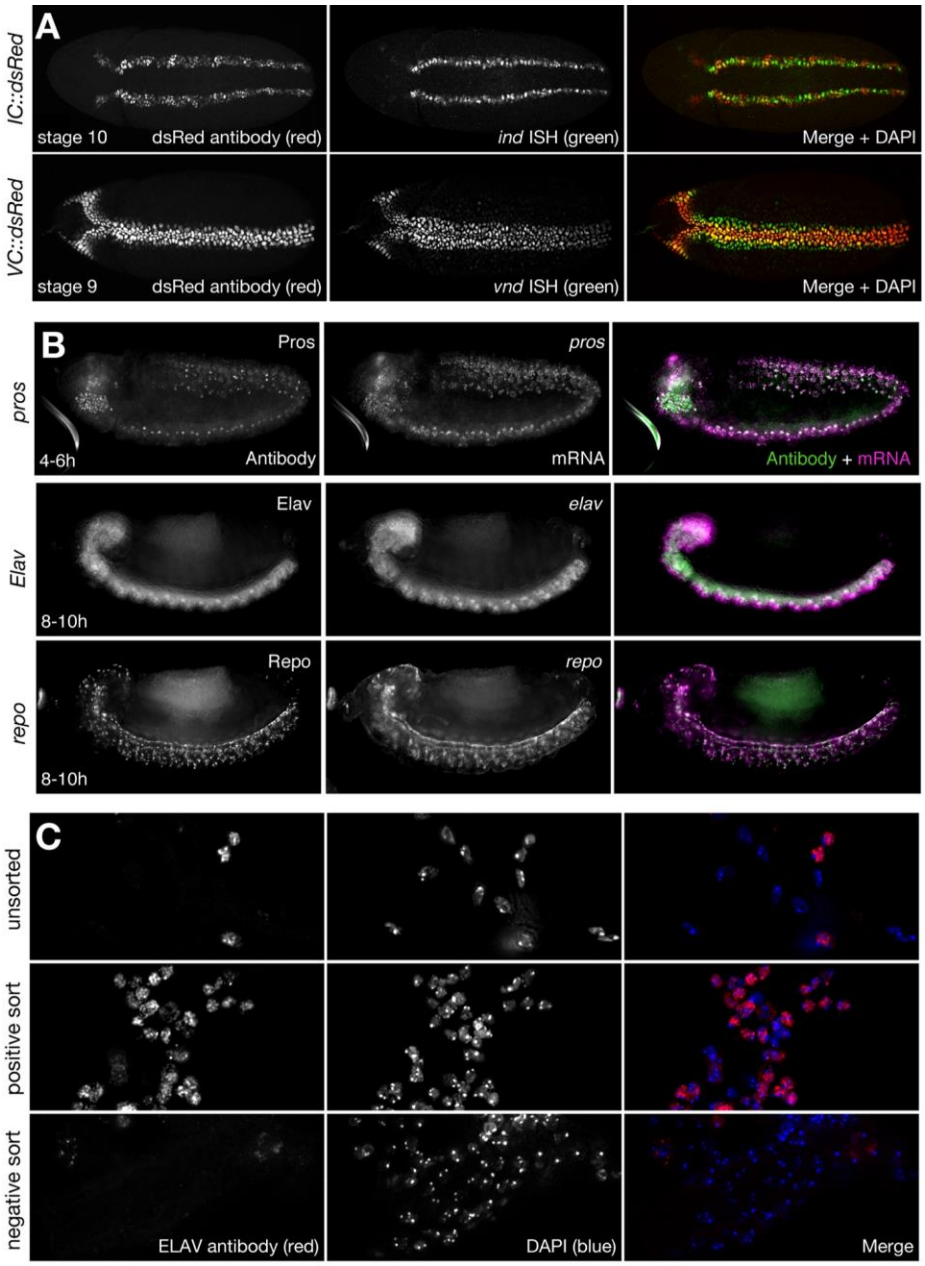


Fig. S1: Neurogenic cell populations can be faithfully marked and purified
(A) Visualization of fluorescent reporter expression in *VC-dsRed* (top panel) and *IC-dsRed* (bottom panel) transgenic embryos. Tissue reporter shown on left and in red by antibody stain, expression of the tissue marker is shown in the middle and in green by *in situ* hybridization. Shown are whole mount embryos, ventral views anterior left. (B) Multiplex whole mount immunohistochemistry (green) and RNA-FISH (magenta) show faithfulness of the antibody sorting markers for neuroblasts (Pros; top panel) at 4-6h, and neurons (Elav; middle panel) and glia (Repo; bottom panel) at 8-10h. (C) Immunohistochemical staining of endogenous Elav protein in cells of dissociated embryos pre- and post-FACS.

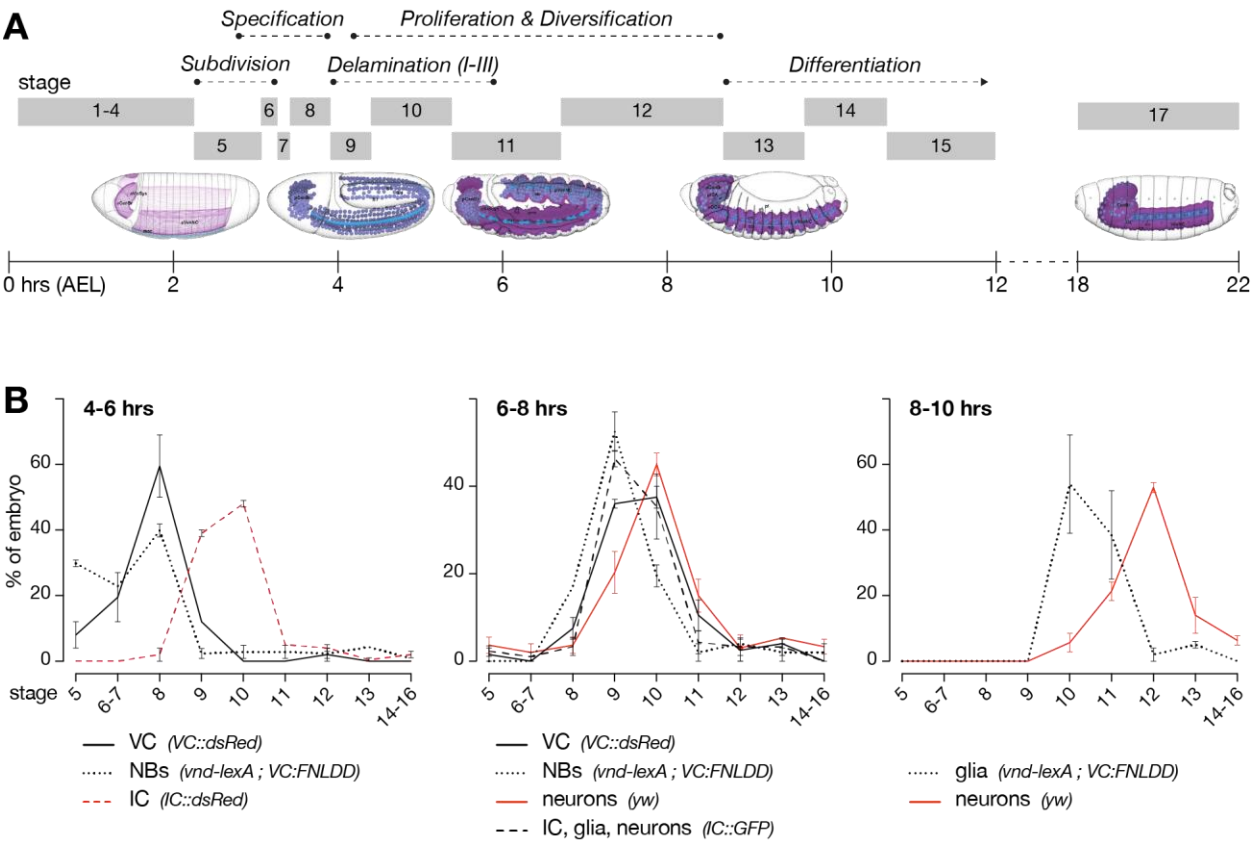


Fig. S2: Timed embryo collections encompass major neurogenic events
(A) Overview of *Drosophila* embryonic neurogenesis, timing and stages. Embryo images show stages of neurogenesis schematically and were obtained from the *Atlas of Drosophila Development* by Volker Hartstein (CSHL Press, 1992, used with permission). (B) Staging of representative samples corresponding to timed embryo collections (n = 2-3 collections per line). Staging according to (Campos-Ortega & Hartenstein 1985). Cell type markers targeted for *DIV*-MARIS indicated for each fly line.

