Conserved transcription factors coordinate synaptic gene expression through repression

Authors and affiliations:
James A. Kentro[1,2,3], Gunjan Singh[1,2], Tuan M. Pham[1,2,4], Justin Currie[1,2,3], Saniya Khullar[5,6], Audrey T. Medeiros[2], Erica Larschan[1,3,*], Kate M. O'Connor-Giles[2,3,*]

1. Department of Molecular Biology, Cell Biology, & Biochemistry, Brown University, Providence, RI, USA
2. Department of Neuroscience, Brown University, Providence, RI, USA
3. Carney Institute for Brain Science, Brown University, Providence, RI, USA
4. Center for Computational Molecular Biology, Brown University, Providence, RI, USA
5. Waisman Center, University of Wisconsin-Madison, Madison, WI, USA
6. Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA

* Co-corresponding authors: erica_larschan@brown.edu, kate_oconnor-giles@brown.edu

The authors declare no competing interests.

Abstract

Chemical synapses are the primary sites of communication in the nervous system. Synapse formation is a complex process involving hundreds of proteins that must be expressed in two cells at the same time. How this spatiotemporal coordination is achieved remains an open question. We find that synaptic genes are broadly and specifically coordinated at the level of transcription across species. Through genomic and functional studies in *Drosophila*, we demonstrate corresponding coordination of chromatin accessibility and identify chromatin regulators DEAF1 and CLAMP as broad repressors of synaptic gene expression. Disruption of either factor causes increased synaptic gene expression across neuronal subtypes and excess synapse formation. We further find that DEAF1, which is linked to syndromic intellectual disability, is both necessary and sufficient to repress synapse formation. Our findings reveal the critical importance of broad temporally coordinated repression of synaptic gene expression in regulating neuronal connectivity and identify two key repressors.

Introduction

The nervous system is remarkable for its cellular diversity and functional complexity. In the human brain, billions of neurons of diverse subtypes form trillions of synapses that connect neurons into the functional circuits that interpret the environment and control behavior. There has been significant progress in understanding how the great diversity of neurons is generated[1–4]. Yet, how neurons acquire shared traits has been less studied.

The chemical synapses that underlie neuronal communication are hallmarks of mature neurons. Synapses are complex intercellular structures designed for the rapid transmission of information in the form of Ca^{2+}-dependent neurotransmitter release and uptake. Building a functional synapse necessitates the simultaneous expression of hundreds of proteins, including neurotransmitter transporters and receptors, regulators of synaptic vesicle release, and scaffolding and adhesion molecules, as well as the proteins that regulate their function and transport to synapses[5,6]. Adding to the challenge, the expression of synaptic genes, which we define broadly here as all genes that encode proteins required for the formation of functional synapses, must not only be coordinated within a given cell, but across synaptically connected neurons. How the spatio-temporal coordination of synaptic gene expression is achieved remains an open question. While this coordination can and likely does occur at multiple levels, including transcription, translation, and protein trafficking, our analysis of *Drosophila* time-series transcriptomic data indicates that hundreds of synaptic genes, including both common and neuronal subtype-specific synaptic genes, are co-regulated at the level of transcription. We show that this broadly coordinated regulation of synaptic genes is also conserved in mouse.

Here, we investigate the mechanisms and regulators that temporally coordinate the expression of synaptic genes across neurons. Consistent with prior studies implicating chromatin regulation in neuronal maturation[7,8], we find that chromatin accessibility is dynamically and coordinately regulated across large numbers of synaptic genes. Using motif analysis, chromatin binding data, and a new machine learning approach for predicting gene regulatory networks from single-cell multiomic data, we identify DEAF1, which is linked to intellectual disability in humans, and CLAMP, a pioneer factor that regulates chromatin accessibility, as predicted negative regulators of coordinated synaptic gene expression. DEAF1 and CLAMP exhibit enriched co-accessibility at synaptic genes and co-bind chromatin at a subset of their binding sites. Through *in vivo* studies, we find that DEAF1 and CLAMP function as direct repressors of a majority of synaptic genes that are directly bound by both factors. Finally, we find that loss of DEAF1 or CLAMP specifically results in excess synapse formation. Conversely, overexpression of DEAF1 blocks synapse formation, indicating an instructive role. Overall, we demonstrate that synaptic gene transcription is broadly co-regulated across species and define the critical importance of repressive mechanisms in sculpting synaptic connections.

Results

Broad temporal coordination of synaptic gene transcription

Coordinated expression of the hundreds of genes required to form functional synapses could occur at multiple levels. To determine if synaptic genes are broadly transcriptionally co-regulated, we examined the developmental transcriptional expression profiles[9] of well-characterized *Drosophila* genes that encode a variety of proteins critical for synapse formation and function and observed a shared expression pattern across embryonic, larval, and pupal development, suggesting coordinated regulation at the level of transcription. We used 20 of these synaptic genes (Figure S1A) to generate a 'synaptic' expression profile footprint and compared it to the developmental expression profiles of all *Drosophila* genes. Plotting the 485 genes with Spearman correlation scores above 0.8 (Table S1) reveals a similar pattern with two peaks of transcription correlating with peak synaptogenesis during embryonic and adult nervous system development (Figure 1A). All 485 genes are expressed in the nervous system and gene ontology (GO) analysis reveals a strong enrichment for synaptic terms (Figure 1B), consistent with coordinated transcriptional regulation of synaptic genes. In contrast, genes with correlation scores below 0.80 include few well-characterized synaptic genes and enrichment for diverse GO terms (Figure S1B).

Most of the 485 coregulated genes encode proteins with well-characterized roles in synapse formation and/or function, including the core proteins that promote Ca^{2+}-dependent synaptic vesicle release at all synapses and the subtype-specific synaptic proteins that regulate the synthesis, transport, and reception of glutamate, acetylcholine, gamma-aminobutyric acid (GABA), serotonin, dopamine, and octopamine/tyramine. Of the 485 co-regulated genes, 168 remain uncharacterized in *Drosophila* and are designated as computed genes (CGs). Among these are genes encoding homologs of ion channels, solute carriers, enzymes, and cytoskeletal proteins linked to synapses in other species as well as genes linked to rare neurodevelopmental disorders. Recent studies in *Drosophila* have identified roles in synapse formation and/or function for a number of previously uncharacterized, highly correlated genes, including TRMT9B/Fid[10,11] and PDZD8[12]. Together, these findings suggest a regulatory mechanism that temporally coordinates the expression of pan-neuronal and cell-type specific synaptic genes across the nervous system.

To assess the specificity of coordinated synaptic gene expression, we investigated genes involved in an earlier stage of neurodevelopment, specifically neurite outgrowth and axon development. As with synaptic genes, we created a composite profile of 20 well-characterized genes with roles in neurite development and axon guidance (Figure S1C). We then calculated a Spearman correlation score for all neuronally expressed genes. The 485 most correlated genes exhibit no overlap with the synaptic gene list and GO analysis reveals enrichment for neuron projection development, axon guidance, and broader developmental parent categories (Figure S1D, Table S2). Like synaptic genes, neurite/axon development genes exhibit two distinct peaks of expression. However, these peaks precede peaks in synaptic gene expression in both embryos and pupae (Figure 1A), consistent with the ordered expression of broad cohorts of functionally related genes across nervous system development.

To further investigate the specificity of the synaptic gene temporal transcriptional profile, we performed an unbiased k-means clustering analysis of all genes expressed in the nervous system based on temporal expression. We identified 12 clusters with distinct temporal expression patterns, one of which closely matches the synaptic temporal expression profile (Figure S1E, cluster 8, light blue). Cluster 8 overlaps with 324/485 coordinately regulated synaptic genes (Figure S1F; padj = $0x10^{00}$). The next most similar profile (cluster 1) overlaps with 85 coordinately regulated synaptic genes (padj = $6x10^{-32}$), while all other clusters have non-significant or zero overlap. We performed Gene Ontology (GO) analysis on the 12 clusters and found that cluster 8 was enriched in synaptic gene ontology categories (Figure S1G). In comparison, cluster 1 is enriched for more general neurodevelopmental GO categories, including axon development, and has earlier peaks of expression (Figure S1E). With clusters enriched for GO categories spanning early to late neuronal development, this analysis reveals a logical progression of gene expression programs across the nervous system, consistent with observations in mammals[13,14].

To independently validate the broad temporally coordinated transcription of synaptic genes, we analyzed time-series single-cell RNA-seq (scRNA-seq) datasets[15–17]. We clustered whole embryo data to obtain embryonic nervous system cells at six timepoints for analysis along with five larval whole brain and nine pupal visual system timepoints. We then plotted the average synaptic gene expression level across timepoints and observed a temporal expression profile that matches our bulk analysis (Figure 1C,D).

We next investigated whether the broadly coordinated expression of pan-neuronal and subtype-specific synaptic genes is conserved in more complex mammalian brains with longer developmental windows. We examined time-series RNA-seq data from the developing mouse brain beginning at embryonic day 10.5[18]. To generate an analogous set of synaptic genes, we created synaptic expression profile footprints for stages E10.5 through E12.5 in whole brain and E13.5 through P0 in forebrain and hindbrain using the orthologs of the synaptic genes used in flies. By comparing these footprints to the temporal expression patterns of all genes in the respective data sets, we identified 742 genes with correlation scores above 0.95 that are expressed together starting at E12 and continuing to rise through embryonic development (Figure 1E). Synaptic gene expression in the hindbrain tapers off earlier than in the forebrain, consistent with regional differences in developmental timing. GO analysis revealed synaptic development and transmission as the most significant categories for genes within this gene set (Figure 1F). As in flies, these genes encode both pan-neuronal and a diverse array of subtype-specific synaptic genes (Table S3). Therefore, broad temporal transcriptional regulation of synaptic gene expression is a conserved feature of nervous system development.

Chromatin accessibility is dynamically and coordinately regulated across synaptic genes

To understand how the expression of hundreds of synaptic genes, which are widely dispersed across chromosomes (Figure S2A), is temporally coordinated, we analyzed existing DNase-seq time-series data from isolated neuronal and mesodermal embryonic tissue to determine if chromatin accessibility is dynamically regulated at synaptic genes[19]. In early developmental stages, promoter regions upstream of synaptic genes are closed relative to a random set of

nervous system expressed genes (Figure 2A, blue vs. grey). Chromatin becomes substantially more accessible at synaptic genes relative to random neuronal genes beginning at 10-12 hours, immediately preceding the coordinated upregulation of synaptic gene transcription (Figure 2B). Using time-series proteomic data available for a subset of synaptic proteins[20], we also observe a coordinated increase in synaptic protein abundance immediately following chromatin opening and transcription initiation, suggesting that the coordinated increase in chromatin accessibility also influences protein production (Figure S2B). This dynamic pattern of chromatin accessibility is specific to the nervous system (Figure S2C). In contrast to synaptic genes, chromatin at the 485 neurite development genes identified above is accessible by 4-6 hours of embryogenesis across cell types (Figure 2A, orange). Therefore, chromatin accessibility across hundreds of synaptic genes precedes their coordinated transcription in a tissue- and stage-specific manner, consistent with findings for other groups of coordinately regulated genes[19,21].

We next investigated chromatin accessibility at these newly accessible sites in later stages of development using larval, pupal, and adult neural ATAC-seq data[22,23]. Across synaptic genes, chromatin accessibility remains broadly accessible throughout these stages although there is modestly decreased accessibility at the 3rd instar larval stage consistent with repression of gene expression (Figure 2C). Notably, the vast majority (95%) of dynamically accessible sites identified near synaptic genes remain accessible in third instar larval brains, indicating that repression of synaptic gene expression at this stage is achieved with only modest changes in chromatin accessibility, possibly to enable rapid activity-dependent expression of synaptic genes at these stages.

DEAF1 motifs and binding are enriched at dynamically regulated accessible regions near synaptic genes

To define transcription factors (TFs) that coordinately regulate synaptic gene expression, we first defined newly accessible regions in 10-12 hour neurons within a 3 kilobase window around each synaptic gene and identified enriched TF binding motifs within these regions. This analysis revealed 21 enriched motifs, with the DEAF1 binding motif most significantly enriched (Figure 2D). Consistent with a specific role at synaptic genes, DEAF1 motif accessibility increases only in central and peripheral neurons (Figure 2E). DEAF1 is a highly conserved, neuronally enriched TF[24–26] that is associated with developmental delay and intellectual disability[27,28]. DEAF1 directly binds DNA through its SAND domain and interacts with positive and negative co-regulators through its MYND domain[29,30].

Multiple TFs are often capable of binding to the same motifs with varying levels of affinity and many sites remain unbound, so to refine our list of candidate regulators we next examined available chromatin binding data for over 450 TFs[31,32]. We compared binding peaks at newly accessible regions within 3kb of synaptic genes to those near three random control sets of neuronally expressed genes to identify TFs with enriched binding at newly accessible regions near synaptic genes. Of all 450 TFs, DEAF1 showed the largest difference in average profile peak height at newly accessible regions near synaptic relative to random neuronally-expressed genes (Figure 2F). DEAF1 binds at the most newly accessible regions around synaptic genes

(~50%) and the largest percentage of synaptic genes (75%, Tables S3,4). Therefore, we hypothesized that DEAF1 directly coordinates synaptic gene expression.

DEAF1 and CLAMP are predicted to co-repress synaptic gene expression

In parallel, we took an independent machine learning approach to identify candidate regulators of coordinated synaptic gene expression using a recently developed tool called NetREm[33]. NetREm analyzes single-cell RNA-seq (scRNA-seq) expression data and incorporates protein-protein interaction (PPI) data as a network constraint among the TF predictors along with binding motif accessibility at target genes from single-cell ATAC-seq (scATAC-seq) data to generate weighted correlation scores between TF-TF pairs. By linking multiple TF-TF pairs, NetREm identifies broader networks that cooperatively or antagonistically co-regulate the expression of target genes.

We first clustered wild-type cells from the recently published scRNA-seq and scATAC-seq third instar larval brain atlas[34]. We annotated clusters broadly to capture substantial differences in synaptic gene expression during neuronal development: neural progenitors, where synaptic genes are inactive; immature neurons, which express only low levels of some synaptic genes; mature neurons, which broadly express synaptic genes; and glia, which were excluded from further analysis (Figure 3A,B). As input for NetREm analysis, we included the 15 TFs bound at the most newly accessible regions near synaptic genes (Table S3) and all TFs identified by the motif analysis above (see Figure 2C) and a second independent motif analysis with a tool that did not include the DEAF1 motif in its database (see methods, Figure S3A) for a total of 149 TFs. NetREm identified 10 TFs predicted to regulate at least 280 synaptic target genes, including DEAF1, which was predicted to regulate 294 targets (64.3%; Figure S3B).

Since DEAF1 is highly enriched at newly accessible sites near synaptic genes (see Figure 2C,D) and among the top candidates that we identified with NetREm, we conducted a second NetREm analysis to predict a DEAF1 co-regulatory network (Figure 3C). Four TFs have consistently high coordination scores (above |0.90|): computed gene CG4360, Chromatin-linked adaptor for MSL proteins (CLAMP), Medea (Med), and NFAT nuclear factor (NFAT). Of these factors, CLAMP and NFAT, also have experimentally verified physical interactions with DEAF1. DEAF1 and NFAT are predicted to act antagonistically in the regulation of synaptic genes, while DEAF1 and CLAMP are predicted to act cooperatively. CLAMP is a functional analog of mammalian CTCF that binds condensed chromatin and regulates the ability of other factors to bind DNA[35]. Motif analysis also revealed CLAMP motifs are the second-most enriched at dynamically regulated sites near synaptic genes (see Figure 2C). Notably, DEAF1 and CLAMP are predicted to function together as negative regulators of synaptic gene expression.

To independently evaluate the predicted relationship between DEAF1, CLAMP and synaptic genes, we analyzed a scRNA-seq and scATAC-seq atlas spanning embryogenesis[15]. Following clustering and annotation of scRNA-seq from embryonic hours 12 to 16 (Figure 3D, S3C-F), we compared *deaf1*, *clamp*, and synaptic gene expression in neural progenitors, immature neurons, and mature neurons. We found that while the expression of both *deaf1* and *clamp* correlate both positively and negatively with a large number of neuronal genes, both factors are strongly anti-

correlated with synaptic genes (Figure 3E), consistent with the NetREm prediction that the two factors function together as repressors of synaptic gene expression.

We next clustered the corresponding scATAC-seq data (Figure 3F) and analyzed neural progenitors, immature neurons, and mature neurons to determine if the accessibility of DEAF1 and CLAMP binding motifs is coordinated at synaptic genes. We grouped regions into sets of consistently co-accessible regions and identified Differentially Accessible Regions (DARs) across cell-types. The DEAF1 binding motif was accessible at 308 synaptic genes (63.5%), the CLAMP binding motif was accessible at 230 synaptic genes (47.4%), and 174 synaptic genes (35.9%) had accessible motifs for both DEAF1 and CLAMP (Fig 3G). Of these 174 synaptic genes, 87 have spatially neighboring co-accessible motifs (Table S5, see Methods). Genes with co-accessible motifs include pan-neuronal synaptic active zone components and subtype specific components of excitatory and inhibitory neurotransmitter systems (Figure 3H, S3G,H). These data are consistent with the model that DEAF1 and CLAMP bind simultaneously to a subset of DARs at synaptic genes to directly regulate synaptic gene expression.

