Gene regulatory dynamics during the development of a paleopteran insect, the mayfly *Cloeon dipterum*

Joan Pallarès-Albanell* 1,2, Laia Ortega-Flores* 1, Tòt Senar-Serra 1,2, Antoni Ruiz 1,2, Josep F. Abril 1,3, Maria Rossello# 1,2, Isabel Almudi# 1,2

1 Department of Genetics, Microbiology and Statistics, Universitat de Barcelona, Diagonal 643, 08028, Barcelona, Spain
2 Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Diagonal, 643, 08028, Barcelona, Spain
3 Institute of Biomedicine of Universitat de Barcelona (IBUB), Universitat de Barcelona, Diagonal, 643, 08028, Barcelona, Spain

*Equally contributed
#correspondence to ialmudi@ub.edu, mariarossello@ub.edu

KEY WORDS: Gene regulation, Paleoptera, mayflies, insects, embryogenesis, ATAC-seq.

Abstract

The evolution of insects has been marked by the appearance of key body plan innovations and novel organs that promoted the outstanding ability of this lineage to adapt to new habitats, boosting the most successful radiation in animals. To understand the origin and evolution of these new structures, it is essential to investigate which are the genes and gene regulatory networks participating during the embryonic development of insects. Great efforts have been made to fully understand, from a gene expression and gene regulation point of view, the development of holometabolous insects, in particular *Drosophila melanogaster*, with the generation of numerous functional genomics resources and databases. Conversely, how hemimetabolous insects develop, and which are the dynamics of gene expression and gene regulation that control their embryogenesis, are still poorly characterized. Therefore, to provide a new platform to study gene regulation in insects, we generated ATAC-seq (Assay for transposase-Accessible Chromatin using sequencing) for the first time during the development of the mayfly *Cloeon dipterum.* This new available resource will allow to better understand the dynamics of gene regulation during hemimetabolan embryogenesis, since *C. dipterum* belongs to the paleopteran order of Ephemeroptera, the sister group to all other winged insects. These new datasets include six different time points of its embryonic development and identify accessible chromatin regions corresponding to both general

and stage-specific promoters and enhancers. With these comprehensive datasets, we characterised pronounced changes in accessible chromatin between stages 8 and 10 of embryonic development, which correspond to the transition from the last stages of segmentation to organogenesis and appendage differentiation. The application of ATAC-seq in mayflies has contributed to identify the epigenetic mechanisms responsible for embryonic development in hemimetabolous insects and it will provide a fundamental resource to understand the evolution of gene regulation in winged insects.

INTRODUCTION

Insects are the most numerous and diverse lineage of animals on the planet (Misof et al. 2014). This huge radiation has been possible due to their extraordinary capabilities to adapt to distinct environments and the myriad of life history traits evolved within this animal class, which resulted in more than thirty extant orders distributed worldwide (Misof et al. 2014; D. Grimaldi & M. S. Engel 2007).

This diversity of forms and lifestyles is the result of changes in the gene regulatory networks (GRNs) controlling the embryonic and postembryonic development of this clade (Carroll 1998; Molina-Gil et al. 2023). In recent decades, the importance of regulatory information responsible for the location and time in which genes and GRNs are functioning, has been widely recognised (Furlong and Levine 2018; Leyhr et al. 2022; Gompel et al. 2005; Andrikou and Arnone 2015). These so called *cis* regulatory elements (CREs) are major players of morphological evolution not only in insects, but also in other animal lineages (Guerreiro et al. 2013; Leyhr et al. 2022). Still, some changes in trans, -in the coding sequences of transcription factors (TFs) and/or signalling molecules- have been also shown to participate in the diversity of phenotypes observed in animals (Galant and Carroll 2002; Santos et al. 2017). Thus, both changes in coding regions and CREs are key to explain the wide range of insect (and animal) morphologies. Comparative genomics projects (Crowley et al. 2023; Formenti et al. 2022; Molik 2022) promoted a better understanding of changes in trans elements, due to their higher degree of conservation between species. By contrast, CREs are more difficult to identify by homology, since they are usually less pleiotropic and for this, they tend to accumulate more changes in their sequences that impede their proper characterisation by sequence similarity. This is even more manifest in insects, with increased rates of genome evolution when comparing with vertebrate lineages (Zdobnov and Bork 2007). In addition, comparative transcriptomics provide information about the spatio-temporal dynamics of gene expression between different lineages (Mantica et al. 2024; Rodríguez-Montes et al. 2023; Levin et al. 2016), which may reflect to some degree differences in GRNs. The advent of new functional genomics approaches based on chromatin accessibility, such as, first, formaldehyde-assisted isolation of regulatory elements (FAIRE-seq (J. M. Simon et al. 2012)) and the more recent Assay for Transposase-Accessible Chromatin (ATAC-seq (Buenrostro et al. 2013)), allow the identification of open chromatin regions that can be assigned as CREs, such as enhancers, promoters and insulators. The advantages of these methods, in comparison to other functional genomics assays used before (e.g. ChIP-seq), are the low input of starting material, the high-throughput protocol without expensive reagents or antibodies and that they are not excessively laborious, among others. Therefore, much more detailed and comprehensive characterisations of CREs have been possible and several FAIRE-seq and ATAC-seq datasets have been

generated during the last decade to address distinct questions related to insect development and evolution. Nonetheless, these works have been mostly done using few established model species, such as the fruitfly *Drosophila melanogaster* (Diptera), the red flour beetle *Tribolium castaneum* (Coleoptera), some butterfly and moth species (Lepidoptera) or certain species of ants and bees (Hymenoptera) (Lowe et al. 2022; Wang et al. 2020; Zhao et al. 2020), all of which belong to a single monophyletic group of insects, the Holometabola (Fig. 1A)(Schmidt-Ott and Lynch 2016). By contrast, the remaining seventeen orders of winged insects, the Hemimetabola (Fig. 1A), continue unexplored, with only adults and one embryonic FAIRE-seq datasets available (Fernández et al. 2020; Richard et al. 2017) to our knowledge.

Both hemimetabolous and hemimetabolous orders shared common phases during their development (Sander 1976; Patel 1994). Indeed, these main events during embryogenesis are also present in non-insect arthropods (Zrzavý and Štys 1997). The early stages of insect development initiate with several rounds of nuclear divisions that migrate to the periphery to form the blastoderm. The blastoderm gives rise to different types of germ band (short, intermediate or long) which are subsequently segmented during the following embryonic stages (Davis and Patel 2002). Then, the segmented embryo undergoes a process of differentiation in which organogenesis and the final development of the juvenile structures take place (Sander 1976).

The mayfly *Cloeon dipterum,* a recently established laboratory system (Almudi et al. 2020; 2019), is in a privileged position to improve the phylogenetic diversity of insect functional genomics resources. Mayflies or Ephemeroptera belong to the Paleoptera group of winged insects, together with Odonata (dragonflies and damselflies)(Fig. 1A) (S. Simon, Blanke, and Meusemann 2018). They are the sister group to all other winged insects and thus, due to this position in the insect phylogenetic tree, they are key to address fundamental questions related to insect physiology, ecology, development and evolution.

Here, we performed ATAC-seq experiments at six different developmental stages, -including some of the hallmark stages of insect embryogenesis mentioned above-, in *C. dipterum* embryos. Our identified accessible chromatin regions provide an exhaustive collection of putative promoters and enhancers along embryogenesis of this hemimetabolan insect. Moreover, by studying the temporal dynamics of these elements, we showed wholesale changes in chromatin accessibility during the transition between the last stages of segmentation and the start of organogenesis and appendage differentiation. Finally, we facilitate the access to these comprehensive datasets through a dedicated web browser (https://genome-euro.ucsc.edu/s/mayfly/Clodip), providing a key resource available for the entire community to understand the evolution of gene regulation during the development of winged insects.

RESULTS AND DISCUSSION

Open Chromatin profiles in *C. dipterum* embryogenesis

To investigate dynamics of gene regulation during the embryogenesis of mayflies, we performed ATAC-seq assays for six different developmental stages: stage (st) 4 (germ band elongation, Fig 1B, B'), st 6 (S-shaped embryo: anatrepsis II Fig. 1C, C'), st 8 (segmentation of the embryo, Fig. 1D, D'), st 10 (revolution: katatrepsis, Fig. 1E, E'), st 12 (start of dorsal closure, Fig. 1F, F'), st 14 (dorsal closure complete, Fig. 1G, G')(Tojo and Machida 1997). Sequences resulting from these experiments were mapped against the *C. dipterum* reference genome assembly (CLODIP2 (Almudi et al. 2020), see methods and Fig. S1) to obtain a non-redundant collection of open chromatin regions throughout the genome that we termed APREs (Accessible Putative Regulatory Elements) (Fig. 1H). After normalization (see methods) we identified a total of 54,547 APREs across the six developmental stages. Of them, 45,649 APREs did not show changes in accessibility in our clustering analyses of the different developmental samples we assayed (i.e. they remained constitutively open or close across these stages) while 8,898 APREs were dynamic and changed their accessibility across developmental stages (Fig. S2 and Table S1).

We next aimed at defining the genomic distributions of the APREs relative to genes and gene annotations. For this, we calculated the proportions of APREs at the "promoter" (i. e. APREs in the immediate vicinity of the annotated transcription start sites (TSSs)), at "proximal regions", spanning up to 5 Kb upstream the TSSs, at "gene bodies" (located between the end of the promoter and the termination site of the gene) and "distal" regions, which comprised genomic regions that do not fall in the previous categories (Fig. 1I, Table S2). We found that both non-dynamic and dynamic APREs were distributed in similar proportions (Fig. S3) and only detected a slight increase in non-dynamic APREs located in promoters with respect to dynamic APREs in promoters (~18% versus ~13%) and an even slighter difference between non-dynamic and dynamic APREs in gene bodies (26% and 29%, respectively, Fig. S3). This proportion of APRE distribution was similar to the distribution of APREs in other invertebrate genomes, such as the chordate amphioxus (*Branchiostoma lanceolatum*) (Marlétaz et al. 2018). These two invertebrate species, *C. dipterum* and *B. lanceolatum,* notably diverged from the distribution found in some vertebrate species, which showed much larger proportion of distal APREs due to the relevance of distal regulation in these vertebrate genomes and the impact of the different rounds of whole genome duplications (Marlétaz et al. 2018). By contrast, when we compared the APRE distribution of *C. dipterum* with the distribution of accessible chromatin in *D. melanogaster,* we observed that fruitflies had a much lower proportion of distal APREs (i.e. a third of the corresponding fractions in *C. dipterum* and

B. lanceolatum), with a higher proportion of APREs located in gene bodies and proximal regions (Fig. 1I)(Bozek et al. 2019). These differences between *C. dipterum* and *D. melanogaster* were most likely due to the higher compaction of the 120 Mb *D. melanogaster* euchromatic genome (Adams et al. 2000), where most of the genomic regulatory blocks ancestral to animals and their associated long range regulatory interactions have been dismantled (Irimia et al. 2012).