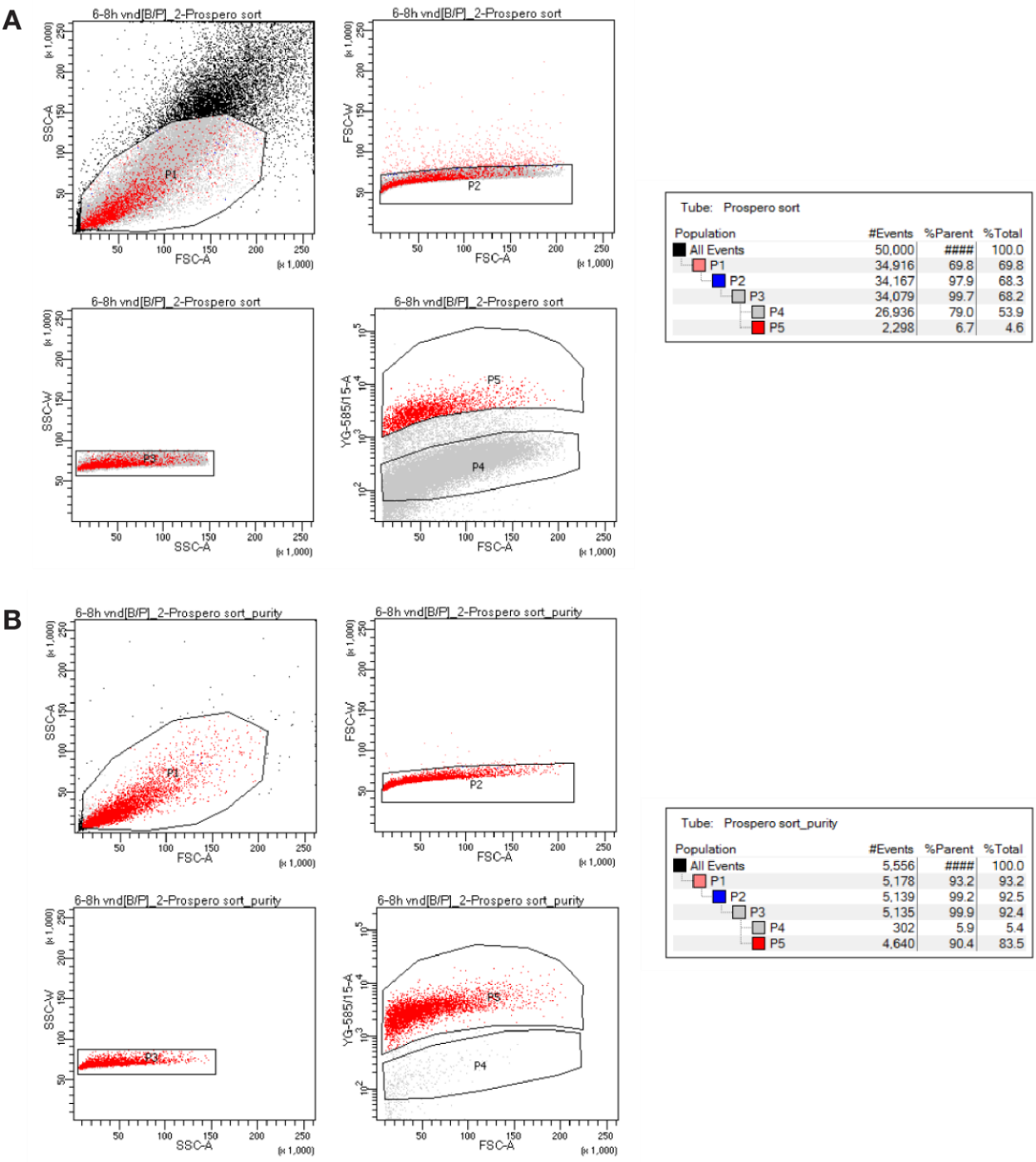
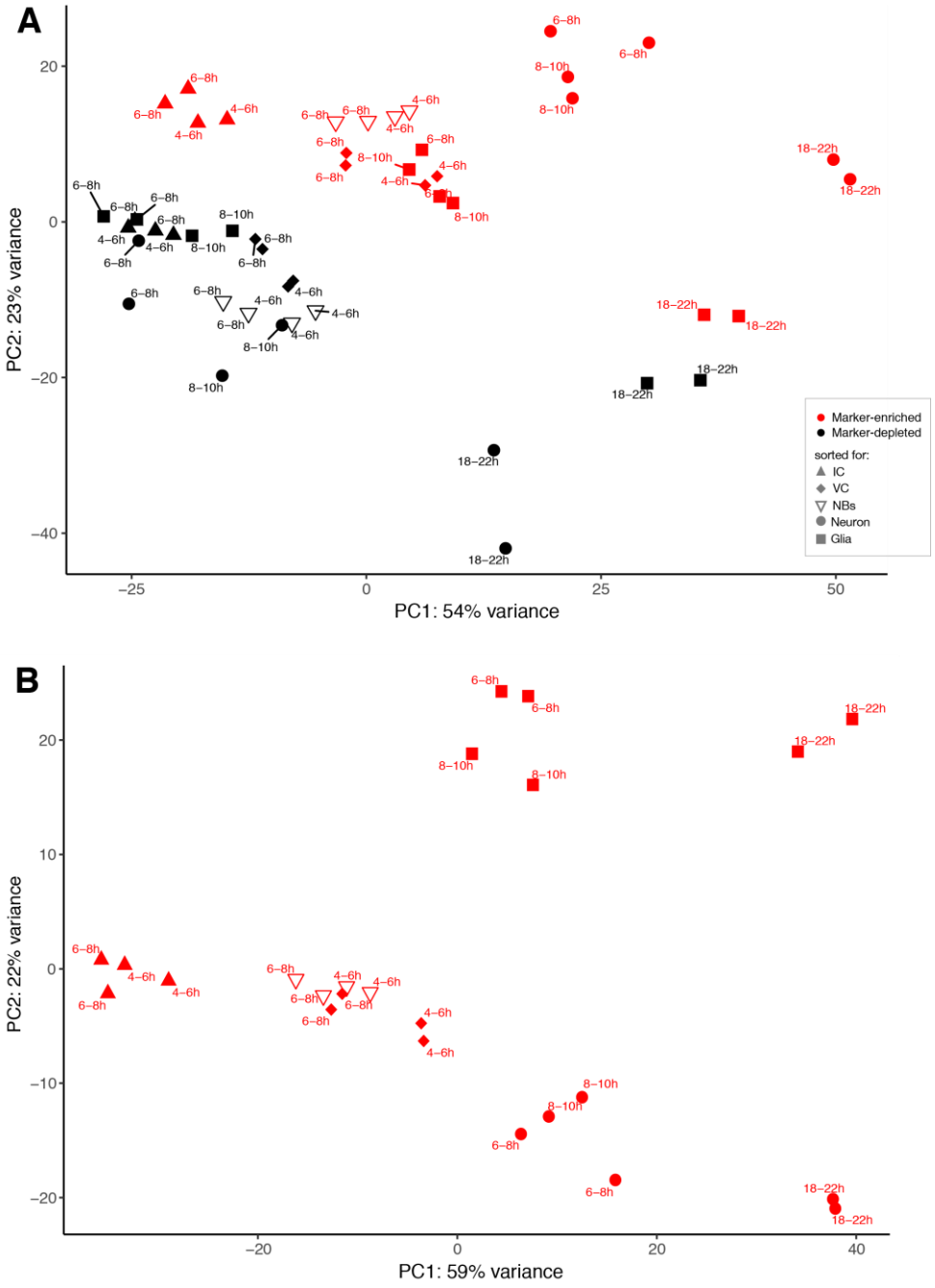
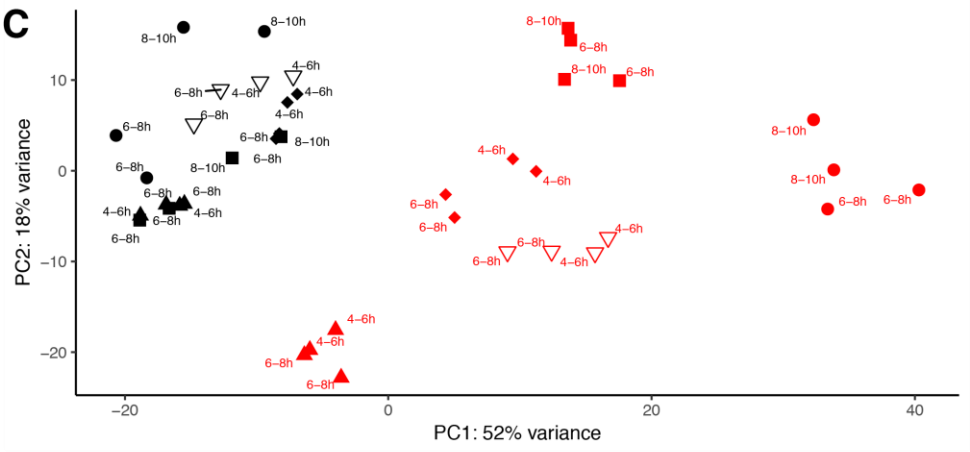
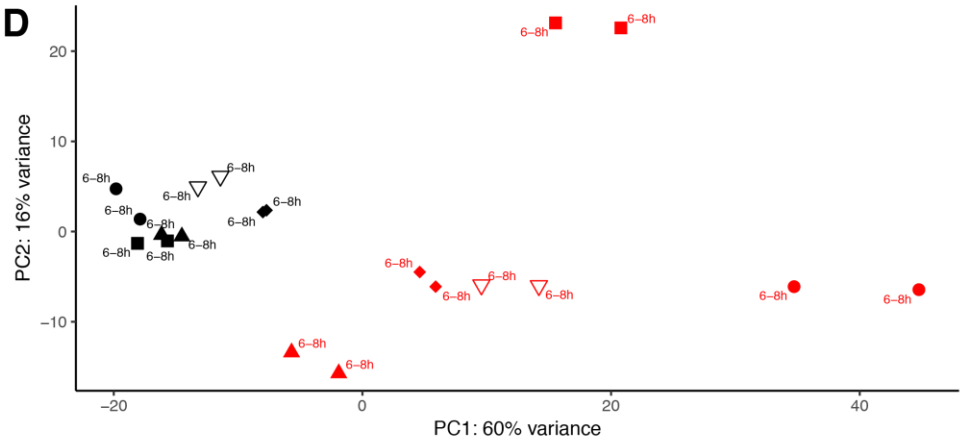


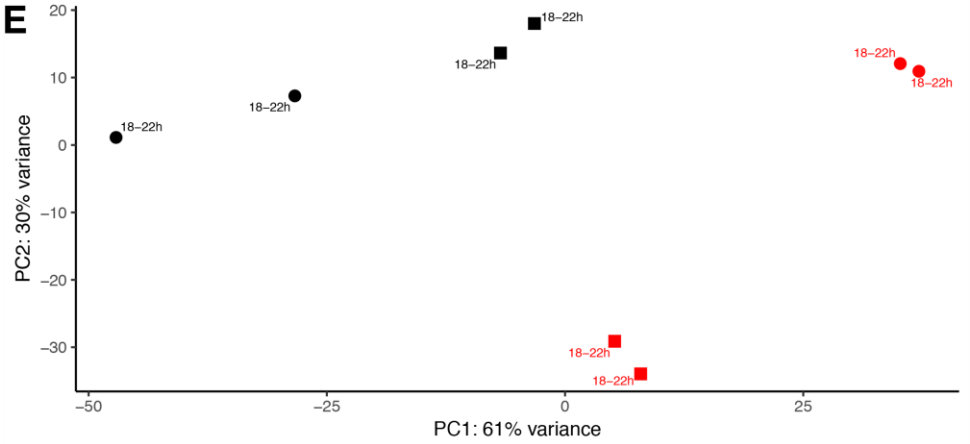
Fig. S3: FACS gating strategy yields high purity of sorted populations

(A) Example FACS gating strategy for sorting of marker-positive (P5; 6.7%) and marker-negative cells (P4; 79%). (B) Re-sort of marker-positive sorted cells shows enrichment of marker-positive (P5; 90.4%) and depletion of marker-negative cells (P4; 5.9%).