DEAF1 and CLAMP colocalize at synaptic genes

CLAMP directs distinct TFs and chromatin regulators to the correct genomic locations to drive context-specific functions, including directing dosage compensation[36] and splicing complexes[37]. Therefore, we hypothesized that colocalization of DEAF1 and CLAMP at synaptic genes may be necessary to repress synaptic gene expression outside of developmental windows of peak expression. DEAF1 and CLAMP have each previously been found to bind to hundreds of sites on polytene chromosomes[38–41], so we co-stained polytene chromosomes to visualize their colocalization across the genome in a tissue in which synaptic genes are repressed (Figure 4A). We observe DEAF1 and CLAMP colocalization at a subset of sites across all chromosomes, consistent with their combined binding defining a subset of co-regulated genes.

The resolution of polytene chromosomes does not allow mapping of specific target genes, so we again turned to DNase-seq accessibility data[19] in the 10-12 hour nervous system, when chromatin at synaptic genes first becomes accessible, to determine if co-accessibility of DEAF1 and CLAMP motifs is enriched at synaptic genes. Measured separately, the accessibility of DEAF1 or CLAMP motifs is approximately 2-fold more enriched at synaptic genes than at random neuronally expressed genes (padj = $3x10^{-8}$ and $3x10^{-7}$, respectively). Notably, accessible peaks containing both motifs (co-accessibility) are 4-fold more enriched at synaptic genes than at other neuronally expressed genes (Figure 4B; padj = $1x10^{-41}$). This 2-fold enrichment of co-accessible vs. individually accessible DEAF1 and CLAMP motifs at synaptic genes further supports the model that DEAF1 and CLAMP coordinately regulate synaptic gene expression.

We next investigated DEAF1 and CLAMP ChIP-seq data[32,42] for co-binding and observed significant overlap at synaptic genes. (Figure 4C). Analysis of ChIP-seq data reveals DEAF1 binding is enriched at coordinately regulated synaptic genes (~75%, padj = $2x10^{-11}$, Fisher's exact test, with Benjamin-Hochberg correction), as is CLAMP binding (~40%, padj = $2x10^{-11}$, Fisher's exact test, with Benjamin-Hochberg correction). At 85% of synaptic genes bound by

CLAMP, we also observe binding by DEAF1 (~34% of synaptic genes, padj = $2x10^{-23}$, Fisher-BH). Together these data suggest that DEAF1 and CLAMP co-regulate synaptic gene expression through co-binding at dynamically accessible sites.

deaf1 or _clamp_ knockdown de-represses synaptic gene expression

Based on our multimodal analysis, we hypothesize that DEAF1 and CLAMP function together to broadly repress synaptic gene expression outside of windows of peak synaptogenesis. If our model is correct, loss of either factor would be expected to dysregulate synaptic gene expression. To test our model *in vivo*, we focused on the third instar larval stage, when synaptic gene expression is normally repressed. We used *elav*-Gal4, which drives expression in all neurons and salivary glands, to express RNAi constructs targeting *deaf1* or *clamp* and confirmed knockdown of each factor on chromatin (Figure S4). We then isolated central brains and ventral ganglia and conducted RNA-seq. Knock down of *deaf1* resulted in the differential expression of 2862 genes relative to RNAi controls, 1558 upregulated and 1304 downregulated. GO analysis of genes significantly upregulated following *deaf1* knockdown revealed enrichment for neurodevelopment terms, including synaptic organization and synaptic signaling (Fig 5A). Consistent with broad DEAF1-mediated repression of synaptic genes, every gene annotated to the “maintenance of presynaptic active zone structure” or “regulation of calcium ion transmembrane transporter activity” terms is significantly upregulated. In contrast, GO enrichment analysis of genes downregulated upon *deaf1* knockdown yielded broad metabolic process terms, including peptide biosynthesis and lipid oxidation (Figure S5A).

Out of 485 coordinately regulated synaptic genes, 207 are upregulated upon *deaf1* knockdown, while only 23 are downregulated (Figure 5C). Among the upregulated genes are those that encode: (1) active zone components, including nSyb, brp, Rim, Syt1, Ca-β; (2) synaptic cell adhesion molecules, including multiple Dpr, Side, Beat, Nlg, CadN proteins; and (3) neurotransmitter transporters and receptors, including vesicular glutamate and monoamine transporters (VGlut and Vmat), Choline acetyltransferase (ChAT), and receptors for acetylcholine, GABA, glutamate, dopamine, as well as the serotonin receptors 5-HT1A and 5-HT1B. Interestingly, DEAF1 has been implicated as a key regulator of 5-HT1A expression in human cell lines[43,44], indicating conserved roles for DEAF1 between *Drosophila* and humans. These data confirm that DEAF1 is a broad negative regulator of synaptic gene expression.

Knockdown of *clamp* in all neurons resulted in the differential expression of 1566 genes relative to RNAi controls, 861 upregulated and 703 downregulated. Similar to *deaf1* knockdown, of 485 synaptic genes, 106 were upregulated and only 27 downregulated (Figure 5C). GO analysis of all upregulated genes shows enrichment of terms for synaptic signaling, calcium transport and homeostasis, and neurotransmitter transport (Figure 5B). In contrast, downregulated genes are enriched for protein catabolism and intracellular transport (Figure S5B). We also observe GO term enrichment of upregulated genes for organic acid metabolism and glucose metabolism, suggesting that CLAMP also has a broader role in negatively regulating gene expression in the nervous system. We observe a similar upregulation of synaptic genes in *clamp* null third instar larval brains[45]. Synaptic genes upregulated upon *clamp* knockdown include 63 genes also upregulated when *deaf1* is knocked down, including genes encoding both cell-type specific

synaptic proteins such as Vmat, ChAT, and receptors for acetylcholine, glutamate, and dopamine as well as common synaptic proteins such as the vesicular calcium sensor Synaptotagmin. Together, these findings demonstrate that disruption of *deaf1* or *clamp* broadly de-represses synaptic gene expression in the larval nervous system.

To address whether DEAF1 and CLAMP regulation of synaptic genes is direct, we assessed the overlap between differentially expressed genes and chromatin binding[32,42]. We observe a striking enrichment of DEAF1 and CLAMP binding at upregulated, but not downregulated genes (Figure 5D). Together, these results demonstrate that DEAF1 and CLAMP directly repress synaptic gene expression across neuronal subtypes.

DEAF1 and CLAMP restrain synapse formation

Synapse formation and refinement are critical developmental processes. In wild-type animals synaptic gene transcription peaks during embryonic nervous system development, then falls to very low levels following this window of robust synaptogenesis (see Figure 1A). This observation suggests that both coordinated activation and broadly coordinated repression of synaptic genes may regulate proper nervous system development.

To determine the role of DEAF1/CLAMP-dependent synaptic gene repression in synaptogenesis, we turned to the well-characterized, highly stereotyped third instar larval neuromuscular junction. Here, glutamatergic motor neurons innervate single muscle cells and form a stereotyped number of synaptic boutons, each containing ~10 individual synapses, allowing us to readily quantify synapse formation in different genetic backgrounds using the neuronal membrane marker HRP and antibodies against the presynaptic active zone protein BRP/CAST/ELKS. Pan-neuronal knockdown of *deaf1* caused a 56% increase in synaptic bouton number and significantly increased active zone number compared to control (Figure 6A-E). We also observe a 54% increase in synaptic boutons in a heteroallelic combination of *deaf1* alleles (*deaf1*$^{k3/S10B}$) that produces rare escapers that survive to third instar larvae (Figure S6A-B). Similarly, pan-neuronal knockdown of *clamp* resulted in a 28% increase in synaptic bouton formation and significantly increased active zone number (Figure 6F-I). In *clamp*^{KG00473} hypomorphs, we observe a 24% increase in synaptic boutons (Figure S6C-D), similar to previous observations[46]. Thus, transcriptional repression of synaptic gene expression by DEAF1 and CLAMP is required to constrain synaptic growth during nervous system development.

Our findings suggest that activation and repression of synaptic gene transcription is dynamically controlled through the temporally regulated actions of TFs acting in opposite directions. This further suggests that dysregulated repression of synaptic genes might inhibit synapse formation. To determine whether DEAF1 misexpression is sufficient to repress synapse formation, we overexpressed *deaf1* in developing neurons and quantified synapse formation at the larval NMJ. We observed a 19% decrease in bouton formation (Figure 6F-H) and striking 52% decrease in active zones (Figure 6I). Thus, the dynamic regulation of synaptic gene activation and repression plays a critical role in sculpting neuronal connectivity during nervous system development.

Discussion

Combining multimodal genomic analysis with predictive modeling with *in vivo* loss- and gain-of-function validation, we reveal a central role for coordinated repression of synaptic genes across the nervous system in controlling synaptogenesis and identify DEAF1 and CLAMP as key repressors.

Neuronal connectivity is tightly controlled through the regulation of synapse formation and maintenance during development. Altered synapse number is a common feature of neurodevelopmental disorders, highlighting the importance of proper regulation. Synapse formation requires the coordinated function of hundreds of proteins, suggesting that synaptic gene expression must be coordinately regulated. Indeed, we find that large numbers of synaptic genes, including those encoding pan-synaptic and subtype-specific synapse components, are co-regulated at the level of transcription. These same synaptic genes are also coordinately activated in mammals, consistent with previous observations of temporal waves of gene expression programs during nervous system development[13,14]. We observe a widespread increase in chromatin accessibility upstream of the majority of synaptic genes over a short window immediately preceding the coordinated transcription of synaptic genes, indicating a key role for dynamic chromatin regulation. In addition to coordinated activation, we observe temporally coordinated repression of synaptic genes following developmental windows of peak synaptogenesis. However, newly accessible chromatin at synaptic genes remains accessible during this period, suggesting these sites may be occupied by multifunctional factors.

Through motif analysis, we identified DEAF1 as the TF most associated with dynamically accessible regions near synaptic genes. DEAF1 interacts with an array of co-regulators that both positively and negatively regulate transcription in different contexts[28,47,48]. Through analysis of single-cell transcriptomic and chromatin accessibility data with NetREm, a new network analysis tool that incorporates protein-protein interaction data, we identified CLAMP as a DEAF1 co-regulator. Notably, CLAMP motifs are the second-most enriched at dynamically regulated sites upstream of synaptic genes. Expression of both DEAF1 and CLAMP are strongly negatively correlated with synaptic gene expression and are predicted to function together in the repression of synaptic gene expression. Consistently, we observe DEAF1 and CLAMP colocalizing on polytene chromosomes and co-binding at a significant fraction of synaptic genes.

While multiple TFs required for the activation of gene expression programs in neurodevelopment have been identified[1,46,49–52], negative regulators have been more difficult to identify. To determine if DEAF1 and CLAMP are negative regulators of synaptic gene expression, we knocked down each factor and conducted RNA-seq at a developmental time point when synaptic gene expression is normally repressed. In both cases, we observed upregulation of synaptic gene expression, confirming their role as repressors of synaptic gene expression. Furthermore, both factors are directly bound at genes upregulated in the absence of each factor genes but not downregulated genes, demonstrating a direct role in synaptic gene repression. Both pan-synaptic and subtype-specific synaptic genes are upregulated, indicating a broad role in repressing synaptic gene expression across neuronal subtypes. Interestingly, a

recent study of dynamically accessible enhancers in mouse nervous system development found that only enhancers positively correlated with gene expression were cell specific[53]. Together with our findings, this suggests a regulatory logic in which negative regulators of synaptic gene expression act more broadly than positive regulators, which may be neuronal subtype specific.

A key question to be addressed in the future is how activation and repression of synaptic genes are coordinated. Our findings indicate that chromatin accessibility is dynamically regulated in embryos and then remains largely open. This constitutively open chromatin may facilitate synaptic plasticity in the brain by allowing for the rapid activity-dependent transcriptional activation of synaptic genes. One possibility is that positive regulators outcompete DEAF1 and CLAMP during activation. We identified several candidate TFs that might play such a roles, including: (1) Trl (GAGA factor), a GA-binding TF similar to CLAMP that has been very well-studied[54–56] as a transcriptional activator and competes with CLAMP for GA-rich binding sites involved in dosage compensation[40] and histone gene regulation[38] and (2) Onecut, which was recently identified as a positive transcriptional regulator of pan-neuronal genes in *C. elegans*[57,58]. Alternatively, regulation of DEAF1 and/or CLAMP themselves could modulate the transition between negative and positive regulation of synaptic gene expression. DEAF1 can both positively and negatively regulate 5-HT_{1A} transcription, with GSK3-dependent phosphorylation promoting its repressive role[43,44]. Therefore, it is possible that DEAF1 remains at its binding sites during activation and plays a permissive role, with its function modulated by phosphorylation state – a model that will be important to test in the future.

Finally, we demonstrate that repression of synaptic gene expression plays an instructive role in synapse formation. Targets of repression include pan-synaptic genes as well as cell-type specific neurotransmitter transporters and receptors consistent with a broad role across cell types. The neurotransmitters released and interpreted by neurons are a central part of their identity and it has long been thought that neuronal identity is conferred by cell-specific combinations of TFs termed terminal selectors[1]. The model that repression of synaptic gene expression before and after developmental windows of synaptogenesis occurs broadly across cell types, while activation of synaptic gene expression relies on the action of cell-specific terminal selectors reconciles these two findings. This model is also consistent with studies in *C. elegans* demonstrating that pan-neuronal gene expression is regulated by a diversity of redundant factors[52]. Recent findings in the developing mouse cortex provide compelling support. There, a similar analysis of dynamic gene expression and chromatin accessibility in single cells revealed that dynamically accessible enhancers associated with positive regulation of gene expression tend to be regulated in a cell-type manner while enhancers associated with negative regulation act more broadly[53]. Loss of either DEAF1 or CLAMP results in excess synapse formation, demonstrating the critical role of synaptic gene repression in sculpting synaptic connections. Conversely, neuronal overexpression of DEAF1 blocks synapse formation. Together, these findings reveal the importance of dynamic modulation of positive and negative regulatory factors across development. Both dominant and recessive pathogenic *DEAF1* alleles cause syndromic intellectual disability [27,59,60], underscoring the importance of decoding the regulatory logic of synaptic gene expression.

Materials and methods

Drosophila stocks:

The following fly lines used in this study were obtained from the Bloomington Drosophila Stock Center (BDSC, NIH P40OD018537): *w*^{1118} (RRID:BDSC_5905), *UAS-deaf1-RNAi* (RRID:BDSC_32512), *deaf1*^{k3} (RRID:BDSC_26279), *deaf1*^{S10B} (RRID:BDSC_91626), *Clamp*^{KG00473} (RRID:BDSC_14352), *UAS-GFP-RNAi* (RRID:BDSC_9330), *pros-Gal4* (RRID:BDSC_80572), *elav*^{C155}*-Gal4* (RRID:BDSC_458). Additional fly lines used in this study were obtained from The Zurich ORFeome Project (FlyORF)[61]: UAS-*deaf1-3xHA* (RRID:Flybase_FBst0501296). *Drosophila melanogaster* stocks were raised on molasses food (Lab Express, Type R) in a 25°C incubator with controlled humidity and 12h light/dark cycle.

Temporal expression profile analysis

Transcriptomic data was accessed through FlyBase[9,62] from Graveley et al., 2011, *Nature*. Data was separated by developmental timepoints. Embryonic 0-2h through pupal p15 developmental timepoints were used for finding Spearman correlations to gene sets with the average profile of 20 synaptic genes using cor() in the R stats package[63]. Tissue-specific data was used to remove any genes not found to be expressed in CNS of P8 pupa or L3 larva. Random control gene sets were collected from genes expressed in the nervous system with Spearman correlations to synaptic genes below 0.80 and with sample_n() in the dplyr package[64].

Clustering analysis

Kmeans clustering was performed with the R stats package. The number of clusters was determined from the gap statistic calculated with packages cluster[65] and factoextra[66]. Gene ontology analysis was performed with the R package gprofiler2[67] and sorted into parent categories with rrvgo[68].

Single-cell temporal profile analysis

Using Seurat[69], all nervous system annotated clusters were subset from 2 hour embryonic datasets[15], as Seurat objects, then merged. Larval nervous system datasets[16] and pupal visual system datasets[17] from each timepoint were loaded as count matrices. Seurat objects were created and filtered with min.cells = 3, min.features = 200, percent.mt < 5, then merged. Each merged object was normalized, 5000 variable genes were used for principle component analysis (PCA) and 50 PCs were used to identify nearest neighbors and RunUMAP with clustering resolutions of 0.47 for embryo and 0.6 for larva and pupa. The R package scCustomize[70] was used with AddModuleScore to average the expression of all synaptic genes in our list for each cell, and the score of each cell was averaged per timepoint.