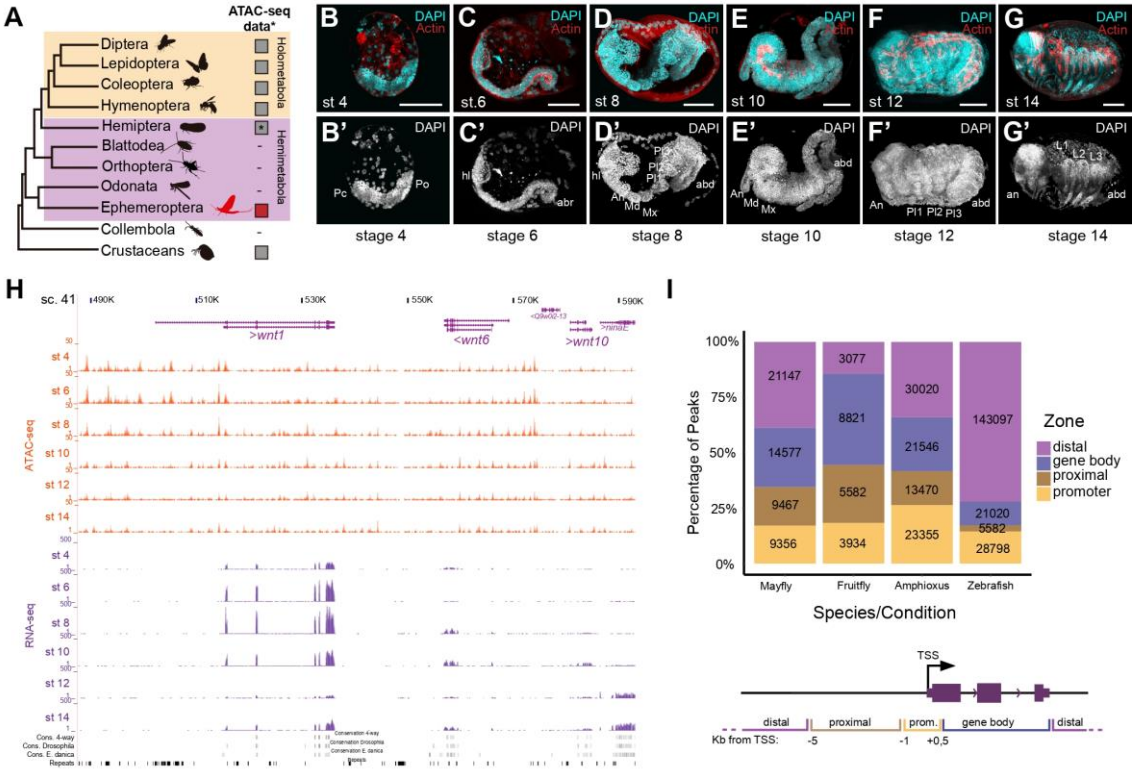


Figure 1. Open Chromatin profiles in *C. dipterum* embryogenesis. (A) Simplified insect phylogeny. Grey squares highlight the availability of ATAC-seq datasets. Asterisk shows lineages in which there is no ATAC-seq information but FAIRE-seq material has been generated. **(B-G)** Embryonic stages used in this study stained with Phalloidin-AlexaFlour-488 (Red) and DAPI (cyan). (B-B') Stage 4 embryo (St 4): germ band initial elongation. (C-C') St 6: S-shaped embryo: anatrepsis II. (D-D') St 8: Segmentation of the embryo. (E-E') St 10 embryo: revolution or katatrepsis. (F-F') St 12: initial dorsalization. (G-G') St 14 embryo: dorsal closure completed. **(H)** Snapshot of the UCSC genome browser showing *wnt1, wnt6, wnt10* cluster and the ATAC-seq tracks in this study and and RNA-seq tracks generated in this study and in (Almudi et al. 2020). **(I)** Percentage of ATAC-seq APREs distributed across *C. dipterum*, *D. melanogaster, B. lanceolatum* and *D. rerio* genomes. Scale bars: 50 um.

ATAC-seq revealed a temporally regulated chromatin profile in the mayfly genome

In order to study changes in chromatin accessibility throughout different samples (i.e. developmental timepoints), we analysed differential APRE activity between consecutive developmental stages. These results showed relatively modest changes between successive timepoints (i.e. 164 APREs at the St 4 to St 6 transition or 365 at St 10 to St 12, Fig. S4, Table S3), with the notable exception of the transition between St 8 and St 10, when more than four thousand APREs (4568) changed their accessibility, from 7,5 to 27 times more than at the other timepoints (Fig. S4). From these, the vast majority (3118) corresponded to APREs that increase accessibility during this St 8-St 10 phase. This large amount of differentially active APREs between these two stages may indicated a major regulatory turnover between an early and late regulatory state during mayfly embryogenesis. In fact, this major shift was also evident when we performed a PCA analysis and clustering of the ATAC-seq datasets, which formed two very distinctive clusters: samples from stages 4, 6 and 8 and samples from 10, 12 and 14 (Fig. S4).

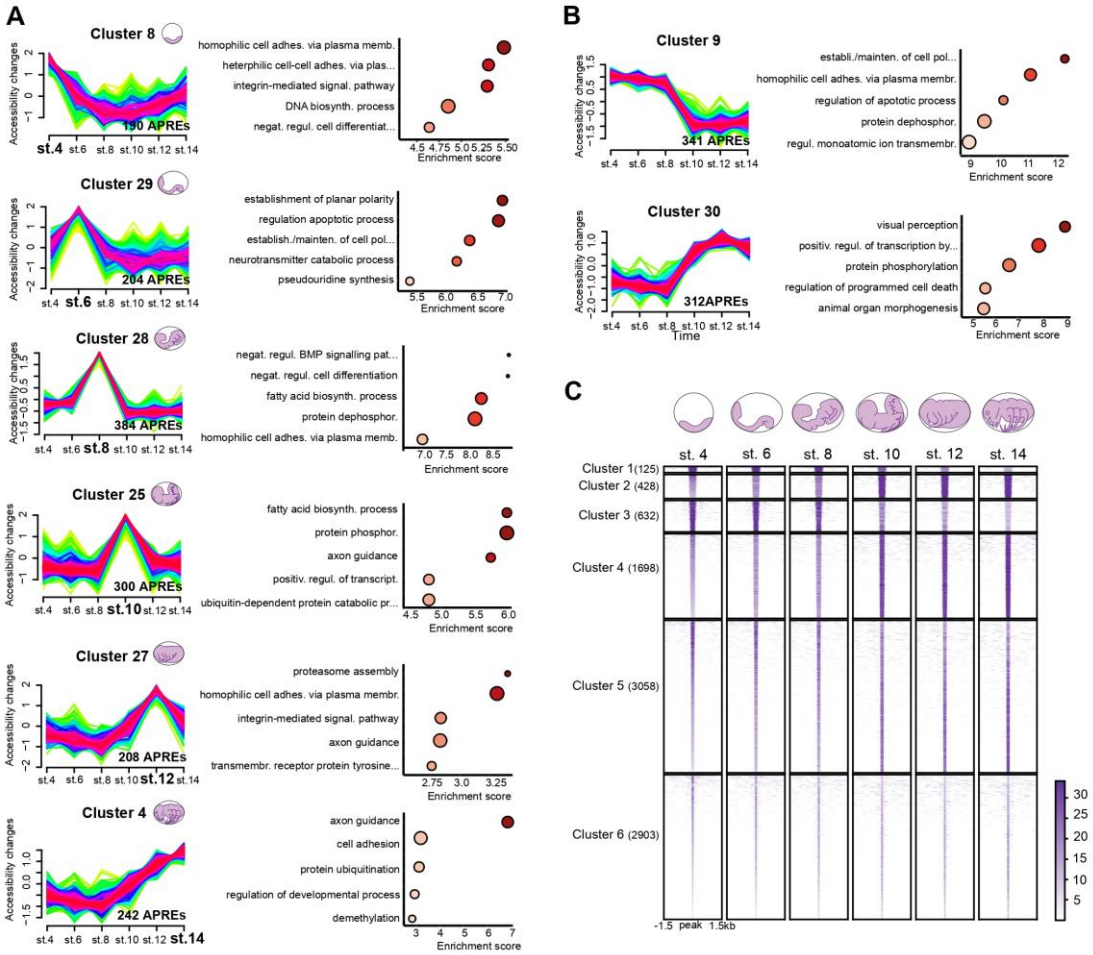
To further explore the dynamics of chromatin accessibility across the selected developmental timepoints, we performed a temporal soft-clustering analysis using Mfuzz (Kumar and Futschik 2007)(see methods and Table S4). Among the different temporal clusters obtained (Fig. S5), we focused on clusters of APREs whose accessibility peaked at a single embryonic stages (e.g. cluster 8 and 16 for st 4, cluster 29 and 21 for st 6, cluster 28, 5, 6, 11, 13, 14 and 26 for st 8, cluster 25, 15 and 18 for st 10, cluster 27 for st 12 and cluster 4 and 2 for st 14; Fig 2A and Fig. S5). We then associated these APREs to their putative target genes and analysed the Gene Ontology (GO) enriched terms for each of these stage-specific clusters, using *D. melanogaster* orthologs (Almudi et al. 2020), since it is the closest organism with functional annotation available (see methods and Table S5). This GO term enrichment analysis exhibited categories highly connected with each of the embryonic stages in which the accessibility of the chromatin was higher (Fig. 2A and Table S5). For instance, cluster 8 (st 4) and cluster 29 (st 6) contained APREs that were associated to genes involved in cell adhesion, planar cell polarity, DNA biosynthetic process and negative regulation of cell differentiation, which are characteristic processes of early embryogenesis in insects (Münster et al. 2019; Brantley and Di Talia 2021). On the other hand, GO terms enriched in clusters corresponding to later stages of embryogenesis (e.g. cl 25, cl 27 and cl 4, Fig 2A and Table S5) revealed processes related to organogenesis, such as axon guidance or regulation of developmental process (Gillott 2005).