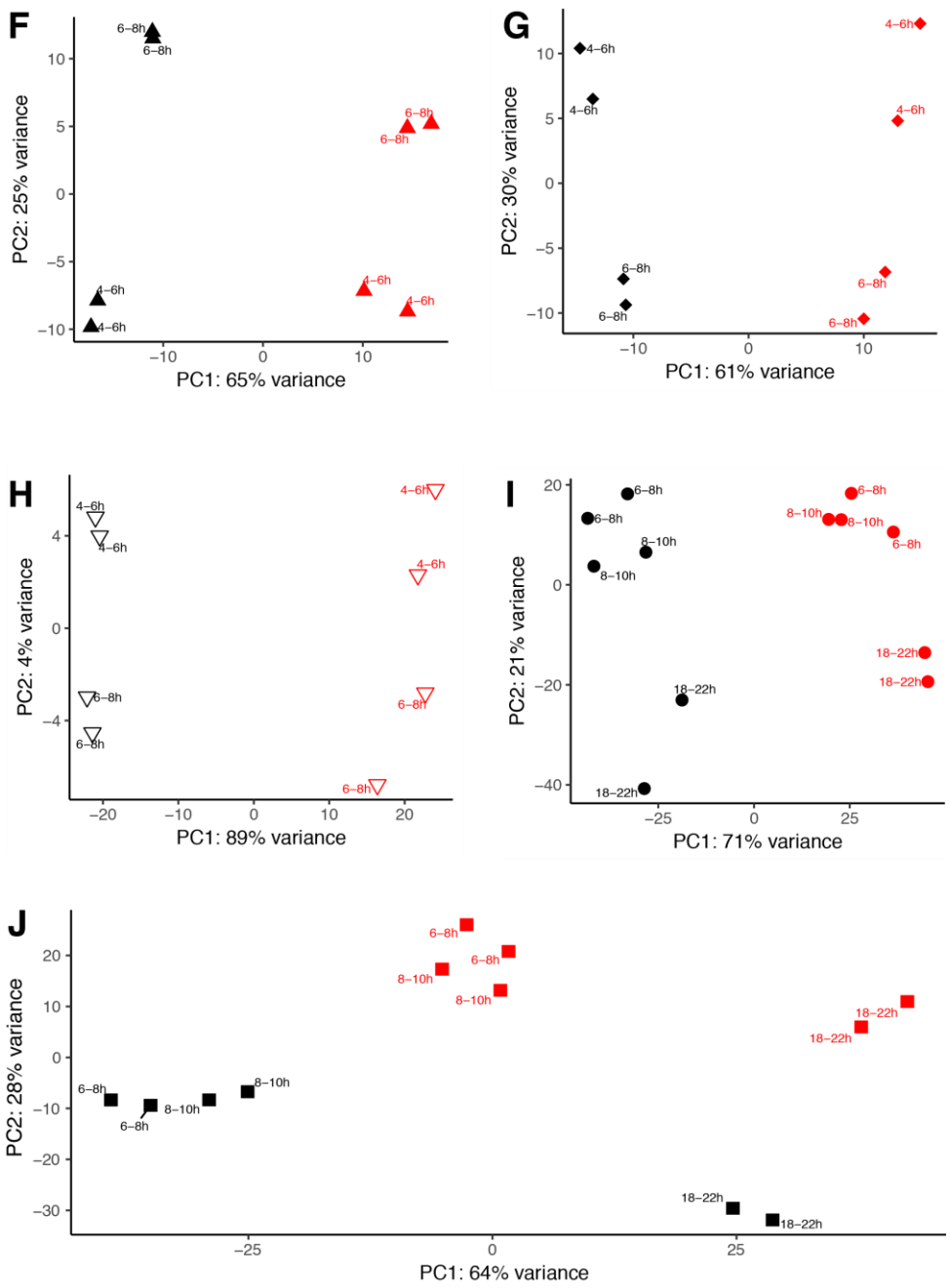


Fig. S4: Principal component analyses separate spatiotemporal transcriptomes according to tissue and developmental age

Principal Component Analysis (PCA) with multiple permutations of datasets generated by *DIV*-SortSeq. Marker-enriched datasets depicted in red, marker-depleted datasets in black. (A) All datasets. (B). Only marker-positive datasets. (C). Only early (4-6, 6-8, 8-10h) datasets. (D) Only 6-8h datasets. (E). Only 18-22h datasets. (F). Only datasets sorted for *ind* expression. (G). All datasets sorted for *vnd* expression. (H) All datasets sorted for *pros* expression. (I) All datasets sorted for *elav* expression. (J) All datasets sorted on *repo* expression.

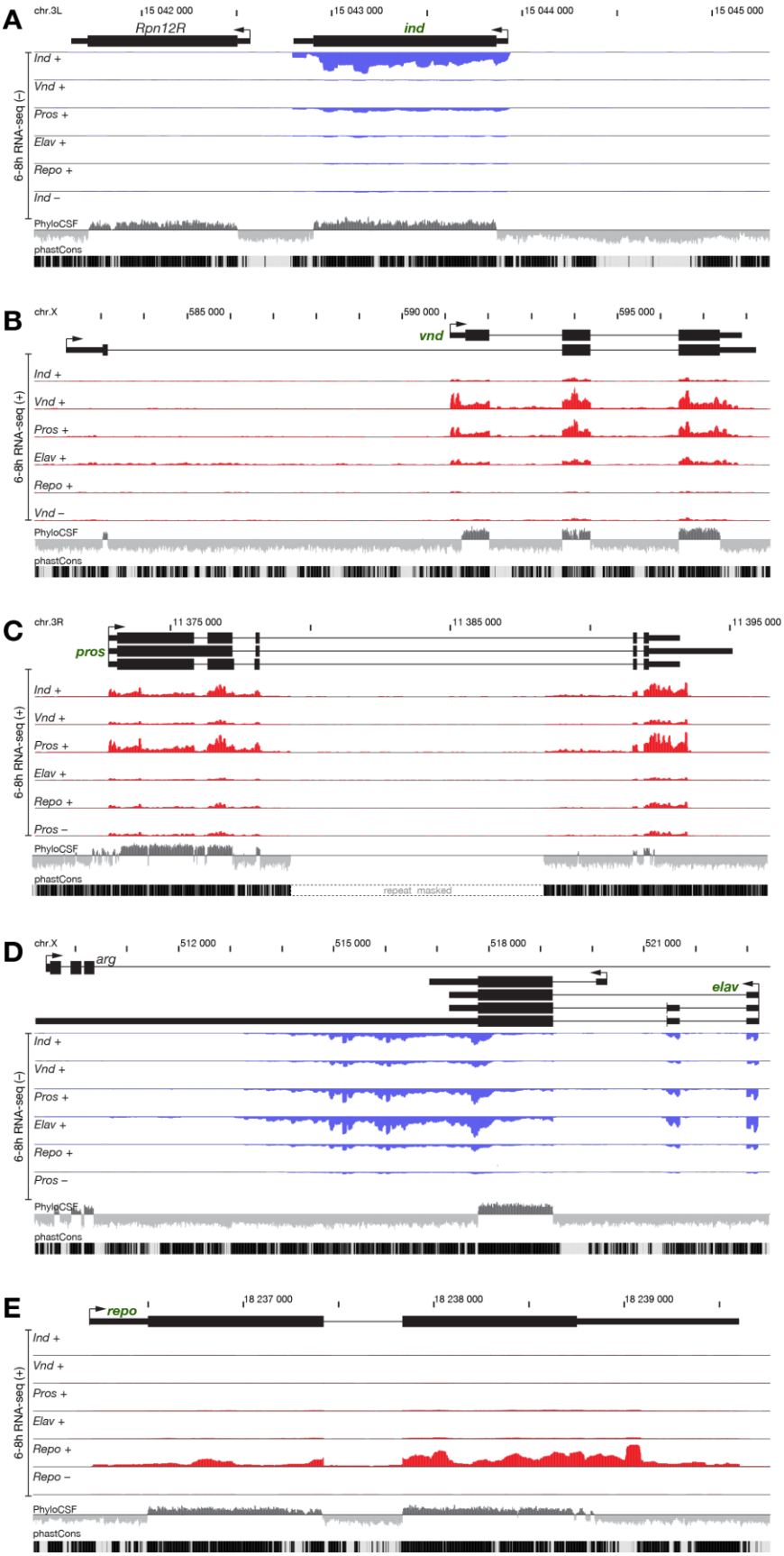


Fig. S5: Cell type specific transcriptomes at neurogenic marker genes
(A-E) Genome browser data of cell type specific transcriptomes around marker gene loci. Shown are annotated transcripts (top), RNA-seq coverage on the plus- (red) or minus-strand (blue) in the indicated sorts at 6-8hrs AEL, as well as coding potential as measured by PhyloCSF scores (all frames overlaid), and conservation amongst *Drosophilids* (phastCons). (A) *ind* locus. (B) *vnd* locus. (C) *pros* locus. (D) *elav* locus. (E) *repo* locus.