Mouse temporal profile analysis

We identified orthologs of the 20 *Drosophila* synaptic expression profile footprint genes ranked as high or moderate according to the DRSC Integrative Ortholog Prediction Tool[71], finding at least two orthologous mouse genes for each fly gene. We used these orthologs to generate a mouse synaptic expression profile footprint for stages E10.5 through E12.5 in whole brain and E13.5 through P0 in forebrain and hindbrain. These data were then analyzed as described above for *Drosophila* to identify coordinately expressed genes.

DNase-seq analysis
DNase-seq bigwig files were downloaded from the Furlong laboratory website. Average profiles of accessibility at synaptic genes were made using deepTools computeMatrix with parameters 'computeMatrix reference-point --referencePoint TSS -a 1000 -b 1000 --binSize 100 --skipZeros --smartLabels --sortRegions descend' and plotProfile '--plotType se --yMin 0 --yMax 350''.

Protein expression data analysis
Protein expression data from Casas-Vila et al., 2017, Genome Research (supplemental table S1). The imputed.log2.LFQ.intensity data for each protein were averaged across the four samples at each timepoint. The data were then subset to synaptic proteins, scores were shifted by the minimum average imputed score and the score of each gene normalized to its maximum expression value.

Motif discovery
DNase-seq bigwig files were converted to bedgraph format with bigWigToBedGraph[72]. Peaks were called with MACS2[73] bdgpeakcall with a cutoff above the standard deviation and parameters '--max-gap 10 --min-length 150'. The neuronal 10-12 hours peak file was subset using BEDTools[74] 'subtract -A' to remove accessibility peaks intersecting those present in neuronal 8-10 hours. Likewise neuronal 10-12 hours peaks were removed from neuronal 8-10 hours peaks to use as background control regions in motif analysis. Peaks newly accessible in the neuronal 10-12 hours data were subset to those within 3kb of synaptic genes or within 3kb of three random control gene sets using bedtools 'window -w 3000 -u'. Homer[75] findMotifGenome.pl was used for motif analysis with the parameters 'dm3 -len 8,10,12 -size given -p 16 -dumpFasta'. The newly accessible neuronal 10-12 hours peak regions were used as input with the peaks unique to neuronal 8-10 hours data as background controls. A second analysis was done without inputting a background region set, defaulting to randomized control sequences. Additional analyses were performed with MEME Suite/5.5.5[76] XSTREME using default settings with Combined Drosophila motif databases and expected motif site distribution set to any number of repetitions. Also, MEME and TomTom were used with MEME parameters '-dna -objfun classic -mod anr -revcomp -nmotifs 21 -maxw 15' and TomTom with default parameters searching the Combined Drosophila motif databases plus JASPAR2018_CORE_insects databases.

Chromatin binding analysis
DNA binding data was downloaded from ENCODE or NCBI GEO databases. From the ENCODE database, files matching DNA binding assay type, *Drosophila melanogaster* organism, and available file types bed narrowPeak and bigWig were downloaded for analysis. For individual datasets from NCBI GEO, preprocessed bigwig, bedgraph, or peak bed files were downloaded for analysis. Average profiles were made from the bigWig files with deepTools computeMatrix reference-point either centered on newly accessible regions within 3kb of synaptic genes or random control gene sets or centered on the TSS of the same gene sets, with parameters '-a 3000 -b 3000 --skipZeros --smartLabels --sortRegions descend' and plotHeatmap with default settings. DNA binding peaks annotated to accessible regions within

3kb each gene set were identified with BEDTools[74] 'intersect -u -sortout', unique intersections were counted and percent of total accessible regions in each set was calculated.

Single-cell analysis

Single-cell data processing for NetREm input

Third instar larval brain scRNA-seq and scATAC-seq data[34] were processed with Seurat[69] and Signac[77], objects were created from counts matrix then filtered, normalized, and clustered using the parameters detailed in Mohana et al.[34]. Genes with highest average expression in each scRNA cluster (gene activity scores for scATAC), were used with enriched marker genes from FindAllMarkers(seurat_obj, only.pos = TRUE, min.pct = 0.25, logfc.threshold = 0.25, test.use = "wilcox") to broadly annotate the 16 RNA and 19 ATAC clusters, which were then merged by annotation.

NetREm

NetREm is a tool that predicts TF-TG regulatory networks (complementary gene regulatory network (GRN)) and TF-TF coordination networks (at the TG-level and overall) for the cell-type in a given context. The TF-TG regulatory network scores TFs by their ability to predict target gene (TG) expression levels. The TG-specific TF-TF coordination B network analyzes how TFs coordinate antagonistically or cooperatively to co-regulate the given TG. The overall TF-TF coordination network, B^_, summarizes the net behavior among TFs to co-regulate TGs overall. As input, NetREm requires gene expression data (single-cell or bulk-level), a prior protein-protein interaction (PPI) network among the predictors. We applied NetREm according to the methods described in the preprint[33], with the following specifications.

For NetREm TF input, we included 15 TFs with the highest percentage of DNA binding intersections in accessible peaks at synaptic genes plus all 141 TFs identified from motif analysis, with 7 overlapping for a total of 149 TFs. Of these, 135 were identified in the 3rd instar larval brain scRNA-seq expression data. For TGs we included all 485 genes from our synaptic genes list of which 457 were identified in the larval data. For PPI Network data, we used the full network of combined scores from Drosophila melanogaster STRINGdb version 12. STRINGdb v12 was subset to include 2,046 unique TF-TF edges of known direct and/or indirect PPIs, comprising 133 of the 135 default TFs examined in this analysis, E(spl)mbeta-HLH and E(spl)mgamma-HLH were not identified in the PPI data and given artificial edge weights of 0.01.

We derive prior reference GRN information that predicts TFs that associate to regulatory elements to regulate TGs. To do this, we utilize various available multiomics data to identify candidate TFs that may bind to regions of open chromatin linked to those TGs. For TG accessibility regions, we identified all chromatin accessibility peaks in the 3rd instar larval brain scATAC-seq data that were found within 3kb of synaptic genes and subset them according to the four cell annotations. We performed motif analysis on these accessible regions using motifmatchr[78] matchMotifs function with the JASPAR2024 motif database and BSgenome.Dmelanogaster.UCSC.dm6 genome and output scores for each peak. We retained the top 50% TFs for each peak based on the highest scores. Only 96 of the 135 TFs were identified by JASPAR2024; thus, we queried additional motif databases using query(MotifDb[79],

andStrings=c("Dmelanogaster"), orStrings=c("jaspar2024", "Dmelanogaster"), notStrings="ctcfl"). These additional motifs are drawn from Cisbp_1.02, FlyFactorSurvey, JASPAR_2014, JASPAR_CORE, Jaspar2016, Jaspar2018, and Jaspar2022. To identify the motifs in the accessibility sequences, we used monaLisa findMotifHits function with parameters 'query=pwm_list, subject=seqs, min.score="75", method="matchPWM", BPPARAM=BiocParallel::SerialParam()'. We found 112 of the 135 TFs using MonaLisa (min. score percentile 75), with 91 TFs in common with the JASPAR2024 database for a total of 117 TFs with some accessible motif. We thus derive our motif-based reference GRN for these TGs in cell-types that have identified accessible regions or motifs.

Then, for the TGs in our motif-based reference GRN, we input the TG-specific cell-type TFs to NetREm as their respective candidate TFs. For the remaining TGs (with unmapped prior GRN information), we use the 135 default TFs; these include 131 in glia, 74 in neural progenitors, 140 in immature neurons, and 97 TGs in mature neurons. Overall, we run NetREm step-by-step for each of these TGs for their corresponding set of TG-specific TFs (if TG is in motif-based reference GRN) or default TFs (if TG is not in reference GRN).

We run NetREm step-by-step for each TG with the following parameters. We randomly selected 80% of the samples in the gene expression data for training (M cells) the model and the remaining 20% for testing. We selected $\beta = 100$ for our PPI network prior constraint to stress identifying groups of functionally coordinating TFs for regulating the TG, and emphasizing known direct and/or indirect PPIs among the TFs. We trained the model on the training gene expression data. The sparsity parameter α is selected using LassoCV with 5-fold cross-validation to optimize the mean square error on training data. We do not fit any y-intercept.

Further, we annotate NetREm's output overall cell-type TF-TF coordination links for TF-TF links in our input PPI network as potential physical interactions if they are known and are found in Drosophila melanogaster STRING v12 database subsetted for direct, physical PPIs. Otherwise, we surmise that these known TF-TF PPIs may likely be via indirect mechanisms. Since few synaptic genes are expressed in glial cells, glial analysis was removed from final comparisons.

Embryo single-cell expression analysis

Embryo scRNA-seq data[15] was initially processed as described above in *Single-cell pattern analysis*, then annotated broadly based on genes with highest average expression per cluster and enriched marker genes. These data were then subset to only neural progenitors, immature neurons, and mature neurons from 06 to 18h timepoints. To compensate for dropout common to all single-cell experiments, we used cell-to-cell similarities based on Pearson correlations of normalized gene expression patterns, performed Network Enhancement[80,81] to denoise the resulting matrix, and multiplied by the original counts matrix. We evaluated the imputed results with cluster analysis, and preservation of marker gene specificity for developmental stages (Figure S3C-I). Mature neuron cell numbers increase and neural progenitors decrease across the span of 10 to 18 hours. Therefore, to more precisely compare expression trends, we subset the time-points from 12 to 16 hours because they contain similar numbers of cells from each of the three annotations.

Imputation of scRNA-seq was done following the protocol recommended by Cheng et al.[81] using NE algorithm (...) with adaptations to impute scRNA-seq. The raw counts were first normalized by $log(\frac{10^5 * C_{ij}}{\sum_i \quad C_{ij}} + 1)$ where C_{ij} is the count value of gene *i* in cell *j*. Then we calculated Pearson correlation between all cells to obtain cell-to-cell similarity matrix. We then apply network enhancement to improve the similarity matrix and the cell-to-cell correlations. The denoised matrix was divided by the maximum value of each cell for normalization while the self-weight was set to 1.5 times the maximum of its neighbors for each cell. The imputed counts were constructed by multiplying the original input matrix by the denoised weighting matrix. Gene-to-gene Pearson correlation matrix was then obtained from the imputed count matrix.

scATAC-seq preprocessing with pyCisTopic

scATAC-seq embryo data[15] was subset from all nervous system annotated clusters in each 2h time point. These data were merged, log normalized, and processed with Signac[77] using default RunTFIDF, FindTopFeatures(min.cutoff = 'q1'), and RunSVD. Nearest neighbors were calculated from dimensions 2:50 with lsi reduction, and clustered with resolution 0.3. Any clusters with only 2 cells were merged with their visually closest cluster in the UMAP space. The resulting 16 clusters were annotated broadly from marker and average gene expression, then merged according to annotation. These data were then subset to only neural progenitors, immature neurons, and mature neurons from 6 to 18h timepoints. Preprocessing was performed using pyCisTopic following the steps outlined in et al.[82]. Specifically, MACS2[73] was used for peak-calling with the following parameters: input_format - “BEDPE”, shift - 73, ext_size - 146, keep_dup - all, q_value - 0.05. This was done separately for each cell type using pseudobulk fragment bed files. Consensus peaks were identified for each cell type and across all cell types using an iterative approach in which overlapping peaks are filtered out based on significance. High quality cells were selected based on the following metrics: total number of unique fragments (>1x10^3.5), TSS enrichment score (> 2), and fraction of reads occurring in called peaks, FRiP (> 0.2) . Predicted doublets were removed using scrublet with expected_doublet_rate set to 0.1 and a doublet score threshold of 0.25. 31,389 high-quality cells remained after these filtering steps. For topic modeling we used the serial Latent Dirichlet Allocation (LDA) algorithm with a collapsed Gibbs sampler. This was done using the run_cgs_models() function from pyCisTopic with the following parameters: n_iter - 200, random_state - 285, alpha - 50, eta - 0.1. We ran topic modeling for each number of topics from 95 to 140 in intervals of 5, and selected 135 as the optimal number of topics using pyCisTopic’s evaluate_models() function which is based on previously published metrics[83–85]. The topic-region distribution was binarized using the binarize_topics() function from pyCisTopic with the method set to ‘otsu’, which converts topics into distinct region sets. Dropout imputation was performed using the impute_accessibility() function with the scale_factor set to 10^6. This step determines the probability of a region being accessible in a given cell by multiplying the region-topic (probability of a region belonging to a topic) and topic-cell (contribution of a topic within a cell) distributions. The imputed accessibility matrix was then used to identify Differentially Accessible Regions (DARs) through a Wilcoxon rank-sum test with the find_diff_features() function. Regions with a logFC > 1.5 and a Benjamini-Hochberg adjusted p-value < 0.05 were

selected as being differentially accessible. After topic modeling and DAR selection, 43,613 regions were selected for downstream analysis.

Motif matching using motifmatchr

Regions within 3kb of a synaptic gene were identified using BEDtools 'window -w 3000', which yielded 3,128 regions. We generated a motif database using the MotifDb[79] R package as follows: query(MotifDb, andStrings = c("Dmelanogaster"), orStrings = c("jaspar2024", "Dmelanogaster"), notStrings = "ctcfl"). This call draws Position Frequency Matrices (PFMs) from Cisbp_1.02, FlyFactorSurvey, JASPAR_2014, JASPAR_CORE, Jaspar2016, Jaspar2018, and Jaspar2022, generating a database of 1437 motifs. PFMs were converted to objects of class PWMatrix from the TFBSTools package using the convert_motifs() function from the universalmotif package. Motif matching was performed using the matchMotifs() function from motifmatchr[78] with 'genome' set to 'dm6' and 'out' set to 'positions'. This generates the coordinates of all hits for each motif in our database.

Polytene chromosome immunostaining and analysis

Polytene chromosome squashes were prepared using the protocol described in Rieder et al.[38]. For immunostaining, we used rabbit anti-CLAMP (1:500, SDIX) and guinea-pig anti-DEAF1 (1:500, gift from William McGinnis[86]) antibodies and species specific Alexa Fluor secondary antibodies at 1:500. All samples were then washed in PTX, stained with DAPI, and mounted in ProLong Diamond (Thermo Fisher Scientific). We imaged slides at 40X on a Zeiss Axio Imager M1 epifluorescence upright microscope using AxioVision v4.8.2 software.

Polytene images were analyzed using NIS Elements. Histone loci were identified by Clamp staining and a 70um line was drawn from the center of the histone locus down the chromosome using the Profile Analysis Tool[87]. Line profile fluorescence intensity was measured in each channel for the region including a neighborhood of ~1um on either side of the line. Intensities were compared by Kruskal-Wallis test followed by Dunn's multiple comparisons test for non-parametric data.

RNA isolation, library preparation, and sequencing

Larvae were sexed and their brains (optic lobes and VNC) were dissected out in PBS in rapid succession. The larval brains (N=12) were then immediately placed in RNase-free 1.5ml tubes (containing 100µl of TRIzol lysis reagent, Invitrogen, USA) on ice. There were eight biological replicates (tubes) for each genotype, with each replicate/tube containing 12 brains. The larval brains in each tube were homogenized with a pestle. QIAGEN RNeasy Plus Universal Mini Kit was used for the RNA extraction and ZYMO RNA Clean & Concentrator-25 kit was used for further purification. The RNA was eluted and confirmed for purity and concentration via 260/280 and 260/230 ratios measured using a Thermo Scientific NanoDrop One[C]. Illumina libraries were prepared and sequenced at Novogene Corporation for 150bp Poly-A selected mRNA and paired-end sequenced using an Illumina NovaSeq 6000, targeting at least 20 million reads per sample.

RNA-seq analysis

Raw sequencing data were run for quality check using FastQC/0.11.9. Adapter content was trimmed with Trim-galore/0.6.10 and Cutadapt/4.6. Hisat2/2.2.1[88] was used for paired-end read alignment to the *Drosophila* reference genome dmel-r6.42, and Samtools/1.16.1[89] flagstat and idxstats for alignment check. From Subread/2.0.2, featureCounts[90] was used to count reads at gene features with the following parameters: -p --countReadPairs -t gene -g gene_id -C -T 8. DESeq2/1.40.2[91] was used for differential gene expression analysis with ashr[92] to quantify false discovery rates (FDR). We consider genes with padj (FDR) values below 0.05 to be differentially expressed. GO analysis was performed with gProfiler2/0.2.2[67]. Jaccard index significance scores were calculated with the R package GeneOverlap/1.36.0[93].