Besides these stage-specific clusters, we also found several clusters that showed more prolonged activity patterns. In this manner, we identified clusters that recapitulated the major developmental shift we had previously observed between st 8 and st 10, with

a set of "early embryogenesis" clusters (e.g. clusters 7, 9, 17, 19, 21 or 26) and another of "late embryogenesis"ones (clusters 3, 22, 23 or 3; Fig. 2B and Fig. S5). Accordingly, we identified enriched GO terms related to early development (e.g. establishment and maintenance of cell polarity or cell adhesion processes) and terms related to late embryogenesis (e.g. visual perception and animal organ morphogenesis), respectively (Fig. 2B and Table S5).

In addition, we also performed *k-means* hard clustering (see methods, Table S6) using the same set of dynamic APREs (Fig. 2C). This analysis was able to recover six clusters with differential dynamics of accessibility, although none of these clusters corresponded to APREs showing stage-specific accessibility. By contrast, we identified two groups of clusters, - cluster 1 and cluster 3, on one hand, and cluster 2 and cluster 4, on the other hand-, that contained APREs accessible during early stages of embryogenesis and late stages of embryogenesis, respectively (Fig. 2C and Table S6), mirroring the st 8-st 10 shift observed in the previous analyses. In order to better characterise these clusters, we carried out TF motif enrichment analysis using Homer software (see methods, Fig. 2D and Table S7). In agreement with previous work in other insects and the results of the GO term enrichment analysis from the Mfuzz clusters (Fig. 2A, B), for APREs in cluster 1 and cluster 3, we identified motifs whose best match were TFs involved in early embryogenesis. These include Hunchback (Hb), a gap gene involved in antero-posterior axis specification (Qian, Capovilla, and Pirrotta 1991), Caudal (Cad), which functions in germ band elongation (Schulz and Tautz 1995; Wu and Lengyel 1998), or Zelda (Zld), a zygotic genome activator that acts during early blastoderm development (Brennan et al. 2023; Liang et al. 2008) (Fig. 2D). By contrast, clusters 2 and 4, whose APREs were open in later stages of development, showed enrichment in motifs that correspond to TFs involved in different processes of organogenesis, such as Nubbin (Nub), a regulator of appendage morphogenesis (Turchyn et al. 2011), Six4, involved in the development of mesodermal structures (Clark et al. 2006), or Mothers against dpp (Mad), that mediates the response to the BMP pathway during the development of diverse insect organs (Sekelsky et al. 1995).

Overall, these analyses revealed two main phases during the embryogenesis of mayflies in which distinct set of regulatory regions are active (Fig. 2B, C) to control different sets of genes and regulatory networks involved in such early or late embryonic processes. These results were consistent with the mid-developmental transition previously described at transcriptomic level for some phyla, including insects (Levin et al. 2016).



D

Early clusters

	Enriched motifs	Best match TF	p-values	% sites
Cluster 1	#2	Slbo	1e-25	77.60 %
	#3	Hb	1e-22	79.20 %
	#4	Zelda	1e-21	69.60 %
	#5	Deaf1	1e-20	34.40 %
	#6	Caudal	1e-18	96.80 %
Cluster 3	#1	Zelda	1e-60	67.56 %
	#2	Dichaete	1e-56	62.97 %
	#3	Pleiohomeotic	1e-49	78.80 %
	#4	Vismay	1e-49	77.53 %
	#5	Runt	1e-43	47.94 %

Late clusters

	Enriched motifs	Best match TF	p-values	% sites
Cluster 2	#1	Nub	1e-51	92.12 %
	#2	Six4	1e-46	86.51 %
	#3	Trl	1e-45	60.58 %
	#5	Dref	1e-40	59.75 %
	#6	Su(H)	1e-39	82.37 %
Cluster 4	#4	Abd-A	1e-69	73.97 %
	#5	Pangolin	1e-66	95.29 %
	#6	Kruppel	1e-65	89.16 %
	#7	Mad	1e-64	80.15 %

Figure 2. ATAC-seq revealed a temporally regulated chromatin profile in the mayfly genome. (A) Mfuzz clusters obtained for the 8,898 dynamic APREs obtained through ATAC-seq experiments in the six selected developmental stages representing stage-specific APRE activity and their associated GO enriched terms. **(B)** Mfuzz clusters representing "early embryogenesis" and "late embryogenesis" APRE activity and their associated GO enriched terms. **(C)** Heatmaps of the 8,898 dynamic APREs clustered using k-means clustering. Six clusters were obtained, with four of them showing a clear dynamic behavior: cluster 1 (n=125) and cluster 3 (n=632) or early activity, cluster 2 (n=428) and cluster 4 (n = 1698) or late activity. **(D)** Motif enrichment analysis of the early (cluster 1, cluster 3) and late (cluster 2, cluster 4) active clusters. Five or four representative motifs of the top-10 were chosen. Motif logos are represented with their position in the top-10, the TF names, the enrichment p-values and the percentage of sites showing the motif.

Chromatin accessibility to understand gene expression dynamics

Since ATAC-seq has been proven to be a powerful method to investigate the regulation of gene expression, we next addressed the relationship between chromatin accessibility and levels of gene expression (Starks et al. 2019). To do this, we measured the levels of gene expression at the same developmental stages of our ATAC-seq datasets (see methods Fig. S6 and Table S8). When focusing on genes associated to the 8,898 dynamic APREs, we first observed that the six stages clustered accordingly to the "early" and "late" embryonic phases that we identified in the ATAC-seq data. Transcriptomes from stages 4, 6 and 8 formed a cluster while transcriptomes from stages 10, 12 and 14 grouped together (Fig. 3A). Moreover, we detected some genes whose expression varied along the developmental timepoints we characterised: cluster 11 and cluster 7 decreased their expression as embryogenesis progressed, while genes from cluster 3, 5 or 9 increased their expression during embryonic development (Fig. 3A).

To further illustrate our results, we investigated the expression pattern of *embryonic lethal abnormal vision* (*elav*) that showed differential APRE accessibility and differential gene expression along embryonic stages (Fig. 3B). *Elav* is a RNA binding protein involved in axon guidance, synapse formation and development and maintenance of neurons (Robinow and White 1988). We performed Hybridization Chain Reaction (HCRs) assays (Bruce et al. 2021) in st 6, st 8 and st 10 embryos to characterise the spatial expression of this gene in these stages in which we observed shifts in chromatin accessibility and expression levels. While embryos at st 6 showed reduced expression of *elav* in some cells in the cephalic region (Fig. 3C), at st 8 *elav* exhibited a broader expression domain in head domains and in the most anterior thoracic segments. At st 10, these neural territories of *elav* expression expanded and elongated

to the abdominal segments (Fig. 3C). As expected, we also observed an increase in the accessibility of chromatin in the locus, especially upstream of the Transcription Start Site (TSS, Fig. 3B).

Previous analysis of co-regulated gene expression across several tissues, using Weighted Gene Correlation Network Analysis (WGCNA), revealed modules of genes specifically expressed in particular adult and nymphal tissues (Almudi et al. 2020). We examined whether chromatin accessibility information from embryos correlated with these modules. We characterised enriched motifs in APREs associated to the genes contained in each of these WGCNA modules and found distinctive enriched binding motifs for some of them (see methods and Table S9). For example, the brain module was enriched in neural motifs, such as Suppressor of Hairy (Su(H)), retained (retn) or Cut (Ct), which are TFs involved in neural or glial development (Grueber and Jan 2004; Shandala et al. 1999), the muscle module had motifs for bagpipe (bap) or twist (twi), required for mesodermal development (Castanon et al. 2001; Cripps and Olson 2002) and specification or the gills module with enrichment in motifs such as defective proventriculus (dve) involved in epithelium patterning or glial cells missing (gmc), which could have a role in the determination of some of the numerous neural cells that we previously identified in these abdominal structures (Almudi et al. 2020; Hosoya et al. 1995) (Fig. 3D, and Table S9). These results suggest that some of the Gene Regulatory Networks involved in the development of nymphal tissues and organs are already functioning during embryogenesis and their regulatory signatures can be detected in our ATAC-seq datasets. Thus, our results can also provide important insights into the regulatory logic of the adult body plan, and therefore also constitute a valuable resource for adult insect biology.

Overall, our ATAC-seq datasets provide a comprehensive resource to help uncovering developmental diversity of insects, since it represents the first publicly available genome-wide collection of putative regulatory elements across embryogenesis in a hemimetabolous lineage using ATAC-seq approaches. Thus, the key phylogenetic position of Ephemeroptera, together with the extensive chromatin accessibility information made available here, will open new venues to address longstanding questions in the fields of developmental and evolutionary biology and comparative genomics.