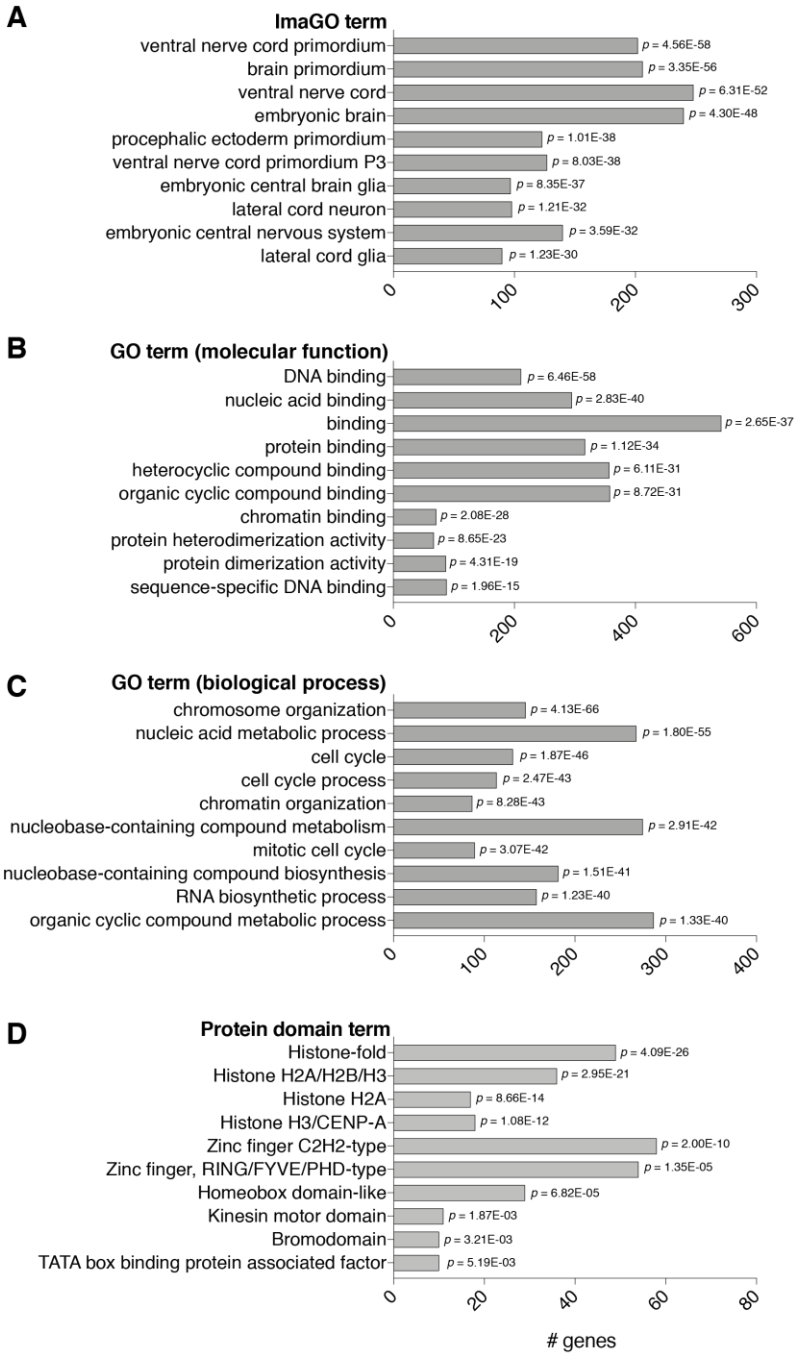


Fig. S6: Annotations of regulated protein-coding genes predict expression and function
Analysis of features of 794 protein-coding genes correlated with known neurogenic genes (Pearson correlation, $r > 0.9$) with Flymine {Lyne:2007jd}. (A) Top ImaGO terms. (B) Top GO terms (molecular function). (C) Top GO terms (biological process). (D) Top protein domain terms. Detailed analysis of all genes can be found in File S2.

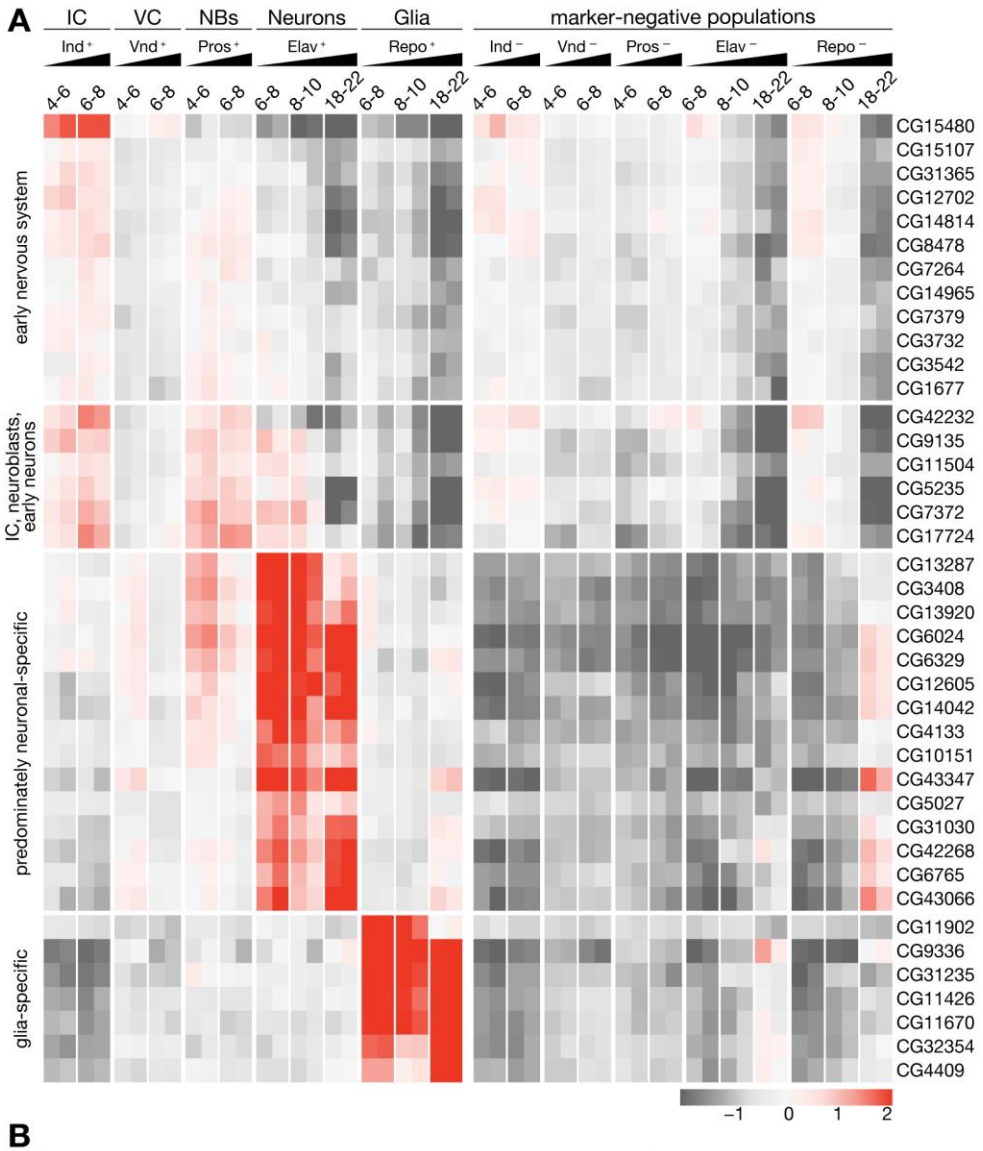
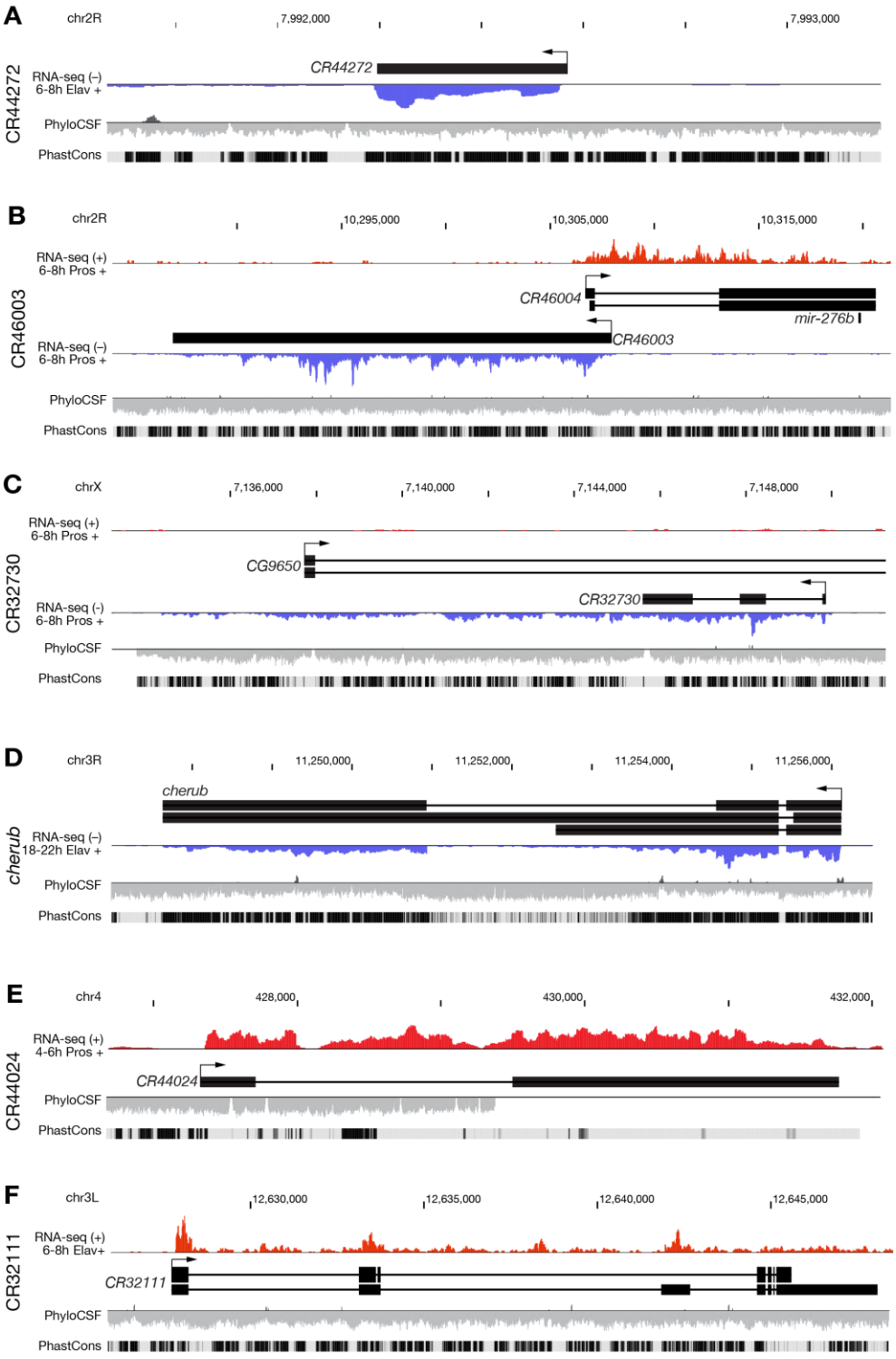


Fig. S7: DIV-SortSeq reveals a multitude of undescribed genes expressed with neurogenic spatiotemporal specificity

(A) Heatmap of expression profiles of protein-coding genes with unknown function enriched similarly as at least one marker in Fig. 2B (Pearson correlation, $r > 0.9$). Row mean-centered expression values calculated via variance-stabilizing transformation of gene-level RNA-seq counts (scale = $\log_2$ ratio to row mean). (B) Colorimetric RNA *in situ* hybridization of transcripts of unknown function identified in Fig. S7A. All embryos are shown in lateral orientation. Images obtained from Berkeley *Drosophila* Genome Project (Hammonds et al. 2013; Tomancak et al. 2002; Tomancak et al. 2007). A list of all 212 CGs with r>0.9 is available in Table S12.