Phenotypic analysis of synapse number

Wandering third-instar larvae were dissected in ice cold Ca^{2+}-free saline and fixed for 6 minutes in Bouin's Fixative (Electron Microscopy Sciences, Cat# 15990). Dissections were washed several times with 1X PBS and then permeabilized with 1X PBS containing 0.1% Triton-X (Fisher Scientific, Cat# BP151-100). Dissected larvae were then blocked either for 1 hour at room temperature or overnight at 4°C in 1X PBS containing 0.1% Triton-X and 1% BSA (Sigma-Aldrich, Cat# A7906). Dissections were incubated in primary antibodies overnight at 4°C and secondary antibodies for 2–4 hours at room temperature or overnight at 4°C. Dissected larvae were mounted in Vectashield (Vector Laboratories, Cat# H-1000-10). The following antibodies were used at the indicated concentrations: mouse anti-Bruchpilot (Brp) at 1:100 (DSHB, Cat# nc82, RRID: AB_2314866), anti-Horseradish Peroxidase (HRP) conjugated to Alexa Fluor 647 at 1:500 (Jackson ImmunoResearch Labs, Cat# 123-605-021, RRID: AB_2338967), and Alexa Fluor 488 goat anti-mouse at 1:500 (Invitrogen, Cat# A-11029, RRID: AB_2534088). NMJs were imaged on a Nikon A1R HD confocal microscope with a 40x oil immersion objective. Boutons were quantified at the highly stereotyped larval NMJ4 in segments A2-A4.

Quantification of boutons was performed using the "multi-point" tool in FIJI/ImageJ (NIH, RRID: SCR_003070). To quantify individual active zones, we used Nikon Elements GA3 software to preprocess images with Guassian and Rolling ball filters and identify Brp puncta using the Bright Spots module. ROIs were drawn using HRP and used as a mask to count Brp puncta. All quantifications were performed masked to genotype. Violin plots were made with R package ggplot2[94] and ggstatsplot[95]. P-values were calculated with Welch's t-test for unequal variances for single comparisons of parametric data or Kruskal-Wallis test followed by Dunn's test and Benjamini-Hochberg p-value adjustment for multiple pairwise comparisons of nonparametric data.

Data and materials availability

Data and materials generated in this study are available upon request from the corresponding authors. Transcriptomic data have been deposited in the NCBI Gene Expression Omnibus (GEO) Database under the accession number GSE280405.

Declaration of interests

The authors declare no competing interests.

Acknowledgements

We thank the Developmental Studies Hybridoma Bank and Bloomington Drosophila Stock Center (NIH P40OD018537) for antibodies and fly stocks. FlyBase[62] provided access to critical reagent and genomic information and databases. We are grateful to Scott Gratz and Karla Kaun for invaluable assistance with figures and members of the Larschan and O'Connor-Giles labs for thoughtful discussions and comments on the manuscript. This work was supported by grants to KOG and EL from the National Institute of Neurological Disorders and Stroke, National Institutes of Health (R21NS125864) and the Brown University Carney Institute for Brain Science, and by NIH training grant T32GM136566 to the Brown Molecular Biology, Cell Biology, and Biochemistry Graduate Program.

Figure Legends

Figure 1. Broad temporal coordination of synaptic gene transcription in *Drosophila* and mouse

(A) Average temporal expression profile of the 485 genes most highly correlated with the synaptic expression profile footprint (blue) is distinct from the average expression profile of 485 genes that regulate neurite development (orange) or a random set of nervous system expressed genes (grey)[9]. (B) GO analysis of the 485 genes shows enrichment for synapse-related terms. Average expression scores of the 485 coordinately expressed synaptic genes across time-series scRNA-seq datasets[15–17] displayed as feature plots (C) and averaged per time point (D) yields a similar temporal expression pattern. (E) Average profile of 742 mouse genes highly correlated with an analogous synaptic expression profile footprint shows coordinated upregulation around embryonic day 12[18]. (F) GO analysis of the 742 genes reveals enrichment of synapse-related terms. Shading around average lines represents the margin of the standard error. GO terms are color coded based on shared hierarchical parent GO terms (synaptic blue, other grey).

Figure 2. DEAF1 motifs and binding are enriched at dynamically regulated accessible regions near synaptic genes

(A) Average chromatin accessibility surrounding TSSs of synaptic genes (aqua) reveals closed chromatin at early stages and a substantial increase in accessibility at 10-12 hours in contrast to neurite development (orange) and random neuronal genes (grey). Based on analysis of tissue-specific DNase-seq time-series data[19]. (B) Coordinated opening of chromatin accessibility at synaptic genes (aqua) precedes the coordinated upregulation of synaptic transcripts (blue). (C) Average chromatin accessibility surrounding TSSs of synaptic genes (aqua) indicates that chromatin remains largely open at synaptic genes in later stages. Based on analysis of larval, pupal, and adult scATAC-seq time-series data[22,23]. (D) Motif analysis of newly accessible regions within 3kb of synaptic genes using Homer[75] identifies the DEAF1 binding motif as the most enriched. (E) Analysis of embryonic ChIP-seq data[31,32] identifies DEAF1 as the most enriched TF binding at newly accessible regions near synaptic genes (blue) compared with random genes (grey). (F) Analysis of embryonic scATAC time-series data[15] demonstrates that after 10h embryonic development, DEAF1 motifs are specifically accessible in the nervous system. Shading around average lines represents the margin of the standard error.

Figure 3. DEAF1 and CLAMP are predicted to co-repress synaptic gene expression

Third instar larval brain scRNA-seq (A) and scATAC-seq (B) data[34] clustered and annotated to isolate neuronal cells for NetREm[33] analysis. (C) TF coordination network centered on DEAF1, showing number of predicted synaptic target genes (ellipse size), predicted positive (red) to negative (dark blue) effect ratio on target genes, experimental physical interaction scores (line width), and calculated average weighted correlation score (line color intensity). (D) Embryonic scRNA-seq time-series data[15] clustered to isolate neuronal cells for expression correlation analysis (E) between DEAF1 and all genes (right) or synaptic genes (left) or CLAMP and all genes (right) or synaptic genes (left). (F) Embryonic scATAC-seq data[15] clustered to isolate neuronal cells for co-accessibility analysis of DEAF1 and CLAMP binding motifs at synaptic

genes (G). (H) Adjacent co-accessible motifs for DEAF1 (gold) and CLAMP (red) at the synaptic gene *brp*.

Figure 4. DEAF1 and CLAMP colocalize at synaptic genes
(A) DEAF1 and CLAMP binding on polytene chromosomes displays widespread binding and colocalization at a subset of sites. DEAF1 and CLAMP motif accessibility (B) and binding (C) at synaptic genes measured individually and together reveals significant enrichment of both factors.

Figure 5. *deaf1* or *clamp* knockdown de-represses synaptic gene expression
GO analysis, showing terms with the highest percentage of genes (recall) upregulated upon neuronal RNAi knockdown of *deaf1* (A) or *clamp* (B). (C) Genes upregulated upon knockdown of *deaf1* or *clamp* are significantly enriched for synaptic genes. No enrichment is observed for random neuronally-expressed genes or among downregulated genes. (D) Genes upregulated upon knockdown of *deaf1* or *clamp* significantly overlap genes with DEAF1 and/or CLAMP binding. No enrichment is observed for downregulated genes. GO terms are color coded based on shared hierarchical parent GO terms (synaptic blue, other grey). Jaccard Analysis.

Figure 6. DEAF1 and CLAMP restrain synapse formation.
(A-E) Representative images (A-C) of NMJ 4 in the indicated genotypes labeled with the neuronal membrane marker HRP (magenta) and active zone marker Brp (green) and quantification of synaptic bouton number (D) and active zone number (E). (F-H) Representative images (F, G) of NMJ 4 in the indicated genotypes labeled with the neuronal membrane marker HRP (magenta) and active zone marker Brp (green) and quantification of synaptic bouton number (H). Scale bar = 10 µm. Data points represent individual NMJs. Kruskal-Wallis test followed by Dunn's multiple comparisons test and Benjamini-Hochberg p-value adjustment. Data is represented as mean ± SEM.

Supplemental Figure Legends

Figure S1. Temporally coordinated gene expression programs
(A) Individual temporal expression profiles of the 20 well-characterized synaptic genes used to generate the synaptic gene expression footprint[9]. (B) Genes with temporal expression correlation values below a Spearman correlation cutoff of 0.80 show little enrichment for synaptic GO terms. (C) Individual temporal expression profiles of the 20 well-characterized neurite development/axon guidance genes used to generate the neurite development gene expression footprint[9]. (D) GO analysis of 485 genes with temporal expression patterns most correlated with the neurite development 'footprint' show enrichment for early neural development terms. (E) Unbiased k-means clustering analysis of all nervous system-expressed Drosophila genes by temporal expression pattern yields one cluster with a pattern closely matching the synaptic pattern (cluster 8, aqua). (F) The 485 synaptic genes mostly map to cluster 8. (G) GO analysis of cluster 8 genes reveals enrichment for the same synaptic terms as the 485 coordinately expressed synaptic genes. GO terms are color coded based on shared hierarchical parent GO terms (synaptic blue, neurite development orange, other grey).

Figure S2. Synaptic genes have temporally coordinated chromatin accessibility and are bound by DEAF1.
(A) Synaptic gene locations are dispersed broadly across the genome. (B) Average expression profiles of the 19 synaptic genes with temporal protein expression data[20] show chromatin accessibility (aqua) and transcription (blue) immediately preceding protein expression (purple). (C) Dynamic chromatin accessibility at synaptic genes TSSs (aqua) is not observed in mesodermal tissue.

Figure S3. DEAF1 is predicted to regulate most synaptic genes
(A) Independent motif analysis using MEME Suite[76], which does not include the DEAF1 motif in its database, identified overlapping and additional enriched binding motifs. (B) Network of top 10 TFs predicted by NetREm, showing number of predicted synaptic target genes (ellipse size), predicted positive (red) to negative (dark blue) effect ratio on target genes, experimental physical interaction scores (line width), and calculated average weighted correlation score (line color intensity). UMAP generated from embryonic scRNA-seq time-series data[15] prior to imputation (C) and following imputation (D). Expression of selected marker genes prior to (E) and following (F) imputation. Co-accessible DEAF1 (gold) and CLAMP (red) motifs at synaptic genes *futsch* (G), and *mAChR-A* (H).

Figure S4. *deaf1* and *clamp* RNAi significantly decrease DEAF1 and CLAMP protein on chromatin
Polytene chromosome spreads in control *gfp*-RNAi (A), *deaf-1*RNAi (B), and *clamp*-RNAi (C) driven by *elav*-Gal4. Close up images (right) show protein in a region that was traced for fluorescence intensity measurement (bottom). Quantification of total DAPI (D), DEAF1 (E), or CLAMP (F) fluorescence along the trace for each genotype normalized to *gfp*RNAi controls. Data points represent animals. Kruskal-Wallis test followed by Dunn's multiple comparisons test. Data is represented as mean ± SEM.

Figure S5. DEAF1 and CLAMP directly regulate repression of synaptic genes
GO analysis, showing terms with the highest percentage of genes (recall) downregulated upon neuronal RNAi knockdown of *deaf1* (A) or *clamp* (B). GO terms are color coded based on shared hierarchical parent GO terms (synaptic blue, other grey).

Figure S6. DEAF1 and CLAMP mutations cause dysregulation of synapse formation
Representative images (A-D) of NMJ 4 in the indicated genotypes labeled with the neuronal membrane marker HRP (magenta) and active zone marker Brp (green) and quantification of synaptic bouton number (F). Scale bar = 10 μm. Data points represent individual NMJs. Welch's t-test for unequal variances. Data is represented as mean ± SEM.

References

1. Hobert, O., and Kratsios, P. (2019). Neuronal identity control by terminal selectors in worms, flies, and chordates. Curr Opin Neurobiol *56*, 97–105.
2. El-Danaf, R.N., Rajesh, R., and Desplan, C. (2023). Temporal regulation of neural diversity in Drosophila and vertebrates. Semin Cell Dev Biol *142*, 13–22.
3. Hamid, A., Gutierrez, A., Munroe, J., and Syed, M.H. (2023). The Drivers of Diversity: Integrated genetic and hormonal cues regulate neural diversity. Semin Cell Dev Biol *142*, 23–35.
4. Xing, Y., Zan, C., and Liu, L. (2023). Recent advances in understanding neuronal diversity and neural circuit complexity across different brain regions using single-cell sequencing. Front Neural Circuits *17*, 1007755.
5. Batool, S., Raza, H., Zaidi, J., Riaz, S., Hasan, S., and Syed, N.I. (2019). Synapse formation: from cellular and molecular mechanisms to neurodevelopmental and neurodegenerative disorders. J. Neurophysiol. *121*, 1381–1397. https://doi.org/10.1152/jn.00833.2018.
6. Wilhelm, B.G., Mandad, S., Truckenbrodt, S., Krohnert, K., Schafer, C., Rammner, B., Koo, S.J., Classen, G.A., Krauss, M., Haucke, V., et al. (2014). Composition of isolated synaptic boutons reveals the amounts of vesicle trafficking proteins. Science *344*, 1023–1028.
7. Gallegos, D.A., Chan, U., Chen, L.F., and West, A.E. (2018). Chromatin Regulation of Neuronal Maturation and Plasticity. Trends Neurosci *41*, 311–324.
8. Kishi, Y., and Gotoh, Y. (2018). Regulation of Chromatin Structure During Neural Development. Front Neurosci *12*, 874.
9. 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.
10. Honjo, K., Mauthner, S.E., Wang, Y., Skene, J.H.P., and Tracey, W.D. (2016). Nociceptor-Enriched Genes Required for Normal Thermal Nociception. Cell Rep. *16*, 295–303. https://doi.org/10.1016/j.celrep.2016.06.003.
11. Hogan, C.A., Gratz, S.J., Dumouchel, J.L., Thakur, R.S., Delgado, A., Lentini, J.M., Madhwani, K.R., Fu, D., and O'Connor-Giles, K.M. (2023). Expanded tRNA methyltransferase family member TRMT9B regulates synaptic growth and function. EMBO Rep *24*, e56808.
12. Thakur, R.S., and O'Connor-Giles, K.M. (2023). PDZD8 promotes autophagy at ER-Lysosome contact sites to regulate synaptogenesis. bioRxiv.
13. Frank, C.L., Liu, F., Wijayatunge, R., Song, L., Biegler, M.T., Yang, M.G., Vockley, C.M., Safi, A., Gersbach, C.A., Crawford, G.E., et al. (2015). Regulation of chromatin accessibility and Zic binding at enhancers in the developing cerebellum. Nat. Neurosci. *18*, 647–656. https://doi.org/10.1038/nn.3995.
14. Bakken, T.E., Miller, J.A., Ding, S.-L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Dalley, R.A., Royall, J.J., Lemon, T., et al. (2016). A comprehensive transcriptional map of primate brain development. Nature *535*, 367–375. https://doi.org/10.1038/nature18637.
15. Calderon, D., Blecher-Gonen, R., Huang, X., Secchia, S., Kentro, J., Daza, R.M., Martin, B., Dulja, A., Schaub, C., Trapnell, C., et al. (2022). The continuum of *Drosophila* embryonic development at single-cell resolution. Science *377*, eabn5800.
16. Corrales, M., Cocanougher, B.T., Kohn, A.B., Wittenbach, J.D., Long, X.S., Lemire, A., Cardona, A., Singer, R.H., Moroz, L.L., and Zlatic, M. (2022). A single-cell transcriptomic atlas of complete insect nervous systems across multiple life stages. Neural Dev *17*, 8.
17. Kurmangaliyev, Y.Z., Yoo, J., Valdes-Aleman, J., Sanfilippo, P., and Zipursky, S.L. (2020). Transcriptional Programs of Circuit Assembly in the Drosophila Visual System. Neuron *108*, 1045-1057 e6.
18. Cardoso-Moreira, M., Halbert, J., Valloton, D., Velten, B., Chen, C., Shao, Y., Liechti, A.,