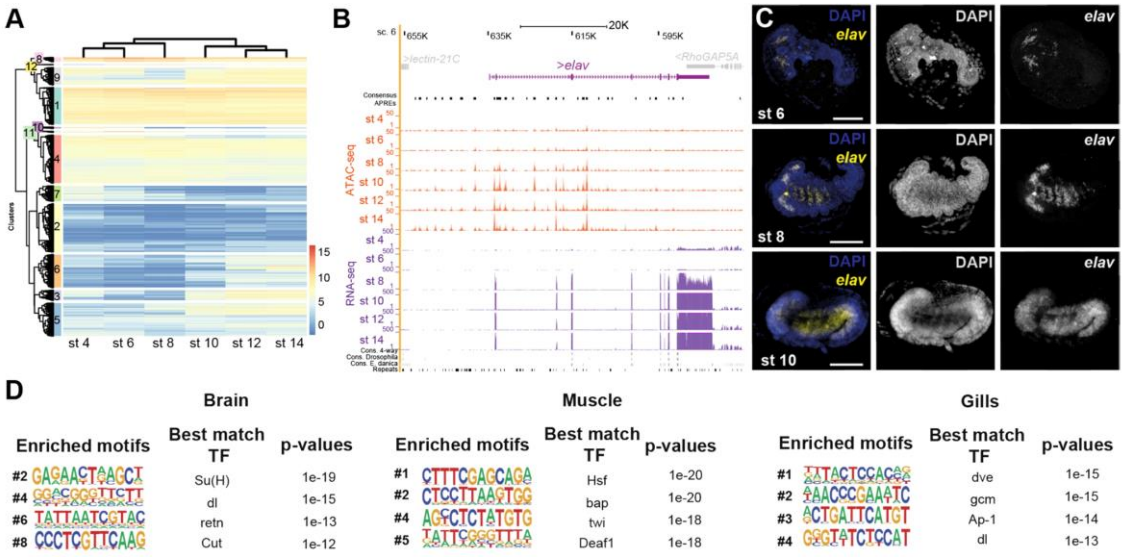


Figure 3. Chromatin accessibility to understand gene expression dynamics. (A) Heatmap showing expression levels of genes associated to dynamic ATAC-seq APRE in the selected embryonic stages. RNA-seq samples clustered according to embryonic stage progression. Secondary clustering showed 12 different gene clusters. **(B)** *elav* genomic regulatory landscape. APRE activity and gene expression increase as embryogenesis progresses. **(C)** HCR hybridization against the gene *elav* at st 6, st 8 and st 10 of embryogenesis. Nuclei were stained with DAPI (dark blue) and *elav* expression pattern is shown in yellow. Scale bars: 50 um. **(C)** Motif enrichment analysis of APREs associated to genes from some tissue-specific WGCNA modules identified in (Almudi et al. 2020). Four representative motifs of the top-10 were chosen. Motif logos are represented with their position in the top-10, the TF names and the enrichment p-values.

MATERIALS & METHODS

Culture maintenance, embryo collection and fixation

Samples were obtained from a *Cloeon dipterum* culture maintained in the laboratory as previously described in (Almudi et al. 2019). Gravid females fertilised different days were collected and dissected to obtain embryos at selected developmental stages: st 4, st 6, st 8, st 10, st 12, st 14. After opening the abdomen of these gravid females, embryos were collected to perform ATAC-seq or RNA-seq procedures and a small subset was collected apart and fixed with 4% Formaldehyde for 1 hour at r.t. to confirm the developmental stage. After 3 x 5' washes with PBS, these fixed embryos were stained with Phalloidin Alexa Fluor™ 488 (A12379) and DAPI to visualise actin filaments and nuclei, respectively. Images were acquired using a Zeiss LSM 880 confocal and were processed with Fiji (Schindelin et al. 2012).

HCR hybridization

HCR hybridization followed a modified version of the Molecular Instruments (Los Angeles, CA, USA) HCR v.3 protocol (Bruce et al. 2021). HCR probe was designed to evade non-specific binding using an open-source probe design program (Kuehn et al. 2022). Briefly, embryos stored in ethanol were rehydrated in stepwise 75/50/25% ethanol in PBTw 0.1%. After 3 x 5' washes in PBTw 0.1%, embryos were permeabilized in Detergent Solution (1.0% SDS, 0.5% Tween, 50.0 mM Tris-HCl (pH 7.5), 1.0 mM EDTA (pH 8.0), and 150.0 mM NaCl) for 30' at room temperature (RT), kept in pre-warmed Probe Hybridization Buffer (Molecular Instruments) for 30' at 37 ºC, and incubated in Probe Solution (4 nM of probe in Probe Hybridation Buffer) overnight at 37 ºC. After 4 x 15' washes in pre-heated wash buffer (Molecular Instruments) at 37 ºC and 2 x 5' washes in 5x SSCTw 0.1% at RT, they were kept in pre-equilibrated Amplification Buffer (Molecular Instruments) for 30' at RT and incubated in hairpin solution (60 nM of each hairpin h1 and h2 (Molecular Instruments) separately in pre-equilibrated Amplification Buffer, heated at 95 ºC for 90 s and cooled down for 30' overnight in the dark at RT. Following 5 x 20' washes in 5x SSCTw 0.1% and 1 x 10' wash in PBTw 0.1 % pH 7.4 in dark at RT, embryos were mounted in Prolongue™ Gold with DAPI (P36941, Invitrogen). Images were acquired using a Zeiss LSM 880 confocal and were processed with Fiji (Schindelin et al. 2012).

RNA-seq sequencing and assembly

Three RNA-seq datasets (including replicates) of st 8 and st 12 embryos were generated using the Illumina technology. Samples were processed immediately after dissection and RNA was extracted using RNeasy Mini Kit (Qiagen) following manufacturers' instructions. Paired-end libraries were generated using Illumina (Novaseq6000) 2x50bp. After quality control, the obtained reads were aligned using the STAR aligner. Initially, a genome index was created using the CLODIP2 reference genome (GCA_902829235.1) with the genomeGenerate mode of STAR. Subsequent alignment of reads to this index was performed using the alignReads mode. Gene expression levels were quantified utilizing the quantMode GeneCounts option within STAR.

ATAC-seq and library preparation

ATAC-seq or assay for transposase-accessible chromatin by sequencing protocol was optimised during these experiments to use on mayflies from (Buenrostro et al. 2015). Briefly, embryos were homogenised in lysis buffer (10 mM Tris–HCl pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% NP-40) to obtain approximately 70000 individual nuclei. After removing lysis buffer, transposition reaction (1.25 µl of Tn5 enzyme in 10 mM Tris–HCl pH 8.0, 5 mM $MgCl_2$, 10% w/v dimethylformamide) was performed for 30 min at 37 ºC and the resulting fragments are purified using MinElute PCR Purification Kit (Qiagen). qPCR was performed to determine the optimal number of cycles necessary for each

library. A unique pair of primers were assigned to each sample (Table S10) and 12 libraries were prepared corresponding to two biological replicates of the six selected developmental time points using a PCR. Libraries were purified using the MinElute PCR Purification Kit (Qiagen). DNA concentration in each sample was calculated with Invitrogen™ Qubit™ 4 Fluorometer using The Qubit 1X dsDNA HS Assay Kit.

ATAC-seq mapping and peak (APRE) calling

For peak (APRE) calling, we followed the pipeline outlined in https://github.com/alexgilgal/Thesis_methods/tree/main/ATAC-seq%20analysis, implementing minor modifications as detailed in our project's GitHub repository (https://github.com/mayflylab/Cdip-RegEmb/tree/main). We employed the ATAC_pipe.pl script for mapping reads, utilizing Bowtie2 (Langmead and Salzberg 2012) to align the reads to the CLODIP2 reference genome (GCA_902829235.1). Following alignment, the resulting BAM files were filtered based on a quality threshold of 10 and a minimum fragment length of 130 bp (Fig. S1).

The processed files were then subjected to peak analysis using the idr_ATAC_script.sh script, which executes peak calling with MACS2 (Zhang et al. 2008) to generate two sets of peaks: conservative peaks, indicating high-confidence regions across biological replicates, and optimal peaks, denoting reproducible events that consider read sampling variability, derived from pseudo-replicates. Subsequent IDR (Li et al. 2011) analysis was performed on both peak sets. Peak statistics—including the number of peaks and rescue ratios—were calculated and documented in a summary file.

APRE classification and gene assignment

APREs are classified and associated with genes based on their proximity to the transcription start sites (TSS). TSSs are defined using the get_TSS.py script (https://github.com/m-rossello/GeneRegLocator/). To classify APREs and link them to genes, we use a custom-made script named make_table_from_zones.py (https://github.com/m-rossello/GeneRegLocator/). This script is designed to delineate regulatory zones around TSSs and associate these zones with APREs from ATAC-seq data. It defines three types of regulatory zones: Promoters, located near the TSS, spanning 1000 bases upstream and 500 bases downstream. Proximal regions, positioned further from the TSS, extending 4000 bases upstream but not overlapping with promoters. Gene bodies, encompassing regions within the gene but excluding the promoter areas (Fig. 1I). Zones are non-overlapping on the same strand, although the same genomic position can exhibit different zones on each strand. Each APRE is associated with one or more genes if it overlaps by more than 70% with a gene zone.

APREs not falling within promoter, proximal, or gene body regions are classified as distal and remain unassociated with any gene (Table S2).

Open chromatin analysis

The counts obtained from consensus APREs were used for all subsequent analyses, following normalization. This normalization involves adjusting the count data using the TMM method (Robinson and Oshlack 2010) to account for differences in library size and composition. To further explore trends and variations across different biological stages, we aggregate the normalized counts by these stages, calculating mean values. Global APRE analysis categorizes APRE as either "open" or "closed" based on a threshold of 10 counts. APREs registering fewer than 10 counts are deemed closed. We define non-dynamic APREs as those that remain consistently open or closed across all examined stages. Conversely, dynamic APREs are characterized by their variability, changing between open and closed states across different stages or samples.

Differential Chromatin Accessibility Analysis

After counts normalization by TMM (Robinson and Oshlack 2010) and sample exploration by PCA and clustering, differential chromatin accessibility analysis was performed in dynamic APREs. For this analysis, we utilize the limma-trend method (Law et al. 2014; Phipson et al. 2016). This approach is applied to the normalized count data of dynamic APREs to identify significant differences in chromatin accessibility between conditions. To visualize the results, we generate volcano plots using the EnhancedVolcano package (Blighe, Rana, and Lewis 2023).

Mfuzz analysis

We conducted the Mfuzz cluster analysis using the mfuzz function from the R package Mfuzz (Kumar and Futschik 2007). Dynamic APREs with mean values computed by developmental stage were analysed. The optimal parameters were systematically determined, setting the fuzzifier value at m=1.5 and the number of clusters at 30.

k-means analysis

We performed k-means hard clustering using the DeepTools package (Ramírez et al. 2016) to analyze dynamic APRE enrichments from ATAC-seq data. This analysis included computing genome region scores with the computeMatrix function. The generated matrices enabled the visualization of heatmaps, which provided insights into the distribution of dynamic APREs across various developmental stages. Additionally, we created profile plots to further explore the chromatin accessibility dynamics.

Gene Ontology Enrichment Analysis

For our Gene Ontology (GO) enrichment analysis, GO annotations are transferred from the UniProt proteome (UP000494165) to the genes associated with the APREs. We perform statistical analysis using the topGO package (Alexa and Rahnenfuhrer 2023), which utilizes the elimination algorithm to identify significantly enriched GO terms within gene clusters. The BH false discovery rate correction method (Benjamini and Hochberg 1995) is employed to control multiple testing. The results are visualized using GO enrichment plots, created with the ggplot2 package.