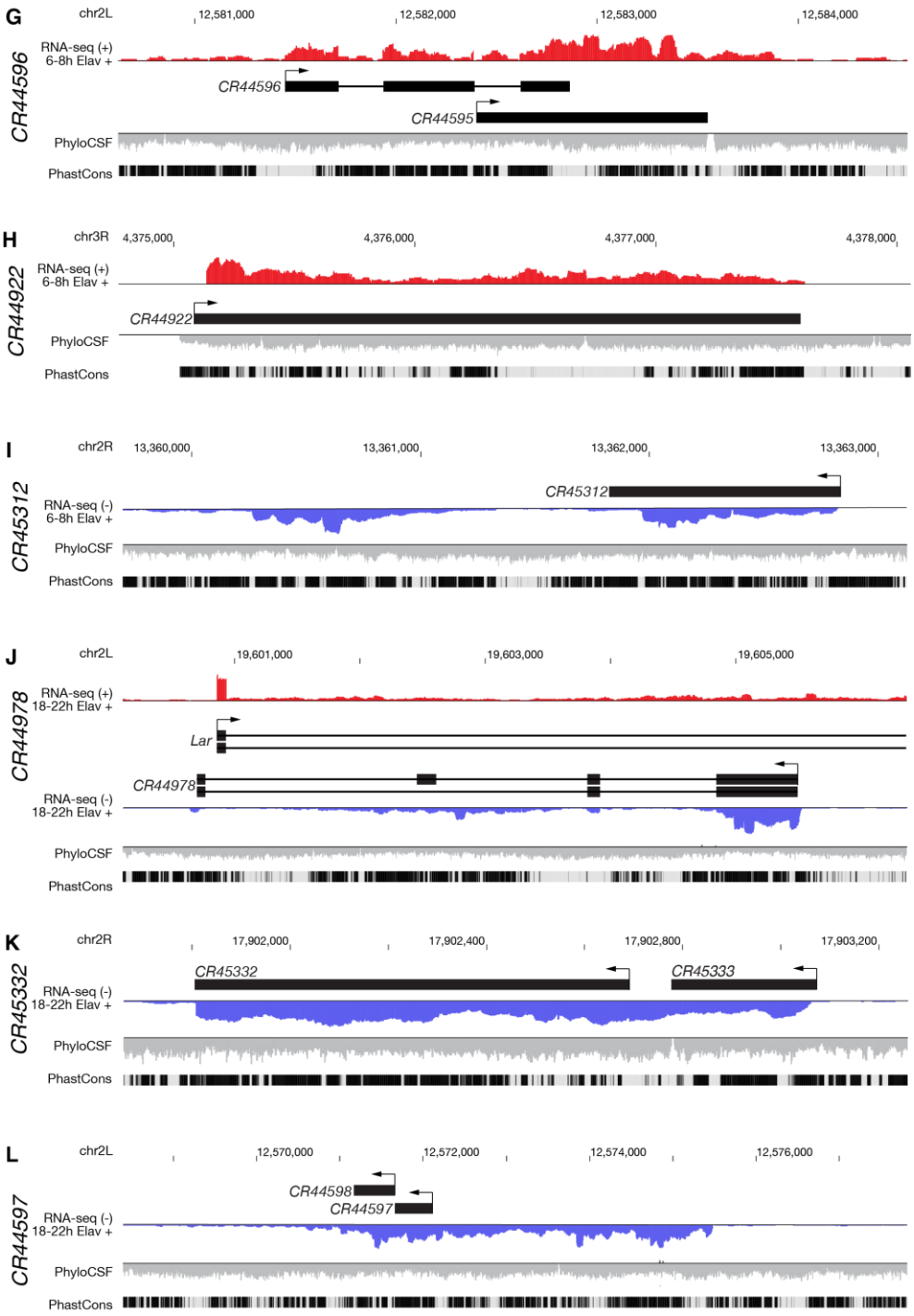


Figure S8: Cell type specific transcriptomes at neurogenic lncRNA loci
Browser images of selected lncRNA loci, display as in Figure S4: The (A) *CR44272,* (B) *CR46003* and *CR46004* (pri-miR-276b), (C) *CR32730,* (D) *cherub,* (E) *CR44024,* (F) *CR32111,* (G) *CR44596,* (H) *CR44922,* (I) *CR45312,* (J) *CR44978,* (K) *CR45332* (This gene model might be misannotated – CR45332 and CR45333 might be a single transcript*),* and the (L) *CR44596* and *CR44597* locus.

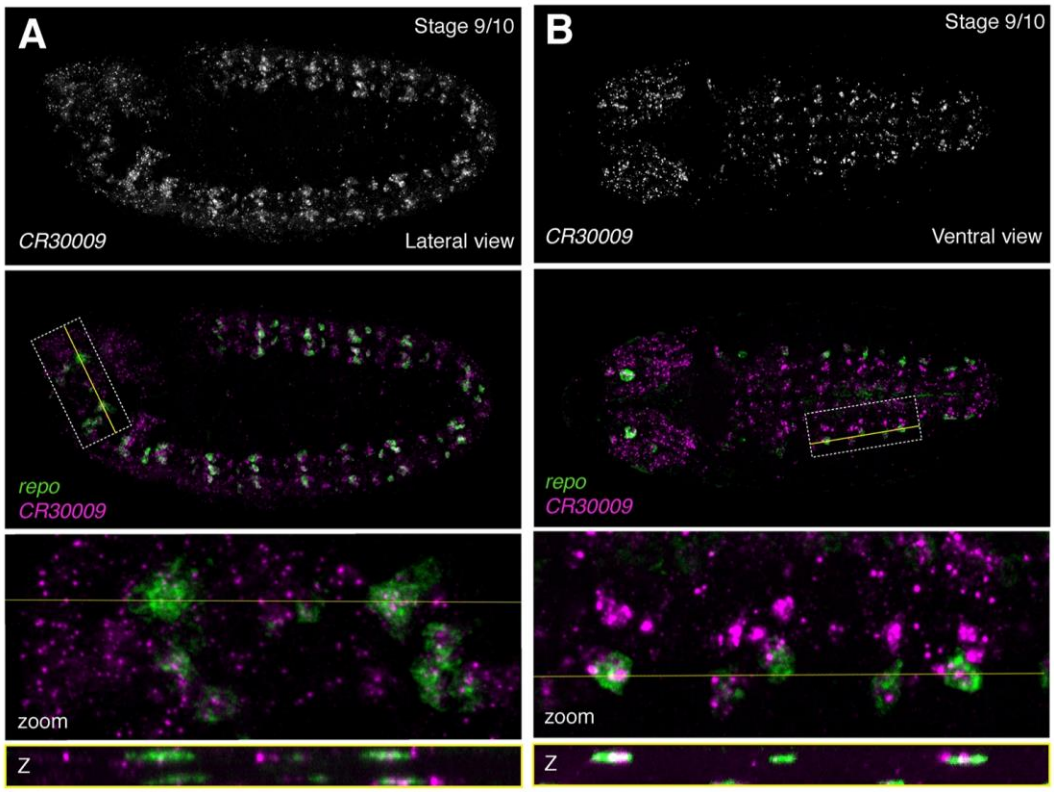


Fig. S9: The lncRNA *CR30009* is expressed in glial subsets in early embryos
RNA fluorescent in situ hybridization (RNA-FISH) against *CR30009* and the glial marker *repo*. (A) Lateral view, stage 9/10. (B) Ventral view, stage 9/10. Top: *CR30009* alone. Second from top: *CR30009* (magenta) overlaid with *repo* (green). Dashed white box indicates region of interest (ROI) and yellow line indicates Z-slice through ROI. Second from bottom: zoom-in of ROI. Bottom: Slice through Z-stack as indicated by yellow line.