Ascencao, K., Rummel, C., Ovchinnikova, S., et al. (2019). Gene expression across mammalian organ development. Nature *571*, 505–509.
19. Reddington, J.P., Garfield, D.A., Sigalova, O.M., Karabacak Calviello, A., Marco-Ferreres, R., Girardot, C., Viales, R.R., Degner, J.F., Ohler, U., and Furlong, E.E.M. (2020). Lineage-Resolved Enhancer and Promoter Usage during a Time Course of Embryogenesis. Dev Cell *55*, 648-664.e9.
20. Casas-Vila, N., Bluhm, A., Sayols, S., Dinges, N., Dejung, M., Altenhein, T., Kappei, D., Altenhein, B., Roignant, J.-Y., and Butter, F. (2017). The developmental proteome of Drosophila melanogaster. Genome Res *27*, 1273–1285.
21. Secchia, S., Forneris, M., Heinen, T., Stegle, O., and Furlong, E.E.M. (2022). Simultaneous cellular and molecular phenotyping of embryonic mutants using single-cell regulatory trajectories. Dev. Cell *57*, 496-511.e8. https://doi.org/10.1016/j.devcel.2022.01.016.
22. Fogarty, E.A., Buchert, E.M., Ma, Y., Nicely, A.B., and Buttitta, L.A. (2024). Transcriptional repression and enhancer decommissioning silence cell cycle genes in postmitotic tissues. G3 Bethesda Md *14*, jkae203. https://doi.org/10.1093/g3journal/jkae203.
23. Merrill, C.B., Titos, I., Pabon, M.A., Montgomery, A.B., Rodan, A.R., and Rothenfluh, A. (2024). Iterative assay for transposase-accessible chromatin by sequencing to isolate functionally relevant neuronal subtypes. Sci. Adv. *10*, eadi4393. https://doi.org/10.1126/sciadv.adi4393.
24. Huggenvik, J.I., Michelson, R.J., Collard, M.W., Ziemba, A.J., Gurley, P., and Mowen, K.A. (1998). Characterization of a nuclear deformed epidermal autoregulatory factor-1 (DEAF-1)-related (NUDR) transcriptional regulator protein. Mol Endocrinol *12*, 1619–1639.
25. Michelson, R.J., Collard, M.W., Ziemba, A.J., Persinger, J., Bartholomew, B., and Huggenvik, J.I. (1999). Nuclear DEAF-1-related (NUDR) Protein Contains a Novel DNA Binding Domain and Represses Transcription of the Heterogeneous Nuclear Ribonucleoprotein A2/B1 Promoter*. J Biol Chem *274*, 30510–30519.
26. 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.
27. Nabais Sa, M.J., Jensik, P.J., McGee, S.R., Parker, M.J., Lahiri, N., McNeil, E.P., Kroes, H.Y., Hagerman, R.J., Harrison, R.E., Montgomery, T., et al. (2019). De novo and biallelic DEAF1 variants cause a phenotypic spectrum. Genet Med *21*, 2059–2069.
28. McGee, S.R., Rajamanickam, S., Adhikari, S., Falayi, O.C., Wilson, T.A., Shayota, B.J., Cooley Coleman, J.A., Skinner, C., Caylor, R.C., Stevenson, R.E., et al. (2023). Expansion and mechanistic insights into de novo DEAF1 variants in DEAF1-associated neurodevelopmental disorders. Hum Mol Genet *32*, 386–401.
29. Bottomley, M.J., Collard, M.W., Huggenvik, J.I., Liu, Z., Gibson, T.J., and Sattler, M. (2001). The SAND domain structure defines a novel DNA-binding fold in transcriptional regulation. Nat Struct Biol *8*, 626–633.
30. Kateb, F., Perrin, H., Tripsianes, K., Zou, P., Spadaccini, R., Bottomley, M., Franzmann, T.M., Buchner, J., Ansieau, S., and Sattler, M. (2013). Structural and functional analysis of the DEAF-1 and BS69 MYND domains. PLoS One *8*, e54715.
31. The modENCODE Consortium, Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., et al. (2010). Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. Science *330*, 1787–1797. https://doi.org/10.1126/science.1198374.
32. Fisher, W.W., Hammonds, A.S., Weiszmann, R., Booth, B.W., Gevirtzman, L., Patton, J.E.J., Kubo, C.A., Waterston, R.H., and Celniker, S.E. (2023). A modERN resource: identification of Drosophila transcription factor candidate target genes using RNAi. Genetics *223*.
33. Khullar, S., Huang, X., Ramesh, R., Svaren, J., and Wang, D. (2024). NetREm Network

Regression Embeddings reveal cell-type transcription factor coordination for gene regulation. bioRxiv, 2023.10.25.563769.
34. Mohana, G., Dorier, J., Li, X., Mouginot, M., Smith, R.C., Malek, H., Leleu, M., Rodriguez, D., Khadka, J., Rosa, P., et al. (2023). Chromosome-level organization of the regulatory genome in the Drosophila nervous system. Cell *186*, 3826-3844 e26.
35. Duan, J., Rieder, L., Colonnetta, M.M., Huang, A., McKenney, M., Watters, S., Deshpande, G., Jordan, W., Fawzi, N., and Larschan, E. (2021). CLAMP and Zelda function together to promote Drosophila zygotic genome activation. Elife *10*.
36. Soruco, M.M., Chery, J., Bishop, E.P., Siggers, T., Tolstorukov, M.Y., Leydon, A.R., Sugden, A.U., Goebel, K., Feng, J., Xia, P., et al. (2013). The CLAMP protein links the MSL complex to the X chromosome during Drosophila dosage compensation. Genes Dev *27*, 1551–1556.
37. Ray, M., Conard, A.M., Urban, J., Mahableshwarkar, P., Aguilera, J., Huang, A., Vaidyanathan, S., and Larschan, E. (2023). Sex-specific splicing occurs genome-wide during early Drosophila embryogenesis. eLife *12*, e87865. https://doi.org/10.7554/eLife.87865.
38. Rieder, L.E., Koreski, K.P., Boltz, K.A., Kuzu, G., Urban, J.A., Bowman, S.K., Zeidman, A., Jordan, W.T., 3rd, Tolstorukov, M.Y., Marzluff, W.F., et al. (2017). Histone locus regulation by the Drosophila dosage compensation adaptor protein CLAMP. Genes Dev *31*, 1494–1508.
39. Urban, J.A., Doherty, C.A., Jordan, W.T., 3rd, Bliss, J.E., Feng, J., Soruco, M.M., Rieder, L.E., Tsiarli, M.A., and Larschan, E.N. (2017). The essential Drosophila CLAMP protein differentially regulates non-coding roX RNAs in male and females. Chromosome Res *25*, 101–113.
40. Kaye, E.G., Booker, M., Kurland, J.V., Conicella, A.E., Fawzi, N.L., Bulyk, M.L., Tolstorukov, M.Y., and Larschan, E. (2018). Differential Occupancy of Two GA-Binding Proteins Promotes Targeting of the Drosophila Dosage Compensation Complex to the Male X Chromosome. Cell Rep. *22*, 3227–3239. https://doi.org/10.1016/j.celrep.2018.02.098.
41. Bag, I., Dale, R.K., Palmer, C., and Lei, E.P. (2019). The zinc-finger protein CLAMP promotes gypsy chromatin insulator function in Drosophila. J Cell Sci *132*.
42. Kuzu, G., Kaye, E.G., Chery, J., Siggers, T., Yang, L., Dobson, J.R., Boor, S., Bliss, J., Liu, W., Jogl, G., et al. (2016). Expansion of GA Dinucleotide Repeats Increases the Density of CLAMP Binding Sites on the X-Chromosome to Promote Drosophila Dosage Compensation. PLoS Genet *12*, e1006120.
43. Czesak, M., Lemonde, S., Peterson, E.A., Rogaeva, A., and Albert, P.R. (2006). Cell-specific repressor or enhancer activities of Deaf-1 at a serotonin 1A receptor gene polymorphism. J Neurosci *26*, 1864–1871.
44. Harkin, E.F., Nasrallah, G., Le Francois, B., and Albert, P.R. (2023). Transcriptional Regulation of the Human 5-HT1A Receptor Gene by Lithium: Role of Deaf1 and GSK3beta. Int J Mol Sci *24*.
45. Tsiarli, M.A., Conard, A.M., Xu, L., Nguyen, E., and Larschan, E.N. (2020). The transcription factor CLAMP is required for neurogenesis in Drosophila melanogaster. Preprint at bioRxiv, https://doi.org/10.1101/2020.10.09.333831 https://doi.org/10.1101/2020.10.09.333831.
46. Haussmann, I.U., White, K., and Soller, M. (2008). Erect wing regulates synaptic growth in Drosophila by integration of multiple signaling pathways. Genome Biol. *9*, R73. https://doi.org/10.1186/gb-2008-9-4-r73.
47. Philippe, T.J., Vahid-Ansari, F., Donaldson, Z.R., Le François, B., Zahrai, A., Turcotte-Cardin, V., Daigle, M., James, J., Hen, R., Merali, Z., et al. (2018). Loss of MeCP2 in adult 5-HT neurons induces 5-HT1A autoreceptors, with opposite sex-dependent anxiety and depression phenotypes. Sci. Rep. *8*, 5788. https://doi.org/10.1038/s41598-018-24167-8.
48. Ruiz, J.L., Yerbanga, R.S., Lefèvre, T., Ouedraogo, J.B., Corces, V.G., and Gómez-Díaz, E.

(2019). Chromatin changes in Anopheles gambiae induced by Plasmodium falciparum infection. Epigenetics Chromatin *12*, 5. https://doi.org/10.1186/s13072-018-0250-9.
49. Kratsios, P., Pinan-Lucarre, B., Kerk, S.Y., Weinreb, A., Bessereau, J.L., and Hobert, O. (2015). Transcriptional coordination of synaptogenesis and neurotransmitter signaling. Curr Biol *25*, 1282–1295.
50. Konstantinides, N., Kapuralin, K., Fadil, C., Barboza, L., Satija, R., and Desplan, C. (2018). Phenotypic Convergence: Distinct Transcription Factors Regulate Common Terminal Features. Cell *174*, 622-635 e13.
51. Haussmann, I.U., and Soller, M. (2010). Differential activity of EWG transcription factor isoforms identifies a subset of differentially regulated genes important for synaptic growth regulation. Dev Biol *348*, 224–230.
52. Stefanakis, N., Carrera, I., and Hobert, O. (2015). Regulatory Logic of Pan-Neuronal Gene Expression in C. elegans. Neuron *87*, 733–750.
53. Noack, F., Vangelisti, S., Raffl, G., Carido, M., Diwakar, J., Chong, F., and Bonev, B. (2022). Multimodal profiling of the transcriptional regulatory landscape of the developing mouse cortex identifies Neurog2 as a key epigenome remodeler. Nat. Neurosci. *25*, 154–167. https://doi.org/10.1038/s41593-021-01002-4.
54. Adkins, N.L., Hagerman, T.A., and Georgel, P. (2006). GAGA protein: a multi-faceted transcription factorThis paper is one of a selection of papers published in this Special Issue, entitled 27th International West Coast Chromatin and Chromosome Conference, and has undergone the Journal's usual peer review process. Biochem. Cell Biol. *84*, 559–558. https://doi.org/10.1139/o06-062.
55. Bayarmagnai, B., Nicolay, B.N., Islam, A.B.M.M.K., Lopez-Bigas, N., and Frolov, M.V. (2012). Drosophila GAGA factor is required for full activation of the dE2f1-Yki/Sd transcriptional program. Cell Cycle *11*, 4191–4202. https://doi.org/10.4161/cc.22486.
56. Li, X., Tang, X., Bing, X., Catalano, C., Li, T., Dolsten, G., Wu, C., and Levine, M. (2023). GAGA-associated factor fosters loop formation in the *Drosophila* genome. Mol. Cell *83*, 1519-1526.e4. https://doi.org/10.1016/j.molcel.2023.03.011.
57. Leyva-Diaz, E., and Hobert, O. (2022). Robust regulatory architecture of pan-neuronal gene expression. Curr Biol.
58. Leyva-Díaz, E. (2023). CUT homeobox genes: transcriptional regulation of neuronal specification and beyond. Front. Cell. Neurosci. *17*. https://doi.org/10.3389/fncel.2023.1233830.
59. Vulto-van Silfhout, A.T., Rajamanickam, S., Jensik, P.J., Vergult, S., de Rocker, N., Newhall, K.J., Raghavan, R., Reardon, S.N., Jarrett, K., McIntyre, T., et al. (2014). Mutations affecting the SAND domain of DEAF1 cause intellectual disability with severe speech impairment and behavioral problems. Am J Hum Genet *94*, 649–661.
60. Chen, L., Jensik, P.J., Alaimo, J.T., Walkiewicz, M., Berger, S., Roeder, E., Faqeih, E.A., Bernstein, J.A., Smith, A.C.M., Mullegama, S.V., et al. (2017). Functional analysis of novel DEAF1 variants identified through clinical exome sequencing expands DEAF1-associated neurodevelopmental disorder (DAND) phenotype. Hum Mutat *38*, 1774–1785.
61. Bischof, J., Björklund, M., Furger, E., Schertel, C., Taipale, J., and Basler, K. (2013). A versatile platform for creating a comprehensive UAS-ORFeome library in Drosophila. Development *140*, 2434–2442. https://doi.org/10.1242/dev.088757.
62. Öztürk-Çolak, A., Marygold, S.J., Antonazzo, G., Attrill, H., Goutte-Gattat, D., Jenkins, V.K., Matthews, B.B., Millburn, G., Dos Santos, G., Tabone, C.J., et al. (2024). FlyBase: updates to the Drosophila genes and genomes database. Genetics *227*.
63. R Core Team (2023). R: A Language and Environment for Statistical Computing.
64. Wickham, H., François, R., Henry, L., Müller, K., and Vaughan, D. (2023). dplyr: A Grammar of Data Manipulation.
65. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2023). cluster: Cluster

Analysis Basics and Extensions.
66. Kassambara, A., and Mundt, F. (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses.
67. Kolberg, L., Raudvere, U., Kuzmin, I., Adler, P., Vilo, J., and Peterson, H. (2023). g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). Nucleic Acids Res *51*, W207–W212.
68. Sayols, S. (2023). rrvgo: a Bioconductor package to reduce and visualize Gene Ontology terms. MicroPublication Biol. https://doi.org/10.17912/micropub.biology.000811.
69. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573-3587 e29.
70. Marsh, S., Salmon, M., and Hoffman, P. (2024). samuel-marsh/scCustomize: Version 2.1.2. Version v2.1.2 (Zenodo). https://doi.org/10.5281/zenodo.10724532 https://doi.org/10.5281/zenodo.10724532.
71. Hu, Y., Comjean, A., Rodiger, J., Liu, Y., Gao, Y., Chung, V., Zirin, J., Perrimon, N., and Mohr, S.E. (2021). FlyRNAi.org-the database of the Drosophila RNAi screening center and transgenic RNAi project: 2021 update. Nucleic Acids Res *49*, D908–D915.
72. 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. https://doi.org/10.1093/bioinformatics/btq351.
73. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.
74. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.
75. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576–589.
76. Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME Suite. Nucleic Acids Res *43*, W39-49.
77. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A., and Satija, R. (2021). Single-cell chromatin state analysis with Signac. Nat. Methods *18*, 1333–1341. https://doi.org/10.1038/s41592-021-01282-5.
78. Schep, A. (2024). motifmatchr: Fast Motif Matching in R.
79. Shannon P, R.M. (2024). MotifDb: An Annotated Collection of Protein-DNA Binding Sequence Motifs.
80. Wang, B., Pourshafeie, A., Zitnik, M., Zhu, J., Bustamante, C.D., Batzoglou, S., and Leskovec, J. (2018). Network enhancement as a general method to denoise weighted biological networks. Nat Commun *9*, 3108.
81. Cheng, Y., Ma, X., Yuan, L., Sun, Z., and Wang, P. (2023). Evaluating imputation methods for single-cell RNA-seq data. BMC Bioinformatics *24*, 302.
82. Bravo Gonzalez-Blas, C., De Winter, S., Hulselmans, G., Hecker, N., Matetovici, I., Christiaens, V., Poovathingal, S., Wouters, J., Aibar, S., and Aerts, S. (2023). SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. Nat Methods *20*, 1355–1367.
83. Mimno, D., Wallach, H., Talley, E., Leenders, M., and McCallum, A. (2011). Optimizing semantic coherence in topic models. In Proceedings of the 2011 conference on empirical methods in natural language processing, pp. 262–272.
84. Arun, R., Suresh, V., Veni Madhavan, C.E., and Narasimha Murthy, M.N. (2010). On

Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In Advances in Knowledge Discovery and Data Mining, M. J. Zaki, J. X. Yu, B. Ravindran, and V. Pudi, eds. (Springer), pp. 391–402. https://doi.org/10.1007/978-3-642-13657-3_43.
85. Cao, J., Xia, T., Li, J., Zhang, Y., and Tang, S. (2009). A density-based method for adaptive LDA model selection. Neurocomputing *72*, 1775–1781. https://doi.org/10.1016/j.neucom.2008.06.011.
86. Veraksa, A., Kennison, J., and McGinnis, W. (2002). DEAF-1 function is essential for the early embryonic development of Drosophila. Genesis *33*, 67–76.
87. Amankwaa, B., Schoborg, T., and Labrador, M. (2022). Drosophila insulator proteins exhibit in vivo liquid–liquid phase separation properties. Life Sci. Alliance *5*. https://doi.org/10.26508/lsa.202201536.
88. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol *37*, 907–915.
89. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. Gigascience *10*.
90. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.
91. 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.
92. Stephens, M. (2017). False discovery rates: a new deal. Biostatistics *18*, 275–294.
93. Shen, L., and Sinai, I.S. of M. at M. (2023). GeneOverlap: Test and visualize gene overlaps https://doi.org/10.18129/B9.bioc.GeneOverlap.
94. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York).
95. Patil, I. (2021). Visualizations with statistical details: The “ggstatsplot” approach. J. Open Source Softw. *6*, 3167. https://doi.org/10.21105/joss.03167.