TFBM enrichment analysis

Transcription factor binding motif (TFBM) enrichment analysis is conducted using the findMotifsGenome.pl tool from the HOMER suite (Heinz et al. 2010). We designate the APREs of interest as the foreground and utilize the remainder of the consensus APREs as the background. The fragment size selected for motif discovery corresponds precisely to the regions of the APREs (-size given). This analysis includes a comparison against collected motifs specific to insects.

Acknowledgements

We thank María José López Fernández, Juan J. Tena, Alejandro Gil and Alberto Pérez-Posada for advice on bioinformatics analyses and Ignacio Maeso and the rest of the members of the Almudi lab for fruitful discussions. This research was funded by the European Research Council (ERC) under the European Union's Horizon Europe research and innovation programme (ERC-CoG2021-101043751 to I.A), by the Spanish Ministry of Science and Innovation (PID2020-116041GB-I00 to IA) and by Ministry of Science, Innovation and Universities and by "European Union NextGeneration EU/PRTR" (CNS2023-145403). M. R. holds a Margarita Salas fellowship, from Spanish Ministry of Science and Innovation and I.A. is the recipient of the "Subprograma Estatal de Incorporación 2017-2020: Beatriz Galindo" (BG20-- 00115) from the Spanish Ministry of Universities.

Competing interests

The authors declare no competing or financial interests.

Author contributions

J.P., L.O-F and M.R. performed most analyses and generated most figures and tables. T.S and A.R performed spatial embryo fixations and HCR assays. L.O-F performed ATAC-seq experiments and generated the libraries. J.F.A and J.P. set up UCSC browser.

I.A. and M.R. coordinated the project. I.A. obtained funding and wrote the main text with inputs from all authors.

Data availability

The datasets generated and analysed during the current study are available in the Sequence Read Archive (SRA) repository. The ATAC-seq dataset, which includes samples from six developmental stages, can be accessed via SRA accession numbers SAMN40277609 through SAMN40277620. Additionally, RNA-seq data for three developmental stages are stored under SRA accession numbers SAMN40990690, SAMN40990692, and SAMN40990693. All datasets are grouped under the project accession number PRJNA1084266. Datasets are also available as a UCSC track hub: https://genome-euro.ucsc.edu/s/mayfly/Clodip. All code used for analysis is available on GitHub at https://github.com/mayflylab/Cdip-RegEmb/

Supplementary information

Figure S1. Read size ATAC-seq libraries
Figure S2. Chromatin state across different developmental stages.
Figure S3. Distribution of APREs per developmental timepoint
Figure S4. PCA analysis ATAC libraries and differential APRE accessibility
Figure S5. Mfuzz clusters
Figure S6. PCA analysis RNA-seq libraries

Table S1. Counts to all the identified consensus APREs before and after normalization and mean aggregation per stage.
Table S2. APREs associated to each gene and genomic zone.
Table S3. Differential APRE activity data
Table S4. Mfuzz output values.
Table S5. Gene ontology enrichment for each Mfuzz cluster.
Table S6. APREs in each cluster in the kmeans clustering.
Table S7. Homer motive enrichment results for each kmeans cluster.
Table S8. RNA counts before and after normalization.
Table S9. Homer motive enrichment results for each WGCNA module.
Table S10. List of primers used for library preparation.

References

Adams, Mark D., Susan E. Celniker, Robert A. Holt, Cheryl A. Evans, Jeannine D. Gocayne, Peter G. Amanatides, Steven E. Scherer, et al. 2000. 'The Genome Sequence of *Drosophila Melanogaster*'. *Science* 287 (5461): 2185–95. https://doi.org/10.1126/science.287.5461.2185.

Alexa, Adrian, and Jorg Rahnenfuhrer. 2023. 'topGO: Enrichment Analysis for Gene Ontology'. [object Object]. https://doi.org/10.18129/B9.BIOC.TOPGO.

Almudi, Isabel, Carlos A. Martín-Blanco, Isabel M. García-Fernandez, Adrián López-Catalina, Kristofer Davie, Stein Aerts, and Fernando Casares. 2019. 'Establishment of the Mayfly Cloeon Dipterum as a New Model System to Investigate Insect Evolution'. *EvoDevo* 10 (1): 6. https://doi.org/10.1186/s13227-019-0120-y.

Almudi, Isabel, Joel Vizueta, Christopher D. R. Wyatt, Alex de Mendoza, Ferdinand Marlétaz, Panos N. Firbas, Roberto Feuda, et al. 2020. 'Genomic Adaptations to Aquatic and Aerial Life in Mayflies and the Origin of Insect Wings'. *Nature Communications* 11 (1): 2631. https://doi.org/10.1038/s41467-020-16284-8.

Andrikou, Carmen, and Maria Ina Arnone. 2015. 'Too Many Ways to Make a Muscle: Evolution of GRNs Governing Myogenesis'. *Zoologischer Anzeiger - A Journal of Comparative Zoology* 256 (May): 2–13. https://doi.org/10.1016/j.jcz.2015.03.005.

Benjamini, Yoav, and Yosef Hochberg. 1995. 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing'. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 57 (1): 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Blighe, Kevin, S. Rana, and M. Lewis. 2023. 'EnhancedVolcano: Publication-Ready Volcano Plots with Enhanced Colouring and Labeling.', 2023. https://doi.org/10.18129/B9.BIOC.ENHANCEDVOLCANO.

Bozek, Marta, Roberto Cortini, Andrea Ennio Storti, Ulrich Unnerstall, Ulrike Gaul, and Nicolas Gompel. 2019. 'ATAC-Seq Reveals Regional Differences in Enhancer Accessibility during the Establishment of Spatial Coordinates in the *Drosophila* Blastoderm'. *Genome Research* 29 (5): 771–83. https://doi.org/10.1101/gr.242362.118.

Brantley, Susanna E., and Stefano Di Talia. 2021. 'Cell Cycle Control during Early Embryogenesis'. *Development* 148 (13): dev193128. https://doi.org/10.1242/dev.193128.

569 Brennan, Kaelan J., Melanie Weilert, Sabrina Krueger, Anusri Pampari, Hsiao-yun Liu,
570 Ally W.H. Yang, Jason A. Morrison, et al. 2023. 'Chromatin Accessibility in the Drosophila
571 Embryo Is Determined by Transcription Factor Pioneering and Enhancer Activation'.
572 *Developmental Cell* 58 (19): 1898-1916.e9.
573 https://doi.org/10.1016/j.devcel.2023.07.007.

574 Bruce, Heather S, Gabby Jerz, Sophia R Kelly, Jenny McCarthy, Aaron Pomerantz,
575 Gayani Senevirathne, Alice Sherrard, Dennis A Sun, Carsten Wolff, and Nipam H. Patel.
576 2021. 'Hybridization Chain Reaction (HCR) In Situ Protocol'. *Protocols.Io*, July.

577 Buenrostro, Jason D, Paul G Giresi, Lisa C Zaba, Howard Y Chang, and William J
578 Greenleaf. 2013. 'Transposition of Native Chromatin for Fast and Sensitive Epigenomic
579 Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position'. *Nature*
580 *Methods* 10 (12): 1213–18. https://doi.org/10.1038/nmeth.2688.

581 Buenrostro, Jason D., Beijing Wu, Howard Y. Chang, and William J. Greenleaf. 2015.
582 'ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide'. *Current*
583 *Protocols in Molecular Biology* 109 (1).
584 https://doi.org/10.1002/0471142727.mb2129s109.

585 Carroll, S. B. 1998. 'From Pattern to Gene, from Gene to Pattern'. *The International*
586 *Journal of Developmental Biology* 42 (3): 305–9.

587 Castanon, Irinka, Stephen Von Stetina, Jason Kass, and Mary K. Baylies. 2001.
588 'Dimerization Partners Determine the Activity of the Twist bHLH Protein during
589 *Drosophila* Mesoderm Development'. *Development* 128 (16): 3145–59.
590 https://doi.org/10.1242/dev.128.16.3145.

591 Clark, Ivan B.N., Joanna Boyd, Graham Hamilton, David J. Finnegan, and Andrew P.
592 Jarman. 2006. 'D-Six4 Plays a Key Role in Patterning Cell Identities Deriving from the
593 Drosophila Mesoderm'. *Developmental Biology* 294 (1): 220–31.
594 https://doi.org/10.1016/j.ydbio.2006.02.044.

595 Cripps, Richard M., and Eric N. Olson. 2002. 'Control of Cardiac Development by an
596 Evolutionarily Conserved Transcriptional Network'. *Developmental Biology* 246 (1): 14–
597 28. https://doi.org/10.1006/dbio.2002.0666.

598 Crowley, Liam, Heather Allen, Ian Barnes, Douglas Boyes, Gavin R. Broad, Christopher
599 Fletcher, Peter W.H. Holland, et al. 2023. 'A Sampling Strategy for Genome Sequencing
600 the British Terrestrial Arthropod Fauna'. *Wellcome Open Research* 8 (March): 123.
601 https://doi.org/10.12688/wellcomeopenres.18925.1.

602 D. Grimaldi & M. S. Engel. 2007. 'D. Grimaldi & M. S. Engel 2005. Evolution of the
603 Insects. Xv + 755 Pp. Cambridge, New York, Melbourne: Cambridge University Press.
604 Price £45.00, US $75.00 (Hard Covers). ISBN 0 521 82149 5.' *Geological Magazine* 144
605 (6): 1035–36. https://doi.org/10.1017/S001675680700372X.

606 Davis, Gregory K., and Nipam H. Patel. 2002. 'Short, Long, and Beyond: Molecular and
607 Embryological Approaches to Insect Segmentation'. *Annual Review of Entomology* 47
608 (1): 669–99. https://doi.org/10.1146/annurev.ento.47.091201.145251.

609 Fernández, Rosa, Marina Marcet-Houben, Fabrice Legeai, Gautier Richard, Stéphanie
610 Robin, Valentin Wucher, Cinta Pegueroles, Toni Gabaldón, and Denis Tagu. 2020.
611 'Selection Following Gene Duplication Shapes Recent Genome Evolution in the Pea
612 Aphid Acyrthosiphon Pisum'. Edited by Rebekah Rogers. *Molecular Biology and*
613 *Evolution* 37 (9): 2601–15. https://doi.org/10.1093/molbev/msaa110.