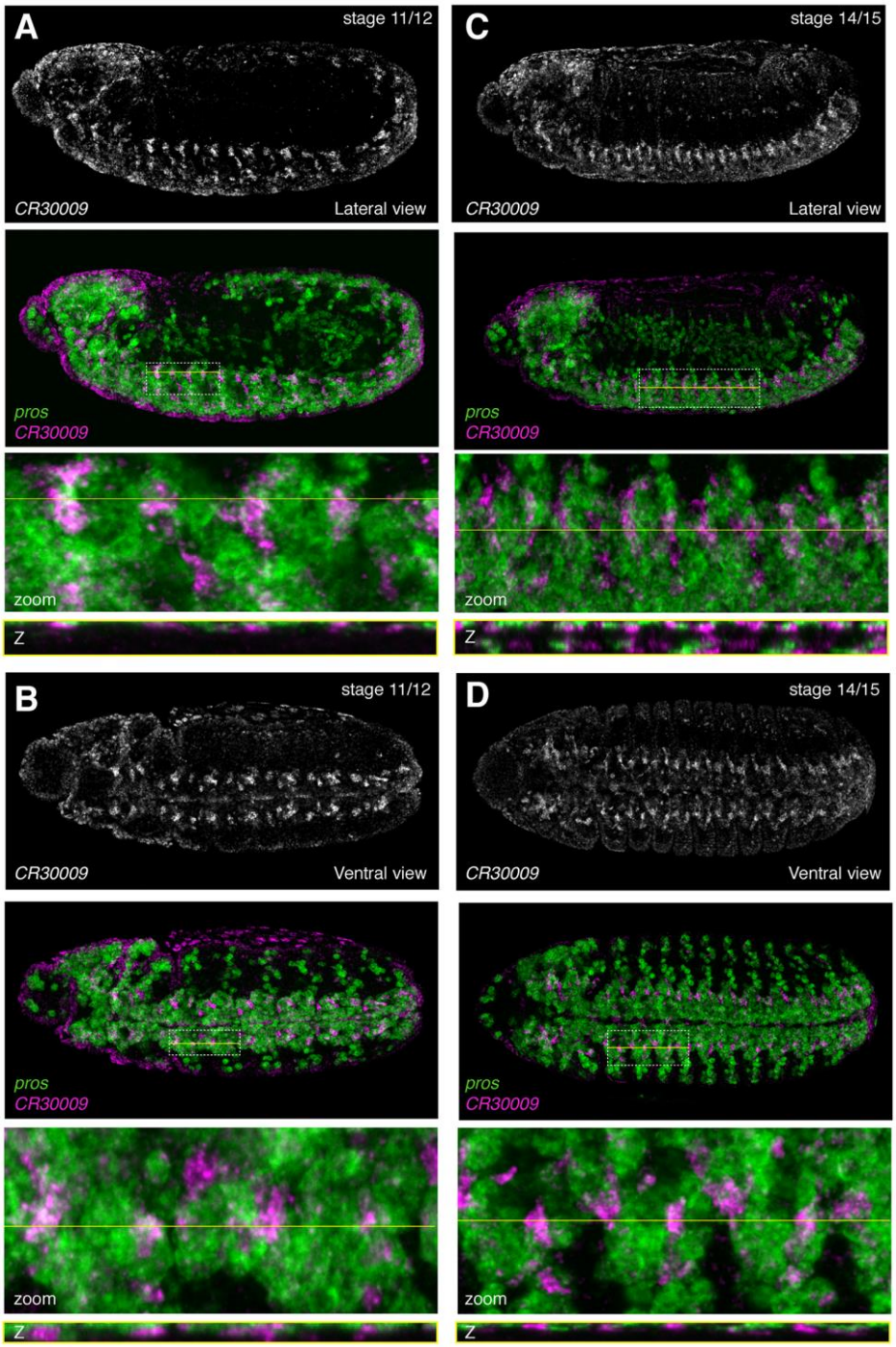


Fig. S10: The lncRNA *CR30009* is expressed in neuroblast subsets

RNA fluorescent in situ hybridization (RNA-FISH) against *CR30009* and the neuroblast marker *pros*. (A) Lateral view, stage 11/12. (B) Lateral view, stage 14/15. (C) Ventral view; stage 11/12. (D). Ventral view; stage 14/15. Top: *CR30009* alone. Second from top: *CR30009* (magenta) overlaid with *pros* (green). Dashed white box indicates region of interest (ROI) and yellow line indicates Z-slice through ROI. Second from bottom: zom-in of ROI. Bottom: Slice through Z-stack as indicated by yellow line.

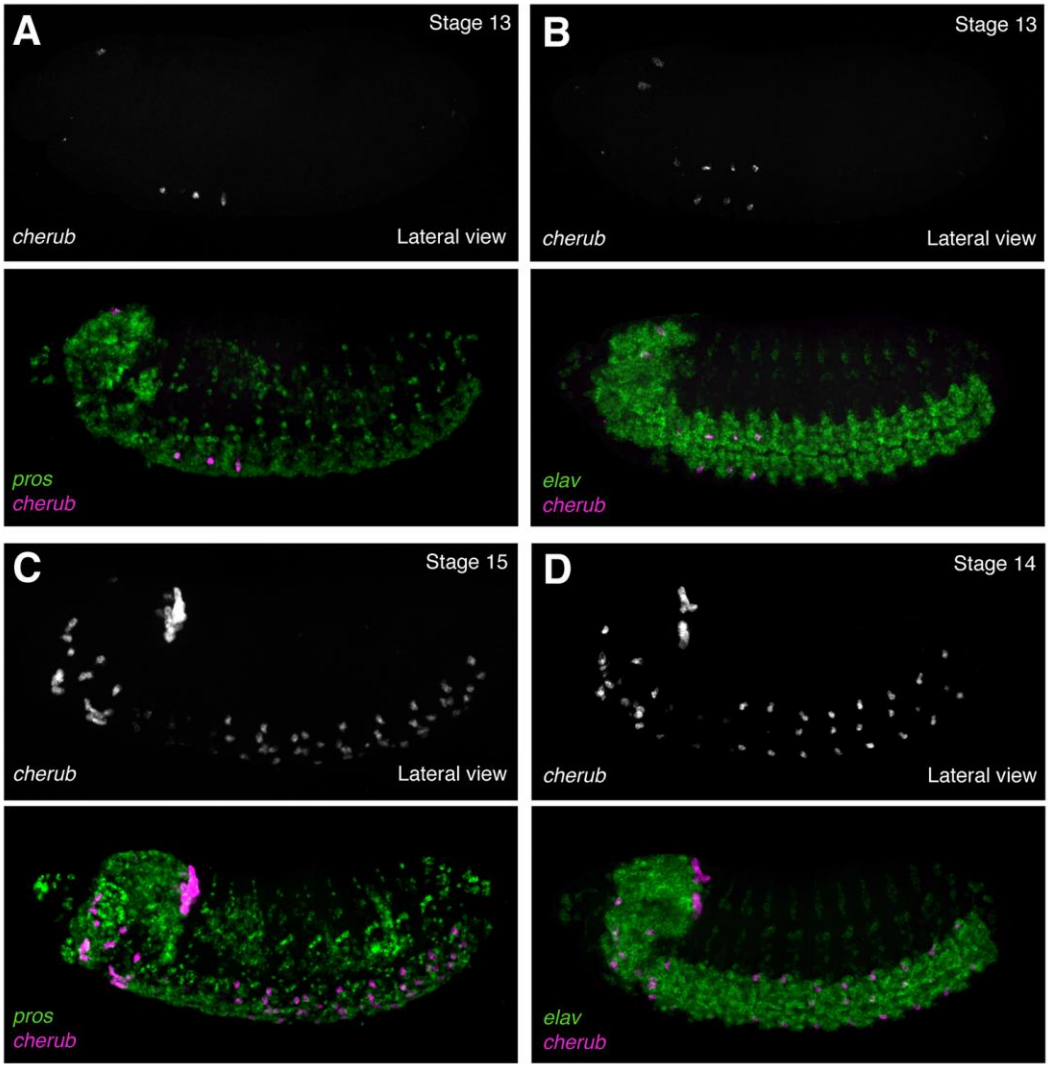


Fig S11: The lncRNA *cherub* is regulated with strict spatiotemporal specificity.
RNA fluorescent in situ hybridization (RNA-FISH) against *cherub*, the neuroblast marker *pros*, and the neuronal marker *elav*. Lateral view. (A) *cherub* with *pros*; stage 13. (B) *cherub* with *elav*; stage 13. (C) *cherub* with *pros*; stage 15. (D) *cherub* with *elav*; stage 14. Top: *cherub* alone. Second from top: *cherub* (magenta) overlaid with marker (green).

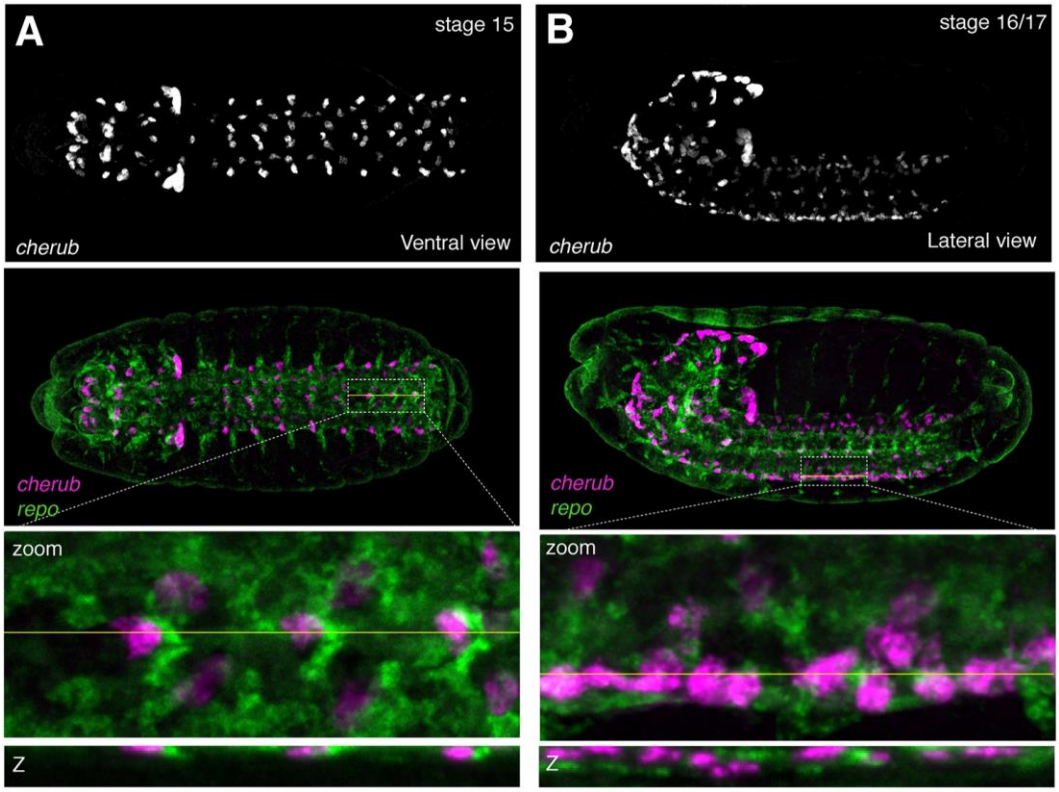


Fig S12: The lncRNA *cherub* is expressed in glial subsets in late embryos
RNA fluorescent in situ hybridization (RNA-FISH) against *cherub* and the glial marker *repo*. (A) Ventral view; stage 15. (B) Lateral view; stage 16/17. Top: *cherub* alone. Second from top: *cherub* (magenta) overlaid with *repo* (green). Dashed white box indicates region of interest (ROI) and yellow line indicates Z-slice through ROI. Second from bottom: zom-in of ROI. Bottom: Slice through Z-stack as indicated by yellow line.