Figure 1

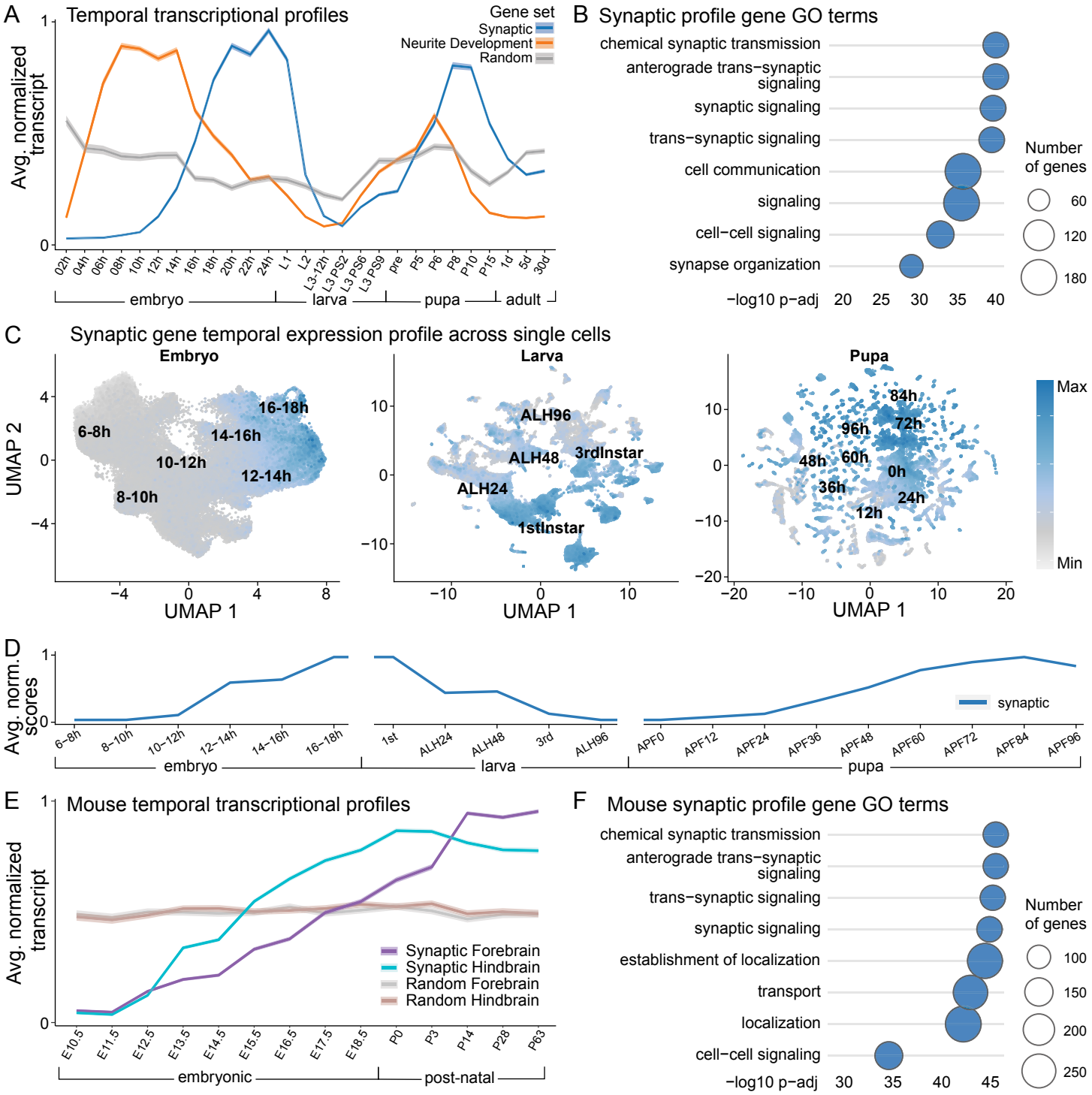
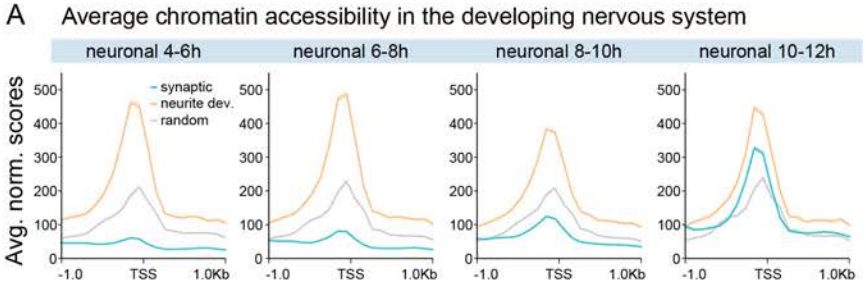
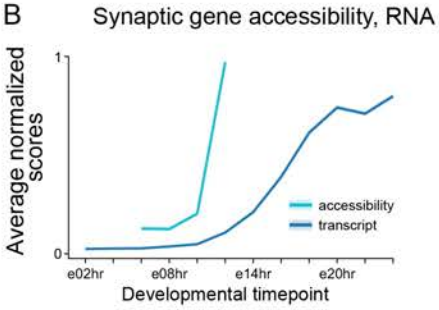
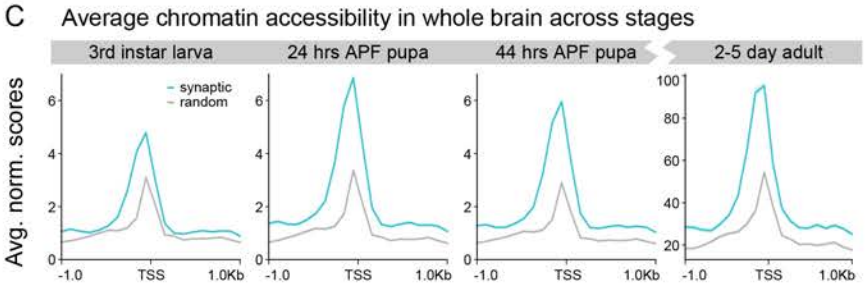
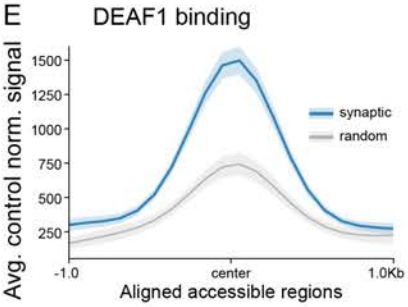

A
Temporal transcriptional profiles
Gene set
Synaptic
Neurite Development
Random
Avg. normalized transcript
embryo
larva
pupa
adult
B
Synaptic profile gene GO terms
chemical synaptic transmission
anterograde trans−synaptic signaling
synaptic signaling
trans−synaptic signaling
cell communication
signaling
cell−cell signaling
synapse organization
−log10 p−adj
Number of genes
C
Synaptic gene temporal expression profile across single cells
Embryo
Larva
Pupa
UMAP 1
UMAP 2
Max
Min
D
Avg. norm. scores
synaptic
E
Mouse temporal transcriptional profiles
Synaptic Forebrain
Synaptic Hindbrain
Random Forebrain
Random Hindbrain
embryonic
post-natal
F
Mouse synaptic profile gene GO terms
chemical synaptic transmission
anterograde trans−synaptic signaling
trans−synaptic signaling
synaptic signaling
establishment of localization
transport
localization
cell−cell signaling
−log10 p−adj
Number of genes

Figure 2









D Enriched motifs in newly accessible regions

Motif	∑-log(*p*)	Best match
	553	Deaf1
	343	GAF/Clamp
	296	pan
	239	pho
	222	byn
	200	Med
	178	run

Motif	∑-log(*p*)	Best match
	166	Mef2
	161	Blimp-1
	154	pnt
	144	hth
	127	bin
	121	kni
	103	Hsf

Motif	∑-log(*p*)	Best match
	94	vvl
	84	CG11617
	74	Top2
	64	odd
	52	hb
	28	B-H1
	18	Adf1

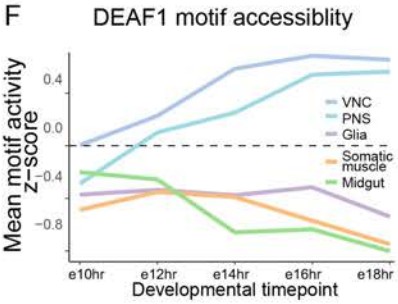


Figure 3

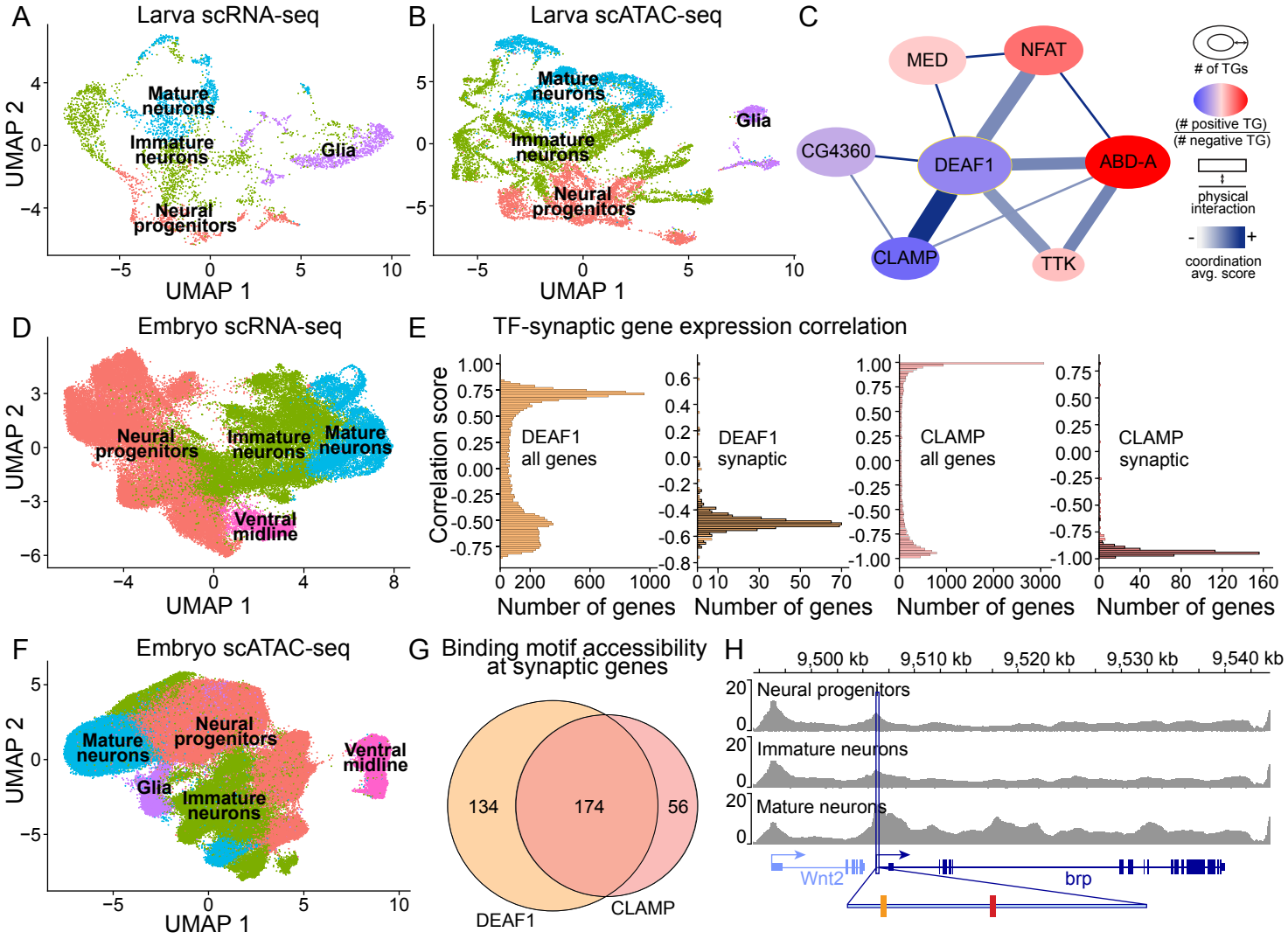

A
Larva scRNA-seq
Mature neurons
Immature neurons
Glia
Neural progenitors
UMAP 2
UMAP 1
B
Larva scATAC-seq
Mature neurons
Immature neurons
Glia
Neural progenitors
UMAP 1
C
MED
NFAT
CG4360
DEAF1
ABD-A
CLAMP
TTK
# of TGs
(# positive TG)
(# negative TG)
physical interaction
coordination avg. score
D
Embryo scRNA-seq
Neural progenitors
Immature neurons
Mature neurons
Ventral midline
UMAP 2
UMAP 1
E
TF-synaptic gene expression correlation
Correlation score
DEAF1 all genes
DEAF1 synaptic
CLAMP all genes
CLAMP synaptic
Number of genes
F
Embryo scATAC-seq
Mature neurons
Neural progenitors
Ventral midline
Glia
Immature neurons
UMAP 2
UMAP 1
G
Binding motif accessibility at synaptic genes
134
174
56
DEAF1
CLAMP
H
9,500 kb
9,510 kb
9,520 kb
9,530 kb
9,540 kb
Neural progenitors
Immature neurons
Mature neurons
Wnt2
brp

Figure 4

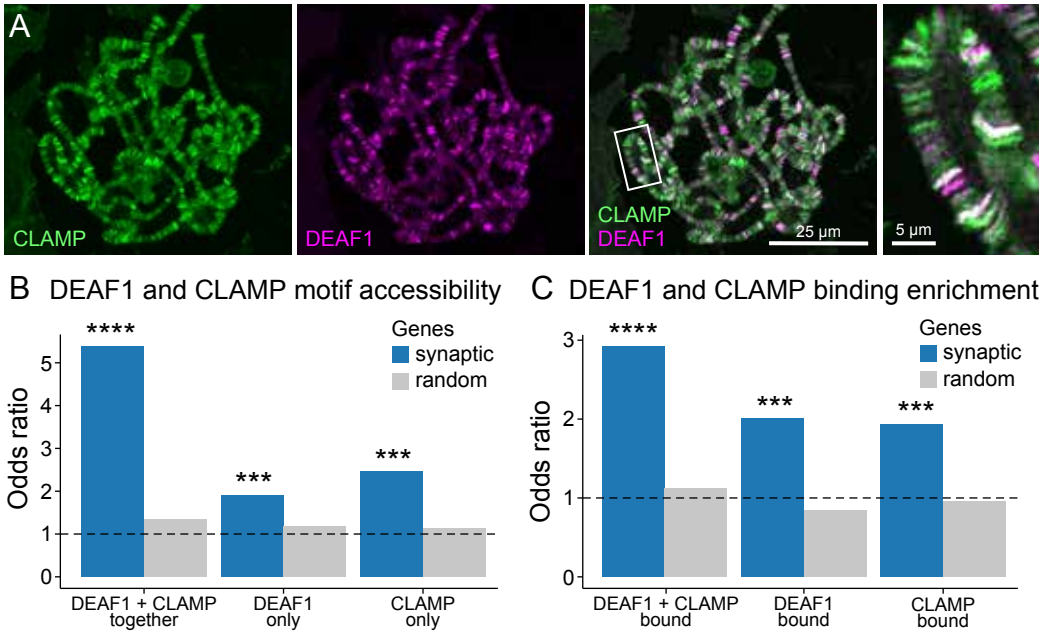

A
CLAMP
DEAF1
CLAMP
DEAF1
25 µm
5 µm
B DEAF1 and CLAMP motif accessibility
Genes
synaptic
random
Odds ratio
****
***
***
DEAF1 + CLAMP together
DEAF1 only
CLAMP only
C DEAF1 and CLAMP binding enrichment
Genes
synaptic
random
Odds ratio
****
***
***
DEAF1 + CLAMP bound
DEAF1 bound
CLAMP bound