614 Formenti, Giulio, Kathrin Theissinger, Carlos Fernandes, Iliana Bista, Aureliano
615 Bombarely, Christoph Bleidorn, Claudio Ciofi, et al. 2022. 'The Era of Reference
616 Genomes in Conservation Genomics'. *Trends in Ecology & Evolution* 37 (3): 197–202.
617 https://doi.org/10.1016/j.tree.2021.11.008.

618 Furlong, Eileen E. M., and Michael Levine. 2018. 'Developmental Enhancers and
619 Chromosome Topology'. *Science* 361 (6409): 1341–45.
620 https://doi.org/10.1126/science.aau0320.

621 Galant, Ron, and Sean B. Carroll. 2002. 'Evolution of a Transcriptional Repression
622 Domain in an Insect Hox Protein'. *Nature* 415 (6874): 910–13.
623 https://doi.org/10.1038/nature717.

624 Gillott, Cedric. 2005. 'Embryonic Development'. In *Entomology*, 597–622.
625 Berlin/Heidelberg: Springer-Verlag. https://doi.org/10.1007/1-4020-3183-1_20.

626 Gompel, Nicolas, Benjamin Prud'homme, Patricia J. Wittkopp, Victoria A. Kassner, and
627 Sean B. Carroll. 2005. 'Chance Caught on the Wing: Cis-Regulatory Evolution and the
628 Origin of Pigment Patterns in Drosophila'. *Nature* 433 (7025): 481–87.
629 https://doi.org/10.1038/nature03235.

630 Grueber, Wesley B, and Yuh Nung Jan. 2004. 'Dendritic Development: Lessons from
631 Drosophila and Related Branches'. *Current Opinion in Neurobiology* 14 (1): 74–82.
632 https://doi.org/10.1016/j.conb.2004.01.001.

633 Guerreiro, Isabel, Andreia Nunes, Joost M. Woltering, Ana Casaca, Ana Nóvoa, Tânia
634 Vinagre, Margaret E. Hunter, Denis Duboule, and Moisés Mallo. 2013. 'Role of a
635 Polymorphism in a Hox/Pax-Responsive Enhancer in the Evolution of the Vertebrate
636 Spine'. *Proceedings of the National Academy of Sciences* 110 (26): 10682–86.
637 https://doi.org/10.1073/pnas.1300592110.

638 Heinz, Sven, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C. Lin, Peter
639 Laslo, Jason X. Cheng, Cornelis Murre, Harinder Singh, and Christopher K. Glass. 2010.
640 'Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-
641 Regulatory Elements Required for Macrophage and B Cell Identities'. *Molecular Cell* 38
642 (4): 576–89. https://doi.org/10.1016/j.molcel.2010.05.004.

643 Hosoya, Toshihiko, Kazunaga Takizawa, Koushi Nitta, and Yoshiki Hotta. 1995. 'Glial
644 Cells Missing: A Binary Switch between Neuronal and Glial Determination in Drosophila'.
645 *Cell* 82 (6): 1025–36. https://doi.org/10.1016/0092-8674(95)90281-3.

646 Irimia, Manuel, Juan J. Tena, Maria S. Alexis, Ana Fernandez-Miñan, Ignacio Maeso,
647 Ozren Bogdanović, Elisa De La Calle-Mustienes, Scott W. Roy, José L. Gómez-
648 Skarmeta, and Hunter B. Fraser. 2012. 'Extensive Conservation of Ancient Microsynteny
649 across Metazoans Due to *Cis* -Regulatory Constraints'. *Genome Research* 22 (12):
650 2356–67. https://doi.org/10.1101/gr.139725.112.

651 Kuehn, Emily, David S. Clausen, Ryan W. Null, Bria M. Metzger, Amy D. Willis, and B.
652 Duygu Özpolat. 2022. 'Segment Number Threshold Determines Juvenile Onset of
653 Germline Cluster Expansion in *Platynereis Dumerilii*'. *Journal of Experimental Zoology*
654 *Part B: Molecular and Developmental Evolution* 338 (4): 225–40.
655 https://doi.org/10.1002/jez.b.23100.

656 Kumar, Lokesh, and Matthias E. Futschik. 2007. 'Mfuzz: A Software Package for Soft
657 Clustering of Microarray Data'. *Bioinformation* 2 (1): 5–7.
658 https://doi.org/10.6026/97320630002005.

659 Langmead, Ben, and Steven L Salzberg. 2012. 'Fast Gapped-Read Alignment with
660 Bowtie 2'. *Nature Methods* 9 (4): 357–59. https://doi.org/10.1038/nmeth.1923.

661 Law, Charity W, Yunshun Chen, Wei Shi, and Gordon K Smyth. 2014. 'Voom: Precision
662 Weights Unlock Linear Model Analysis Tools for RNA-Seq Read Counts'. *Genome*
663 *Biology* 15 (2): R29. https://doi.org/10.1186/gb-2014-15-2-r29.

664 Levin, Michal, Leon Anavy, Alison G. Cole, Eitan Winter, Natalia Mostov, Sally Khair,
665 Naftalie Senderovich, et al. 2016. 'The Mid-Developmental Transition and the Evolution
666 of Animal Body Plans'. *Nature* 531 (7596): 637–41. https://doi.org/10.1038/nature16994.

667 Leyhr, Jake, Laura Waldmann, Beata Filipek-Górniok, Hanqing Zhang, Amin Allalou, and
668 Tatjana Haitina. 2022. 'A Novel Cis-Regulatory Element Drives Early Expression of
669 Nkx3.2 in the Gnathostome Primary Jaw Joint'. *eLife* 11 (November): e75749.
670 https://doi.org/10.7554/eLife.75749.

671 Li, Qunhua, James B. Brown, Haiyan Huang, and Peter J. Bickel. 2011. 'Measuring
672 Reproducibility of High-Throughput Experiments'. *The Annals of Applied Statistics* 5 (3).
673 https://doi.org/10.1214/11-AOAS466.

674 Liang, Hsiao-Lan, Chung-Yi Nien, Hsiao-Yun Liu, Mark M. Metzstein, Nikolai Kirov, and
675 Christine Rushlow. 2008. 'The Zinc-Finger Protein Zelda Is a Key Activator of the Early
676 Zygotic Genome in Drosophila'. *Nature* 456 (7220): 400–403.
677 https://doi.org/10.1038/nature07388.

678 Lowe, Robert, Marek Wojciechowski, Nancy Ellis, and Paul J Hurd. 2022. 'Chromatin
679 Accessibility-Based Characterisation of Brain Gene Regulatory Networks in Three
680 Distinct Honey Bee Polyphenisms'. *Nucleic Acids Research* 50 (20): 11550–62.
681 https://doi.org/10.1093/nar/gkac992.

682 Mantica, Federica, Luis P. Iñiguez, Yamile Marquez, Jon Permanyer, Antonio Torres-
683 Mendez, Josefa Cruz, Xavier Franch-Marro, et al. 2024. 'Evolution of Tissue-Specific
684 Expression of Ancestral Genes across Vertebrates and Insects'. *Nature Ecology &*
685 *Evolution*, April. https://doi.org/10.1038/s41559-024-02398-5.

686 Marlétaz, Ferdinand, Panos N. Firbas, Ignacio Maeso, Juan J. Tena, Ozren Bogdanovic,
687 Malcolm Perry, Christopher D. R. Wyatt, et al. 2018. 'Amphioxus Functional Genomics
688 and the Origins of Vertebrate Gene Regulation'. *Nature* 564 (7734): 64–70.
689 https://doi.org/10.1038/s41586-018-0734-6.

690 Misof, Bernhard, Shanlin Liu, Karen Meusemann, Ralph S. Peters, Alexander Donath,
691 Christoph Mayer, Paul B. Frandsen, et al. 2014. 'Phylogenomics Resolves the Timing
692 and Pattern of Insect Evolution'. *Science* 346 (6210): 763–67.
693 https://doi.org/10.1126/science.1257570.

694 Molik, David C. 2022. 'An Outsider's Perspective on Why We Climb Mountains and Why
695 Projects Like the I5k Matter'. Edited by Phyllis Weintraub. *Journal of Insect Science* 22
696 (4): 2. https://doi.org/10.1093/jisesa/ieac038.

697 Molina-Gil, Sara, Sol Sotillos, José Manuel Espinosa-Vázquez, Isabel Almudi, and
698 James C.-G. Hombría. 2023. 'Interlocking of Co-Opted Developmental Gene Networks
699 in Drosophila and the Evolution of Pre-Adaptive Novelty'. *Nature Communications* 14
700 (1): 5730. https://doi.org/10.1038/s41467-023-41414-3.

701 Münster, Stefan, Akanksha Jain, Alexander Mietke, Anastasios Pavlopoulos, Stephan
702 W. Grill, and Pavel Tomancak. 2019. 'Attachment of the Blastoderm to the Vitelline
703 Envelope Affects Gastrulation of Insects'. *Nature* 568 (7752): 395–99.
704 https://doi.org/10.1038/s41586-019-1044-3.

705 Patel, Nipam H. 1994. 'Developmental Evolution: Insights from Studies of Insect
706 Segmentation'. *Science* 266 (5185): 581–90. https://doi.org/10.1126/science.7939712.

707 Phipson, Belinda, Stanley Lee, Ian J. Majewski, Warren S. Alexander, and Gordon K.
708 Smyth. 2016. 'Robust Hyperparameter Estimation Protects against Hypervariable Genes
709 and Improves Power to Detect Differential Expression'. *The Annals of Applied Statistics*
710 10 (2). https://doi.org/10.1214/16-AOAS920.

711 Qian, S., M. Capovilla, and V. Pirrotta. 1991. 'The Bx Region Enhancer, a Distant Cis-
712 Control Element of the Drosophila Ubx Gene and Its Regulation by Hunchback and Other
713 Segmentation Genes.' *The EMBO Journal* 10 (6): 1415–25.
714 https://doi.org/10.1002/j.1460-2075.1991.tb07662.x.