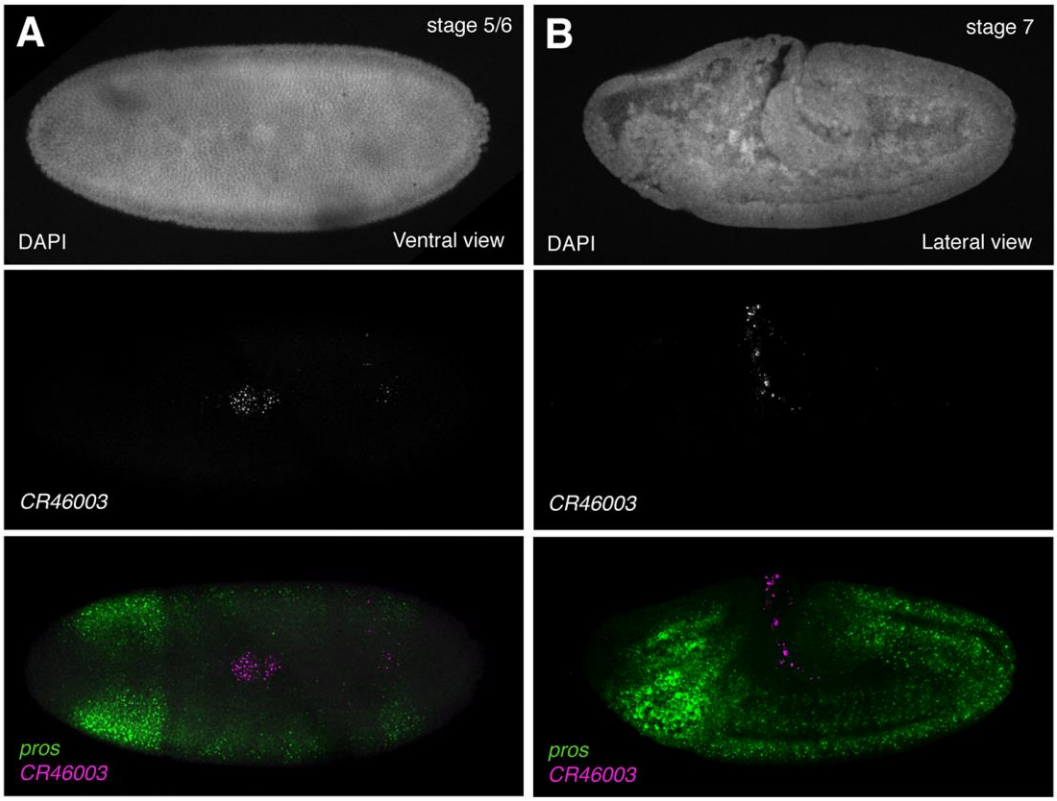


Fig S13: The lncRNA *CR46003* is expressed in early embryogenesis
RNA fluorescent *in situ* hybridization (RNA-FISH) against *CR46003* with the neuroblast marker *pros*. (A) Dorsal view; stage 5/6. (B) Lateral view; stage 7/8. Top: DAPI, middle: *CR46003* alone. Bottom: *CR46003* (magenta) overlaid with *pros* (green).

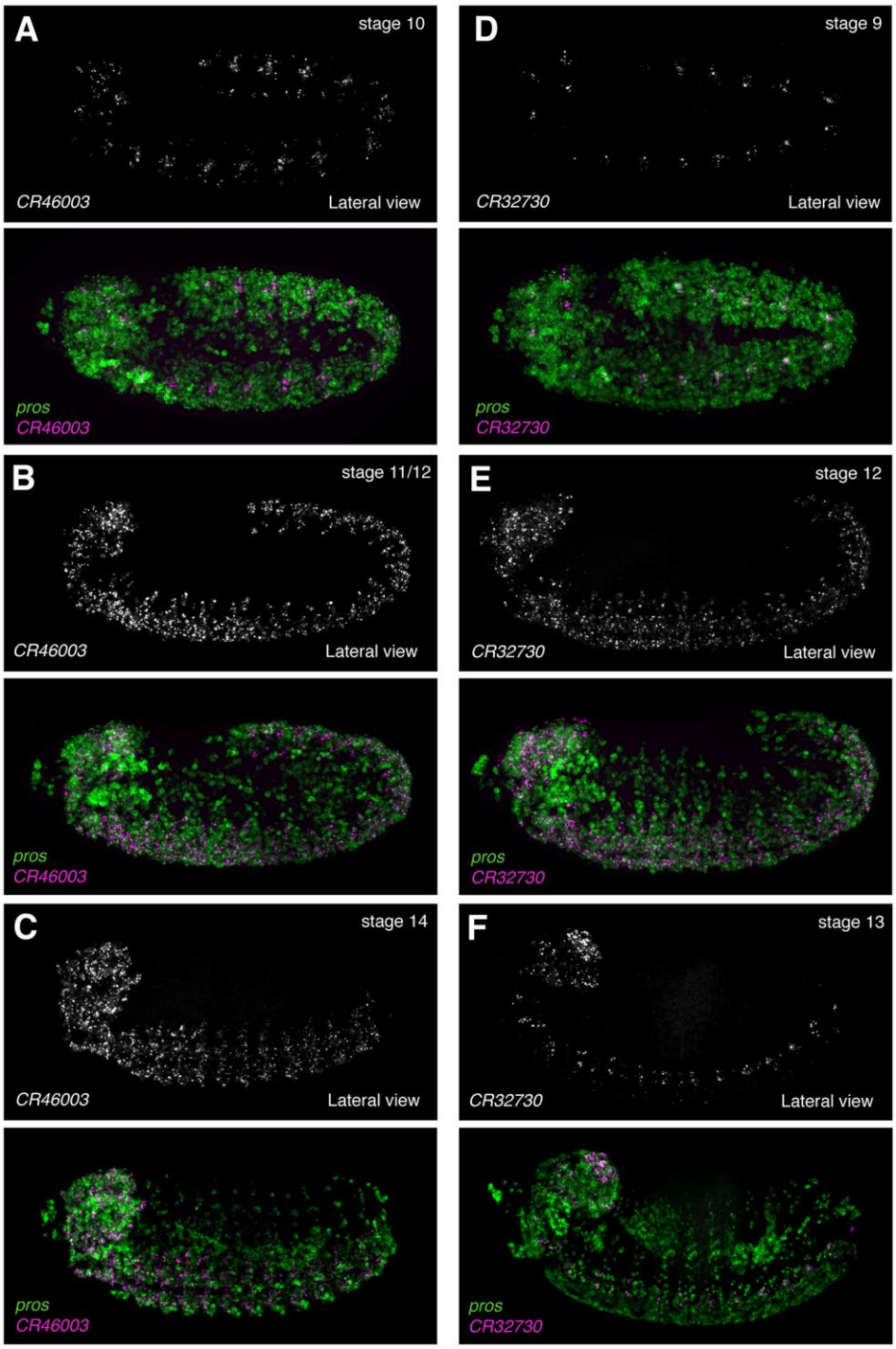


Fig S14: The lncRNAs *CR46003* and *CR32730* are expressed with similar spatiotemporal specificity.

RNA fluorescent *in situ* hybridization (RNA-FISH) against *CR46003* and *CR32730* together with the neuroblast marker *pros*. Lateral view. (A) *CR46003*; stage 10. (B) *CR46003*; stage 11/12. (C) *CR46003*; stage 14. (D) *CR32730*; stage 9. (E) *CR32730*; stage 12. (F) *CR32730*; stage 13. Top: lncRNA alone. Second from top: lncRNA (magenta) overlaid with *pros* (green).