Figure 5

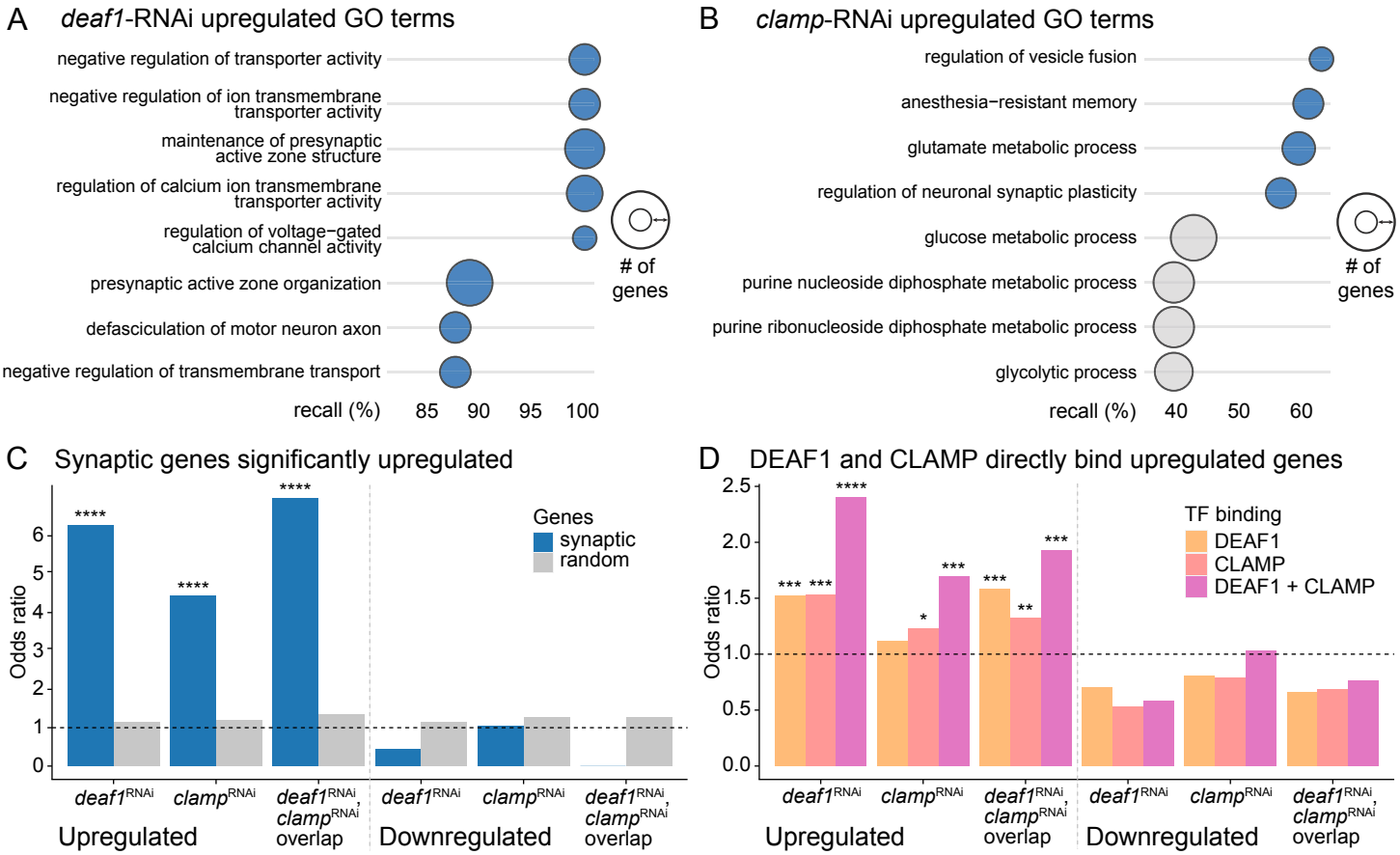

A
deaf1-RNAi upregulated GO terms
negative regulation of transporter activity
negative regulation of ion transmembrane transporter activity
maintenance of presynaptic active zone structure
regulation of calcium ion transmembrane transporter activity
regulation of voltage−gated calcium channel activity
presynaptic active zone organization
defasciculation of motor neuron axon
negative regulation of transmembrane transport
recall (%)
85
90
95
100
# of genes
B
clamp-RNAi upregulated GO terms
regulation of vesicle fusion
anesthesia−resistant memory
glutamate metabolic process
regulation of neuronal synaptic plasticity
glucose metabolic process
purine nucleoside diphosphate metabolic process
purine ribonucleoside diphosphate metabolic process
glycolytic process
recall (%)
40
50
60
# of genes
C
Synaptic genes significantly upregulated
Odds ratio
Genes
synaptic
random
deaf1RNAi
clampRNAi
deaf1RNAi, clampRNAi
Upregulated
overlap
Downregulated
D
DEAF1 and CLAMP directly bind upregulated genes
TF binding
DEAF1
CLAMP
DEAF1 + CLAMP

Figure 6

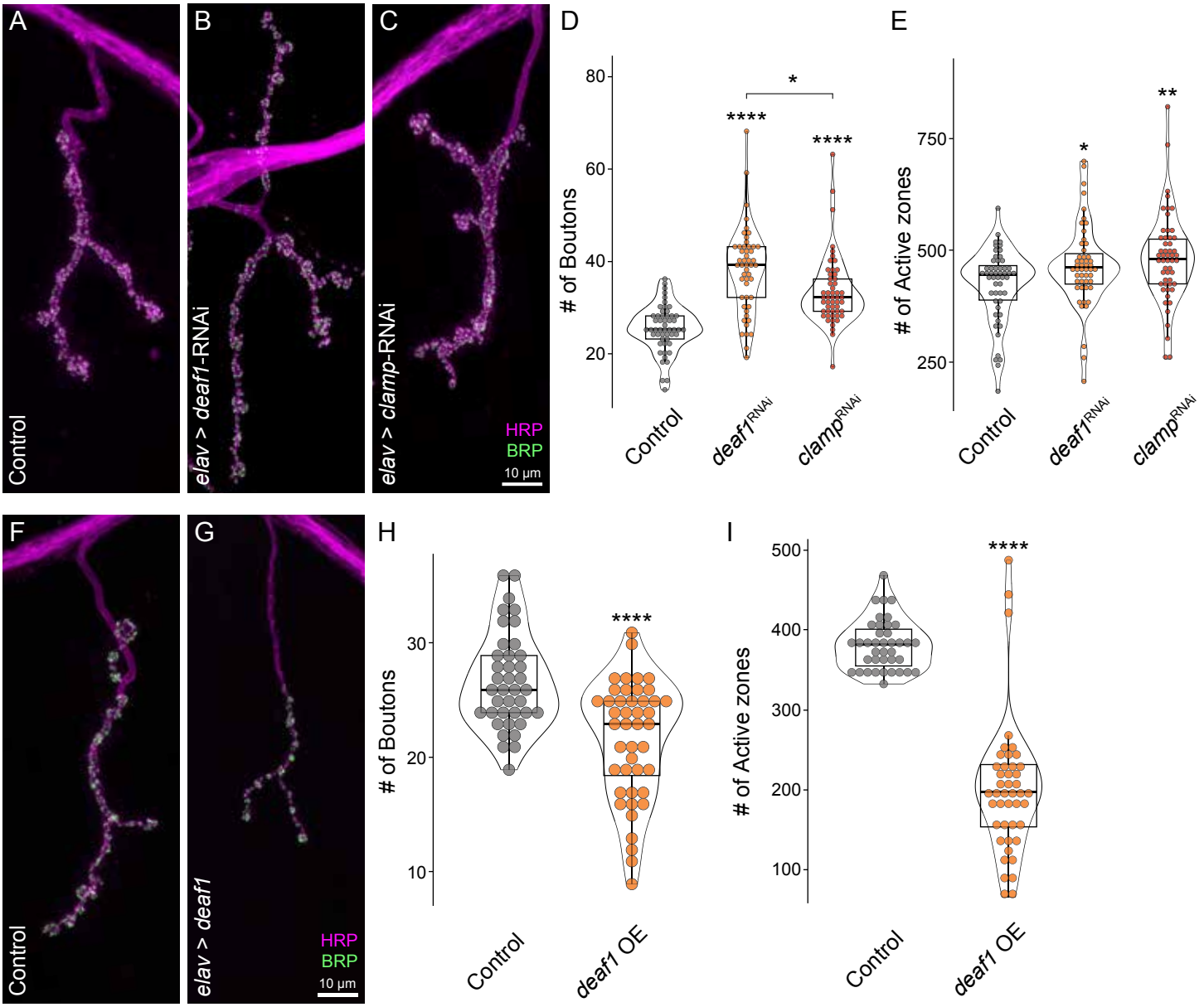

A
Control
B
elav > deaf1-RNAi
C
elav > clamp-RNAi
HRP
BRP
10 µm
D
# of Boutons
80
60
40
20
*
****
****
Control
deaf1RNAi
clampRNAi
E
# of Active zones
750
500
250
*
**
Control
deaf1RNAi
clampRNAi
F
Control
G
elav > deaf1
HRP
BRP
10 µm
H
# of Boutons
30
20
10
****
Control
deaf1 OE
I
# of Active zones
500
400
300
200
100
****
Control
deaf1 OE

Figure S1

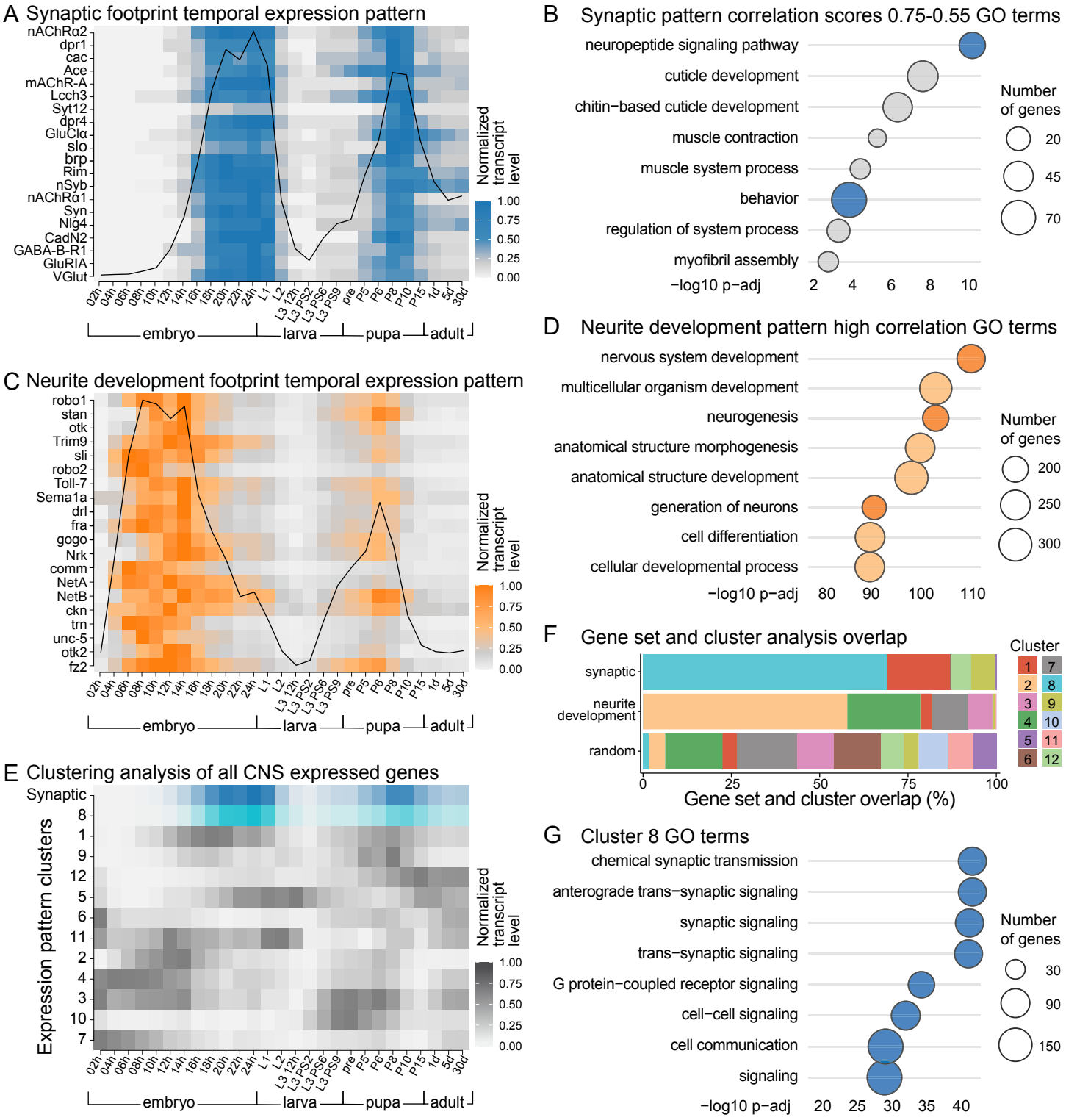

A Synaptic footprint temporal expression pattern
nAChRα2
dpr1
cac
Ace
mAChR-A
Lcch3
Syt12
dpr4
GluClα
slo
brp
Rim
nSyb
nAChRα1
Syn
Nlg4
CadN2
GABA-B-R1
GluRIA
VGlut
Normalized transcript level
1.00
0.75
0.50
0.25
0.00
02h 04h 06h 08h 10h 12h 14h 16h 18h 20h 22h 24h L1 L2 L3 12h L3 PS2 L3 PS6 L3 PS9 pre P5 P6 P8 P10 P15 1d 5d 30d
embryo
larva
pupa
adult
B Synaptic pattern correlation scores 0.75-0.55 GO terms
neuropeptide signaling pathway
cuticle development
chitin−based cuticle development
muscle contraction
muscle system process
behavior
regulation of system process
myofibril assembly
−log10 p−adj
2 4 6 8 10
Number of genes
20
45
70
C Neurite development footprint temporal expression pattern
robo1
stan
otk
Trim9
sli
robo2
Toll-7
Sema1a
drl
fra
gogo
Nrk
comm
NetA
NetB
ckn
trn
unc-5
otk2
fz2
Normalized transcript level
1.00
0.75
0.50
0.25
0.00
02h 04h 06h 08h 10h 12h 14h 16h 18h 20h 22h 24h L1 L2 L3 12h L3 PS2 L3 PS6 L3 PS9 pre P5 P6 P8 P10 P15 1d 5d 30d
embryo
larva
pupa
adult
D Neurite development pattern high correlation GO terms
nervous system development
multicellular organism development
neurogenesis
anatomical structure morphogenesis
anatomical structure development
generation of neurons
cell differentiation
cellular developmental process
−log10 p−adj
80 90 100 110
Number of genes
200
250
300
E Clustering analysis of all CNS expressed genes
Synaptic
Expression pattern clusters
8
1
9
12
5
6
11
2
4
3
10
7
Normalized transcript level
1.00
0.75
0.50
0.25
0.00
02h 04h 06h 08h 10h 12h 14h 16h 18h 20h 22h 24h L1 L2 L3 12h L3 PS2 L3 PS6 L3 PS9 pre P5 P6 P8 P10 P15 1d 5d 30d
embryo
larva
pupa
adult
F Gene set and cluster analysis overlap
synaptic
neurite development
random
0 25 50 75 100
Gene set and cluster overlap (%)
Cluster
1 7
2 8
3 9
4 10
5 11
6 12
G Cluster 8 GO terms
chemical synaptic transmission
anterograde trans−synaptic signaling
synaptic signaling
trans−synaptic signaling
G protein−coupled receptor signaling
cell−cell signaling
cell communication
signaling
−log10 p−adj
20 25 30 35 40
Number of genes
30
90
150

Figure S2

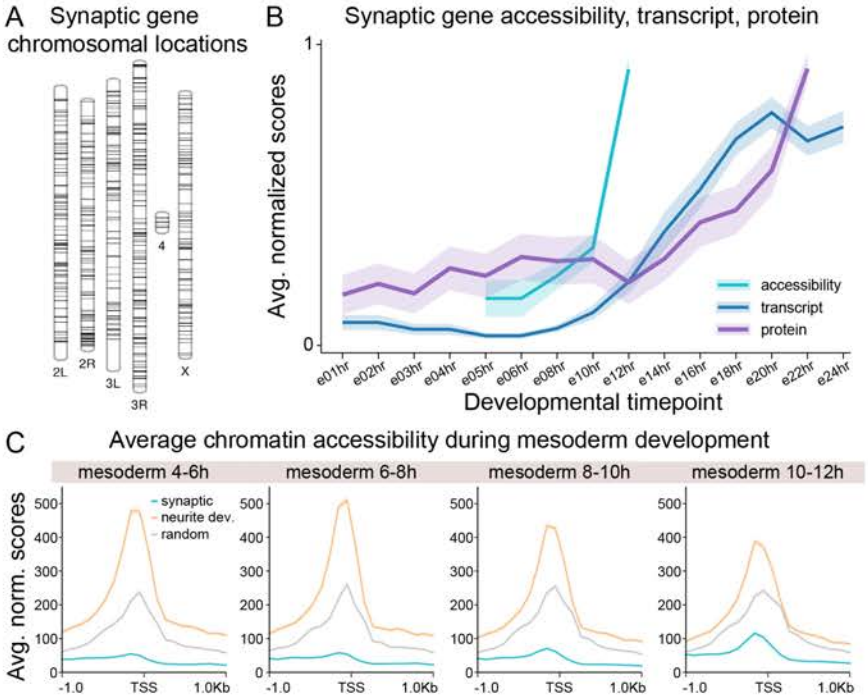
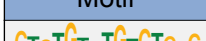
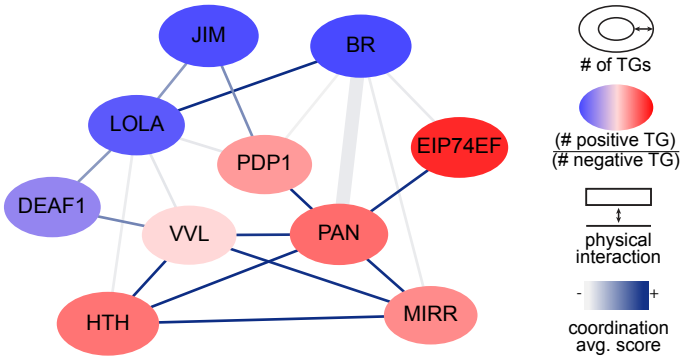