715 Ramírez, Fidel, Devon P Ryan, Björn Grüning, Vivek Bhardwaj, Fabian Kilpert, Andreas
716 S Richter, Steffen Heyne, Friederike Dündar, and Thomas Manke. 2016. 'deepTools2: A
717 next Generation Web Server for Deep-Sequencing Data Analysis'. *Nucleic Acids*
718 *Research* 44 (W1): W160–65. https://doi.org/10.1093/nar/gkw257.

719 Richard, Gautier, Fabrice Legeai, Nathalie Prunier-Leterme, Anthony Bretaudeau, Denis
720 Tagu, Julie Jaquiéry, and Gaël Le Trionnaire. 2017. 'Dosage Compensation and Sex-
721 Specific Epigenetic Landscape of the X Chromosome in the Pea Aphid'. *Epigenetics &*
722 *Chromatin* 10 (1): 30. https://doi.org/10.1186/s13072-017-0137-1.

723 Robinow, Steven, and Kalpana White. 1988. 'The Locus Elav of Drosophila
724 Melanogaster Is Expressed in Neurons at All Developmental Stages'. *Developmental*
725 *Biology* 126 (2): 294–303. https://doi.org/10.1016/0012-1606(88)90139-X.

726 Robinson, Mark D, and Alicia Oshlack. 2010. 'A Scaling Normalization Method for
727 Differential Expression Analysis of RNA-Seq Data'. *Genome Biology* 11 (3): R25.
728 https://doi.org/10.1186/gb-2010-11-3-r25.

729 Rodríguez-Montes, Leticia, Svetlana Ovchinnikova, Xuefei Yuan, Tania Studer, Ioannis
730 Sarropoulos, Simon Anders, Henrik Kaessmann, and Margarida Cardoso-Moreira. 2023.
731 'Sex-Biased Gene Expression across Mammalian Organ Development and Evolution'.
732 *Science* 382 (6670): eadf1046. https://doi.org/10.1126/science.adf1046.

733 Sander, Klaus. 1976. 'Specification of the Basic Body Pattern in Insect Embryogenesis'.
734 In *Advances in Insect Physiology*, 12:125–238. Elsevier. https://doi.org/10.1016/S0065-
735 2806(08)60255-6.

736 Santos, M. Emília, Augustin Le Bouquin, Antonin J. J. Crumière, and Abderrahman Khila.
737 2017. 'Taxon-Restricted Genes at the Origin of a Novel Trait Allowing Access to a New
738 Environment'. *Science*.

739 Schindelin, Johannes, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark
740 Longair, Tobias Pietzsch, Stephan Preibisch, et al. 2012. 'Fiji: An Open-Source Platform
741 for Biological-Image Analysis'. *Nature Methods* 9 (7): 676–82.
742 https://doi.org/10.1038/nmeth.2019.

743 Schmidt-Ott, Urs, and Jeremy A Lynch. 2016. 'Emerging Developmental Genetic Model
744 Systems in Holometabolous Insects'. *Developmental Mechanisms, Patterning and*
745 *Evolution* 39 (August): 116–28. https://doi.org/10.1016/j.gde.2016.06.004.

746 Schulz, Cordula, and Diethard Tautz. 1995. 'Zygotic *Caudal* Regulation by *Hunchback*
747 and Its Role in Abdominal Segment Formation of the *Drosophila* Embryo'. *Development*
748 121 (4): 1023–28. https://doi.org/10.1242/dev.121.4.1023.

749 Sekelsky, J J, S J Newfeld, L A Raftery, E H Chartoff, and W M Gelbart. 1995. 'Genetic
750 Characterization and Cloning of Mothers against Dpp, a Gene Required for
751 Decapentaplegic Function in Drosophila Melanogaster.' *Genetics* 139 (3): 1347–58.
752 https://doi.org/10.1093/genetics/139.3.1347.

753 Shandala, Tatiana, R. Daniel Kortschak, Stephen Gregory, and Robert Saint. 1999. 'The
754 *Drosophila Dead Ringer* Gene Is Required for Early Embryonic Patterning through
755 Regulation of *Argos* and *Buttonhead* Expression'. *Development* 126 (19): 4341–49.
756 https://doi.org/10.1242/dev.126.19.4341.

757 Simon, Jeremy M, Paul G Giresi, Ian J Davis, and Jason D Lieb. 2012. 'Using
758 Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) to Isolate Active
759 Regulatory DNA'. *Nature Protocols* 7 (2): 256–67.
760 https://doi.org/10.1038/nprot.2011.444.

761 Simon, Sabrina, Alexander Blanke, and Karen Meusemann. 2018. 'Reanalyzing the
762 Palaeoptera Problem – The Origin of Insect Flight Remains Obscure'. *Wings and*
763 *Powered Flight: Core Novelties in Insect Evolution* 47 (4): 328–38.
764 https://doi.org/10.1016/j.asd.2018.05.002.

765 Starks, Rebekah R., Anilisa Biswas, Ashish Jain, and Geetu Tuteja. 2019. 'Combined
766 Analysis of Dissimilar Promoter Accessibility and Gene Expression Profiles Identifies
767 Tissue-Specific Genes and Actively Repressed Networks'. *Epigenetics & Chromatin* 12
768 (1): 16. https://doi.org/10.1186/s13072-019-0260-2.

769 Tojo, Koji, and Ryuichiro Machida. 1997. 'Embryogenesis of the mayflyEphemera
770 Japonica McLachlan (Insecta: Ephemeroptera, Ephemeridae), with Special Reference
771 to Abdominal Formation'. *Journal of Morphology* 234 (1): 97–107.
772 https://doi.org/10.1002/(SICI)1097-4687(199710)234:1<97::AID-JMOR9>3.0.CO;2-K.

773 Turchyn, Nataliya, John Chesebro, Steven Hrycaj, Juan P. Couso, and Aleksandar
774 Popadić. 2011. 'Evolution of Nubbin Function in Hemimetabolous and Holometabolous
775 Insect Appendages'. *Developmental Biology* 357 (1): 83–95.
776 https://doi.org/10.1016/j.ydbio.2011.06.014.

777 Wang, Mingyue, Yang Liu, Tinggang Wen, Weiwei Liu, Qionghua Gao, Jie Zhao, Zijun
778 Xiong, et al. 2020. 'Chromatin Accessibility and Transcriptome Landscapes of
779 Monomorium Pharaonis Brain'. *Scientific Data* 7 (1): 217.
780 https://doi.org/10.1038/s41597-020-0556-x.

781 Wu, Linda H., and Judith A. Lengyel. 1998. 'Role of *Caudal* in Hindgut Specification and
782 Gastrulation Suggests Homology between *Drosophila* Amnioproctodeal Invagination and
783 Vertebrate Blastopore'. *Development* 125 (13): 2433–42.
784 https://doi.org/10.1242/dev.125.13.2433.

785 Zdobnov, Evgeny M., and Peer Bork. 2007. 'Quantification of Insect Genome
786 Divergence'. *Trends in Genetics* 23 (1): 16–20. https://doi.org/10.1016/j.tig.2006.10.004.

787 Zhang, Yong, Tao Liu, Clifford A Meyer, Jérôme Eeckhoute, David S Johnson, Bradley
788 E Bernstein, Chad Nusbaum, et al. 2008. 'Model-Based Analysis of ChIP-Seq (MACS)'.
789 *Genome Biology* 9 (9): R137. https://doi.org/10.1186/gb-2008-9-9-r137.

790 Zhao, Xiaomeng, Long Su, Weilin Xu, Sarah Schaack, and Cheng Sun. 2020. 'Genome-
791 Wide Identification of Accessible Chromatin Regions in Bumblebee by ATAC-Seq'.
792 *Scientific Data* 7 (1): 367. https://doi.org/10.1038/s41597-020-00713-w.

793 Zrzavý, J., and P. Štys. 1997. 'The Basic Body Plan of Arthropods: Insights from
794 Evolutionary Morphology and Developmental Biology'. *Journal of Evolutionary Biology*
795 10 (3): 353–67. https://doi.org/10.1046/j.1420-9101.1997.10030353.x.