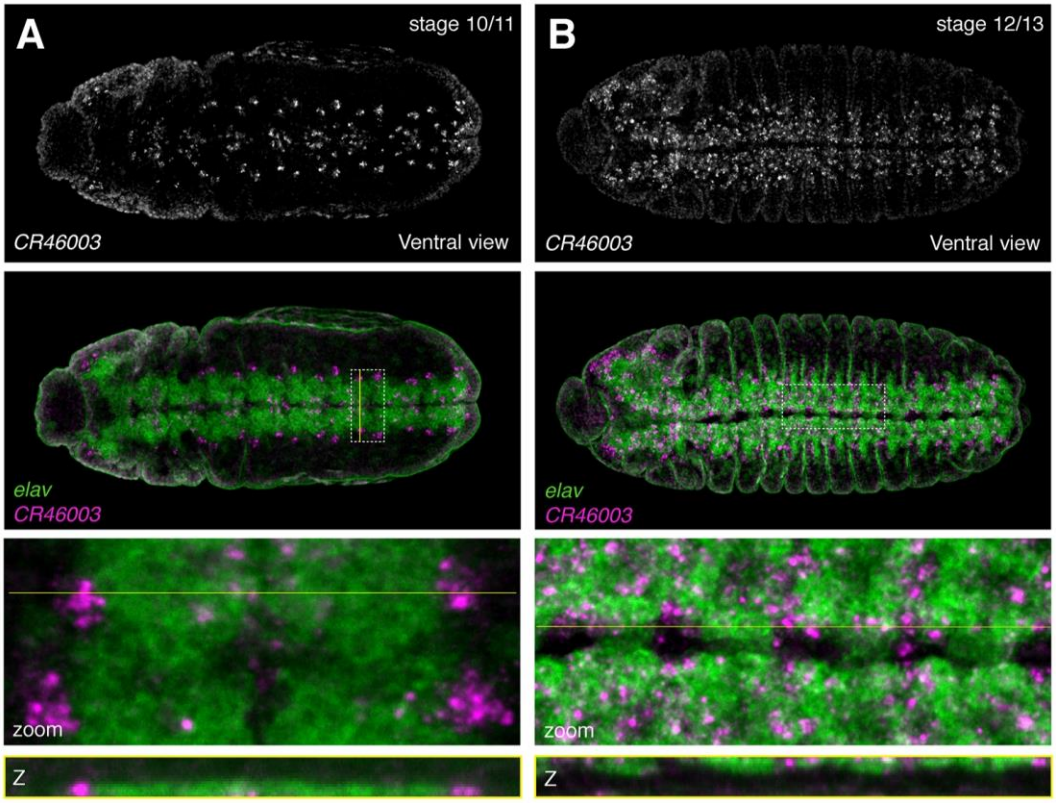


Fig S15: The lncRNA *CR46003* is expressed in neuronal subsets

RNA fluorescent in situ hybridization (RNA-FISH) against *CR46003* with the neuronal marker, *elav*. Ventral view. (A) Stage 10/11. (B) Stage 12/13. Top: *CR46003* alone. Second from top: *CR46003* (magenta) overlaid with *elav* (green). Dashed white box indicates region of interest (ROI) and yellow line indicates Z-slice through ROI. Second from bottom: zom-in of ROI. Bottom: Slice through Z-stack as indicated by yellow line.

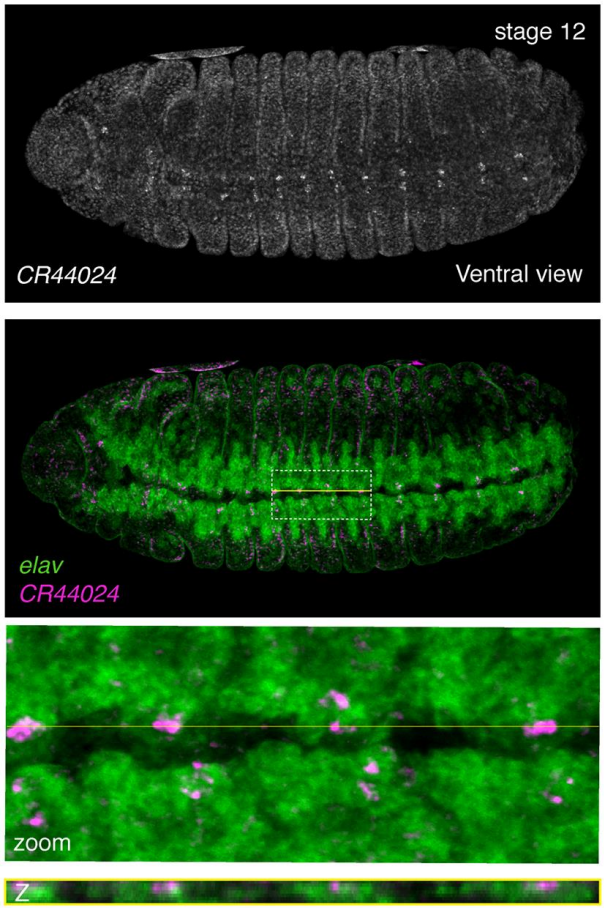


Fig S16: The lncRNA *CR44024* is expressed from stage 12 in neuronal subsets of the ventral nerve cord.

RNA fluorescent in situ hybridization (RNA-FISH) against *CR44024* with the neuronal marker, *elav*. Ventral view, stage 12. Top: *CR44024* alone. Second from top: *CR44024* (magenta) overlaid with *elav* (green). Dashed white box indicates region of interest (ROI) and yellow line indicates Z-slice through ROI. Second from bottom: zom-in of ROI. Bottom: Slice through Z-stack as indicated by yellow line.

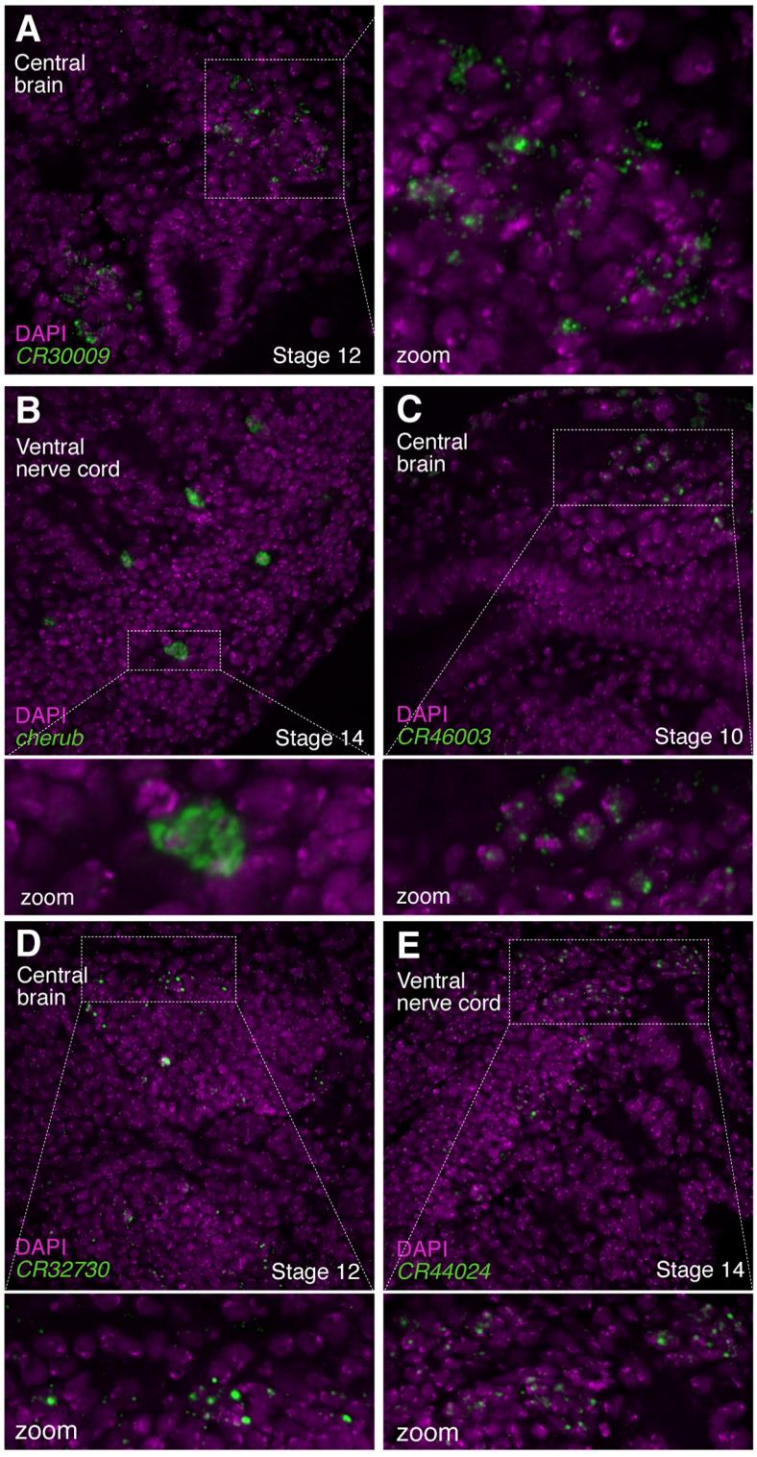


Fig S17: LncRNAs exhibit varying patterns of subcellular distribution

RNA fluorescent in situ hybridization (RNA-FISH) against lncRNAs (green) overlaid with the nucleic acid marker, DAPI (magenta). (A) *CR30009*; central brain, stage 12. (B) *cherub*; ventral nerve cord, stage 14. (C) *CR46003*; central brain, stage 10. (D) *CR32730*; central brain, stage 12. (E) *CR44024;* ventral nerve cord, stage 14. (A) Left panel: Dashed white box indicates region of interest (ROI). Right panel: zom-in of ROI. (B-E) Top panel: Dashed white box indicates region of interest (ROI). Bottom panel: zoom-in of ROI.

Supplementary References

Barolo, S., Castro, B. & Posakony, J.W., 2004. New Drosophila transgenic reporters: insulated P-element vectors expressing fast-maturing RFP. *BioTechniques*, 36(3), pp.436–40– 442.

Bhatt, D.M. et al., 2012. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell*, 150(2), pp.279–290.

Hrvatin, S. et al., 2014. MARIS: Method for Analyzing RNA following Intracellular Sorting K. Aalto-Setala, ed. *PLoS ONE*, 9(3), pp.e89459–6.

Karaiskos, N. et al., 2017. The Drosophila embryo at single-cell transcriptome resolution. *Science*, 358(6360), pp.194–199.

Markstein, M. et al., 2004. A regulatory code for neurogenic gene expression in the Drosophila embryo. *Development*, 131(10), pp.2387–2394.

Pawlicki, J.M. & Steitz, J.A., 2008. Primary microRNA transcript retention at sites of transcription leads to enhanced microRNA production. *The Journal of cell biology*, 182(1), pp.61–76.

Rosner, M. & Hengstschläger, M., 2012. Detection of cytoplasmic and nuclear functions of mTOR by fractionation. *Methods in molecular biology (Clifton, N.J.)*, 821(Chapter 8), pp.105–124.

Rubin, G.M. & Spradling, A.C., 1982. Genetic transformation of Drosophila with transposable element vectors. *Science*, 218(4570), pp.348–353.

Schmittgen, T.D. & Livak, K.J., 2008. Analyzing real-time PCR data by the comparative C(T) method. *Nature Protocols*, 3(6), pp.1101–1108.

Shechner, D.M. et al., 2015. Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nature Methods*, 12(7), pp.664–670.

Stapleton, M. et al., 2002. The Drosophila gene collection: identification of putative full-length cDNAs for 70% of D. melanogaster genes. *Genome Research*, 12(8), pp.1294–1300.

Vandesompele, J. et al., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology*, 3(7), p.RESEARCH0034.