A Synaptic gene chromosomal locations
2L
2R
3L
3R
4
X
B Synaptic gene accessibility, transcript, protein
Avg. normalized scores
1
0
accessibility
transcript
protein
e01hr
e02hr
e03hr
e04hr
e05hr
e06hr
e08hr
e10hr
e12hr
e14hr
e16hr
e18hr
e20hr
e22hr
e24hr
Developmental timepoint
C Average chromatin accessibility during mesoderm development
mesoderm 4-6h
mesoderm 6-8h
mesoderm 8-10h
mesoderm 10-12h
Avg. norm. scores
synaptic
neurite dev.
random
500
400
300
200
100
0
-1.0
TSS
1.0Kb

Figure S3

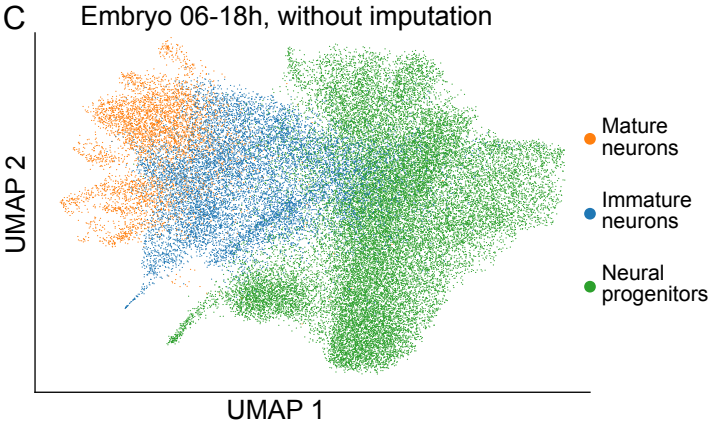
A

Motif	E-value	# of sites	Best match
	4.7×10^{-249}	334	klu
	3.4×10^{-206}	445	CLAMP/GAF
	3.7×10^{-111}	591	kni/jim
	1.0×10^{-34}	330	lola
	3.0×10^{-29}	178	Mad
	5.6×10^{-24}	70	l(3)neo38
	3.8×10^{-14}	110	Adf1
	4.7×10^{-5}	101	slp1

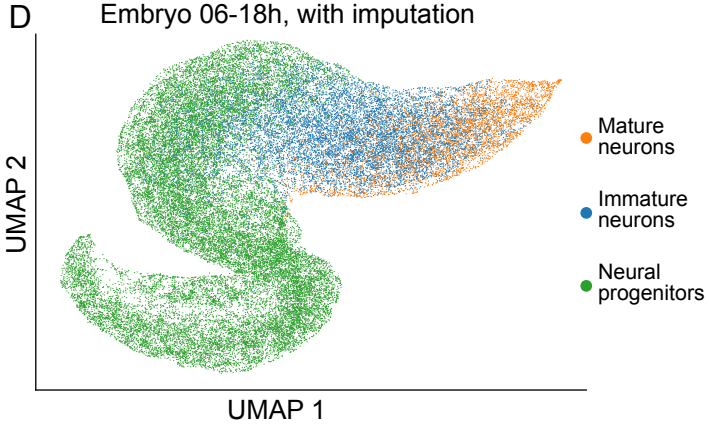
B



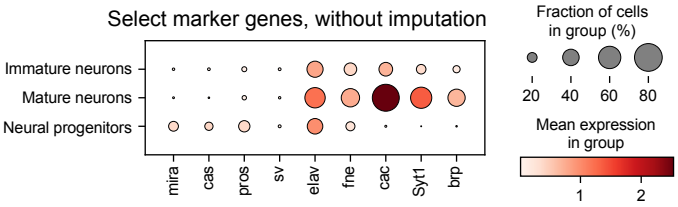
C



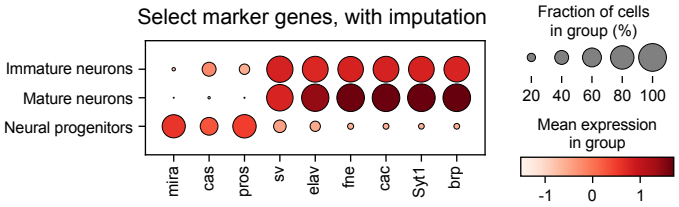
D



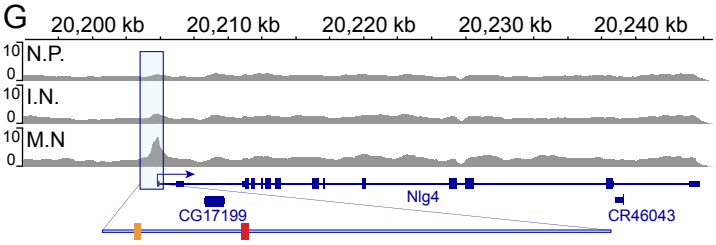
E



F



G



H

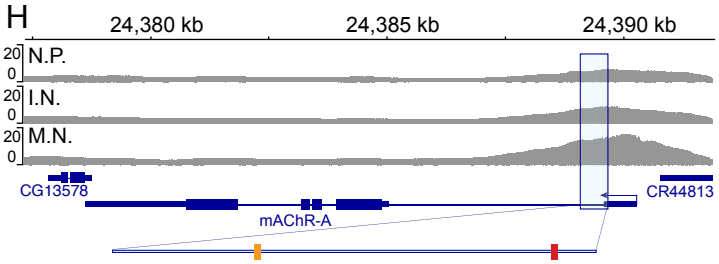


Figure S4

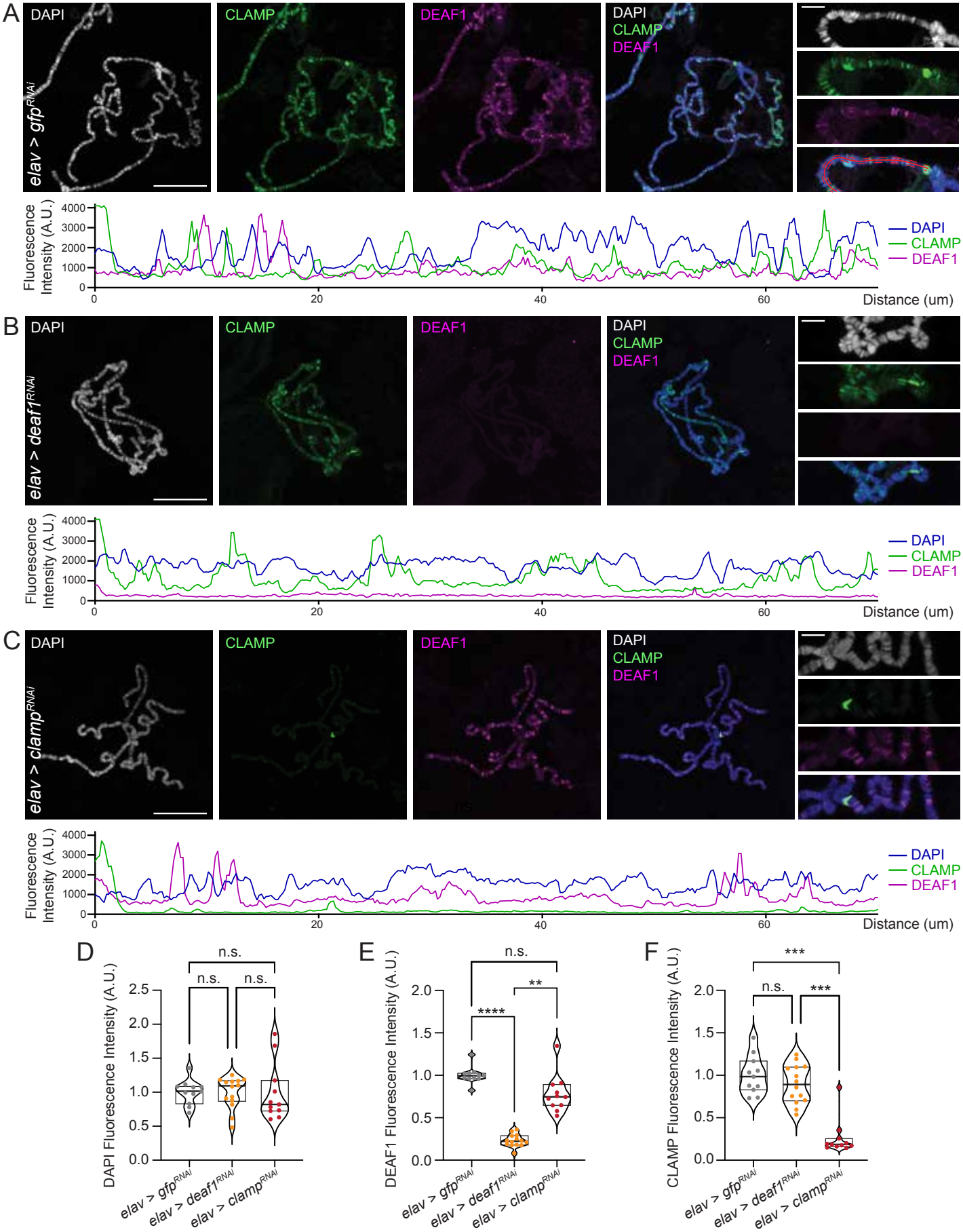

A
DAPI
CLAMP
DEAF1
DAPI
CLAMP
DEAF1
elav > gfp^RNAi
Fluorescence Intensity (A.U.)
Distance (um)
B
elav > deaf1^RNAi
C
elav > clamp^RNAi
D
DAPI Fluorescence Intensity (A.U.)
n.s.
E
DEAF1 Fluorescence Intensity (A.U.)
****
**
F
CLAMP Fluorescence Intensity (A.U.)
***
elav > gfp^RNAi
elav > deaf1^RNAi
elav > clamp^RNAi

Figure S5

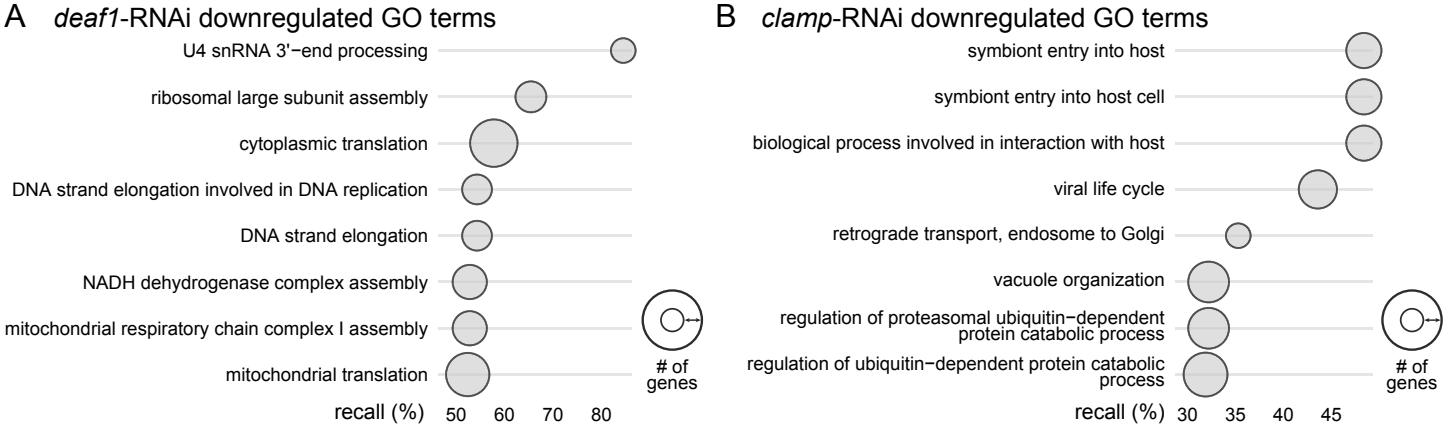
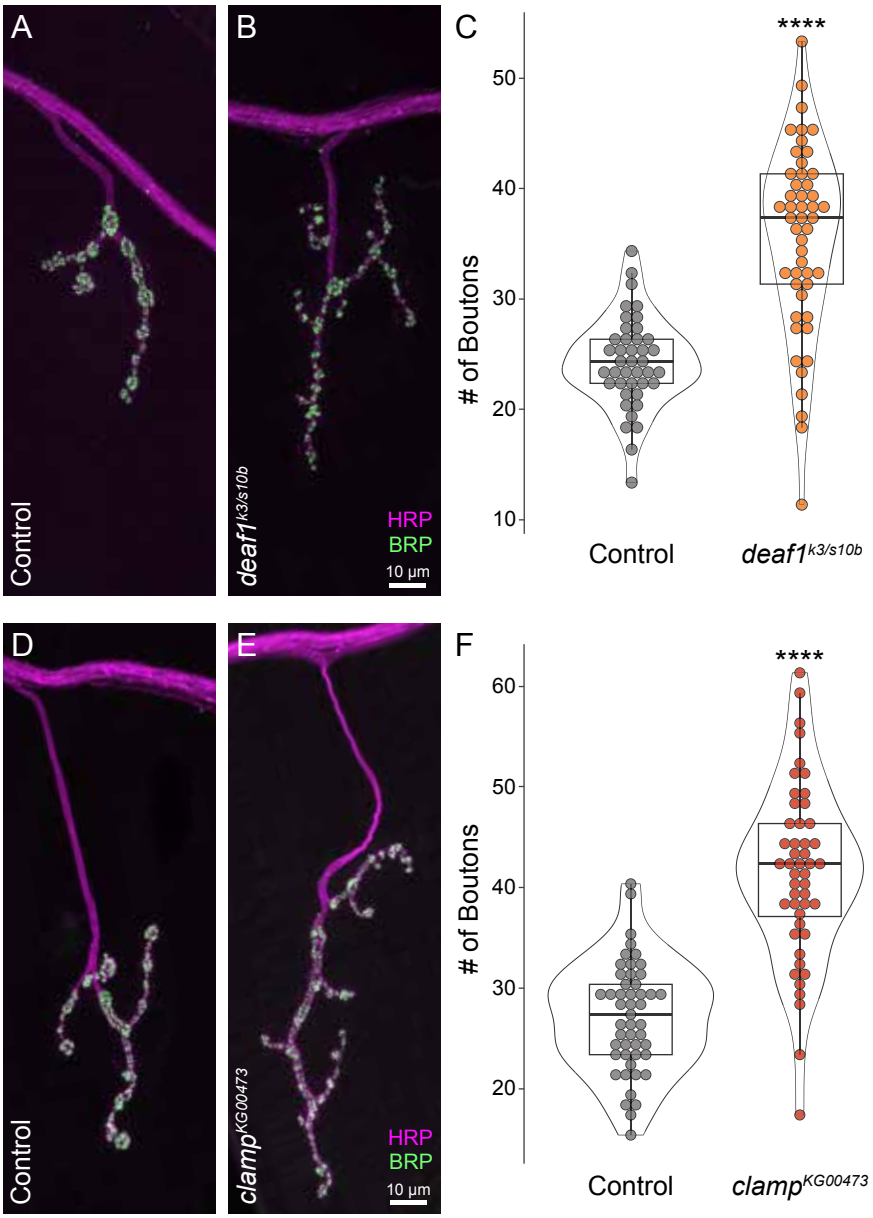

A
deaf1-RNAi downregulated GO terms
U4 snRNA 3'−end processing
ribosomal large subunit assembly
cytoplasmic translation
DNA strand elongation involved in DNA replication
DNA strand elongation
NADH dehydrogenase complex assembly
mitochondrial respiratory chain complex I assembly
mitochondrial translation
recall (%)
50
60
70
80
# of genes
B
clamp-RNAi downregulated GO terms
symbiont entry into host
symbiont entry into host cell
biological process involved in interaction with host
viral life cycle
retrograde transport, endosome to Golgi
vacuole organization
regulation of proteasomal ubiquitin−dependent protein catabolic process
regulation of ubiquitin−dependent protein catabolic process
recall (%)
30
35
40
45
# of genes

Figure S6


A
Control
B
deaf1k3/s10b
HRP
BRP
10 µm
C
****
# of Boutons
50
40
30
20
10
Control
deaf1k3/s10b
D
Control
E
clampKG00473
HRP
BRP
10 µm
F
****
# of Boutons
60
50
40
30
20
Control
clampKG00473