796

797

SUPPLEMENTARY FIGURES

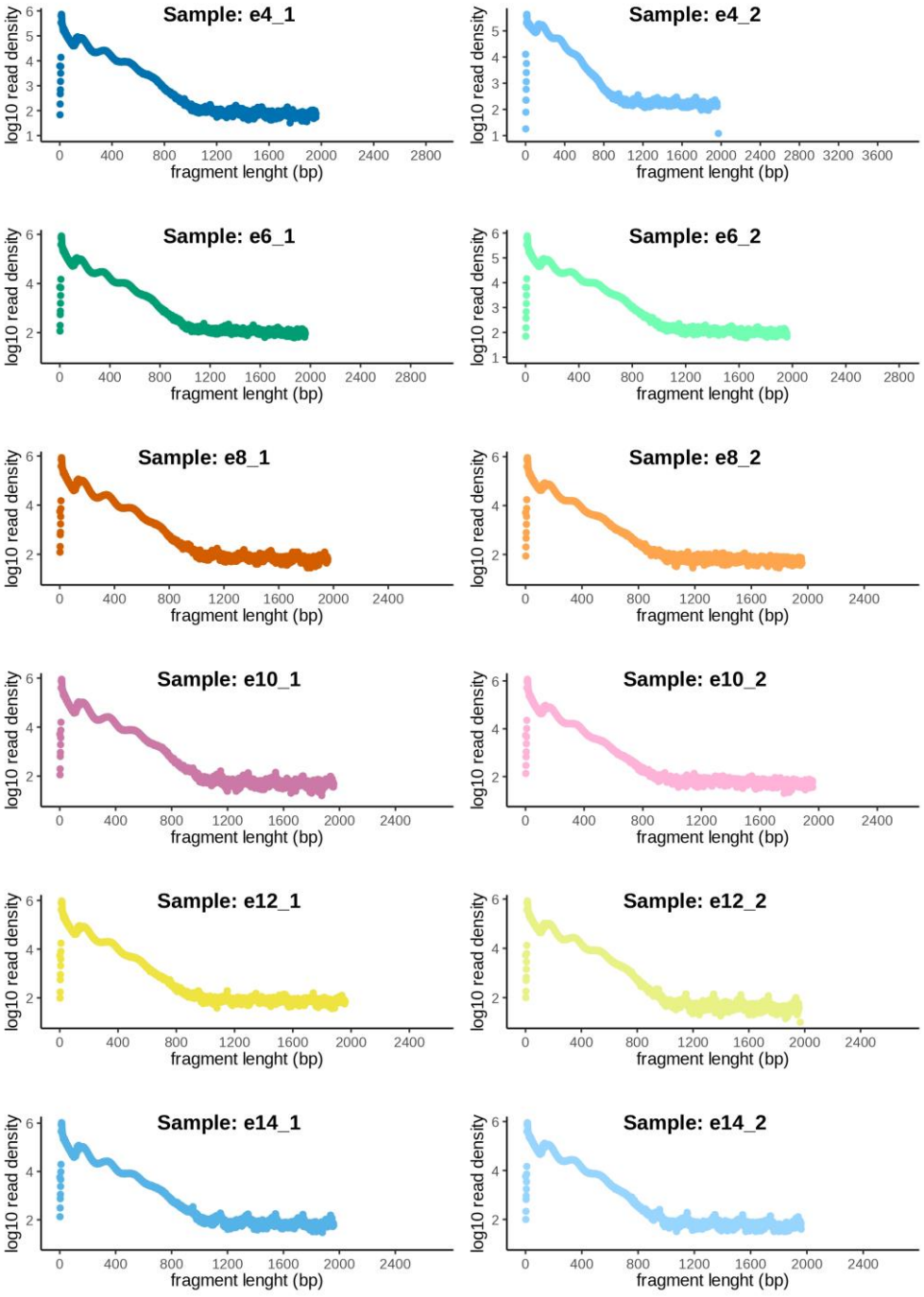


Figure S1. Read size distribution in ATAC-seq libraries.

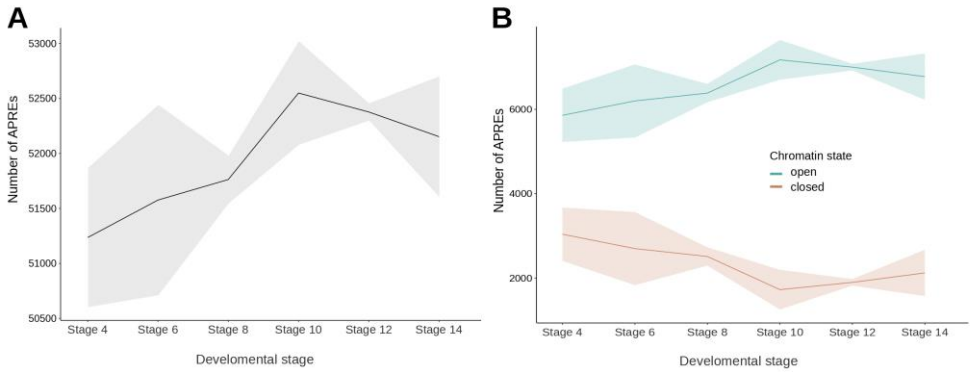


Figure S2. Chromatin state across different developmental stages. (A) Total number of APREs identified as open at each developmental stage. **(B)** Chromatin APREs that dynamically opening and closing across the stages.

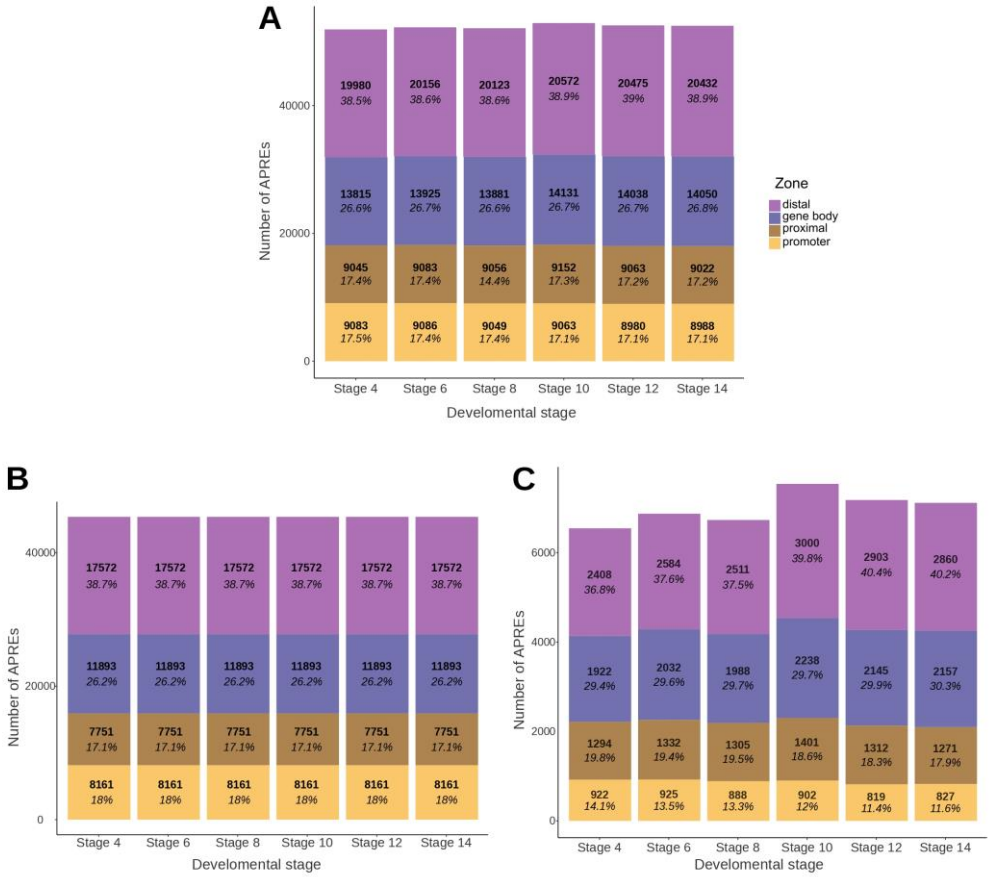


Figure S3. Distribution of APREs per developmental timepoint. (A) Number of APREs in each genomic zone distributed across each developmental stage. **(B)** Non-dynamic APREs per genomic zone distributed across each stage. **(C)** Dynamic APREs per genomic zone distributed across each stage.

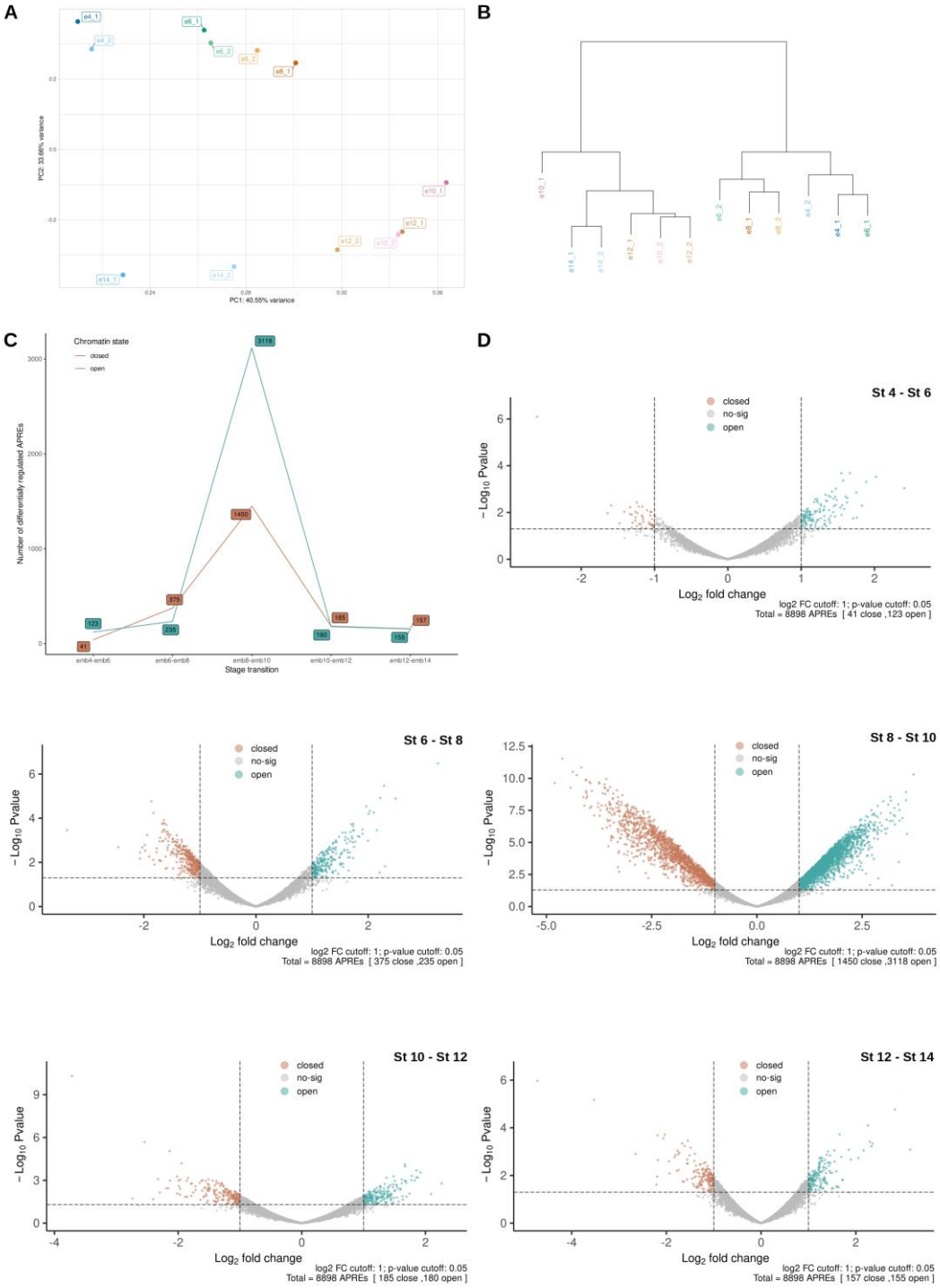


Figure S4. PCA analysis of ATAC libraries and differential APRE accessibility. (A) PCA showing the distribution of samples across the first two principal components. **(B)** Hierarchical clustering of each sample. **(C)** Number of differentially accessible chromatin regions between the different stage transitions. **(D)** Volcano plot of all differentially accessible regions for each stage transition.

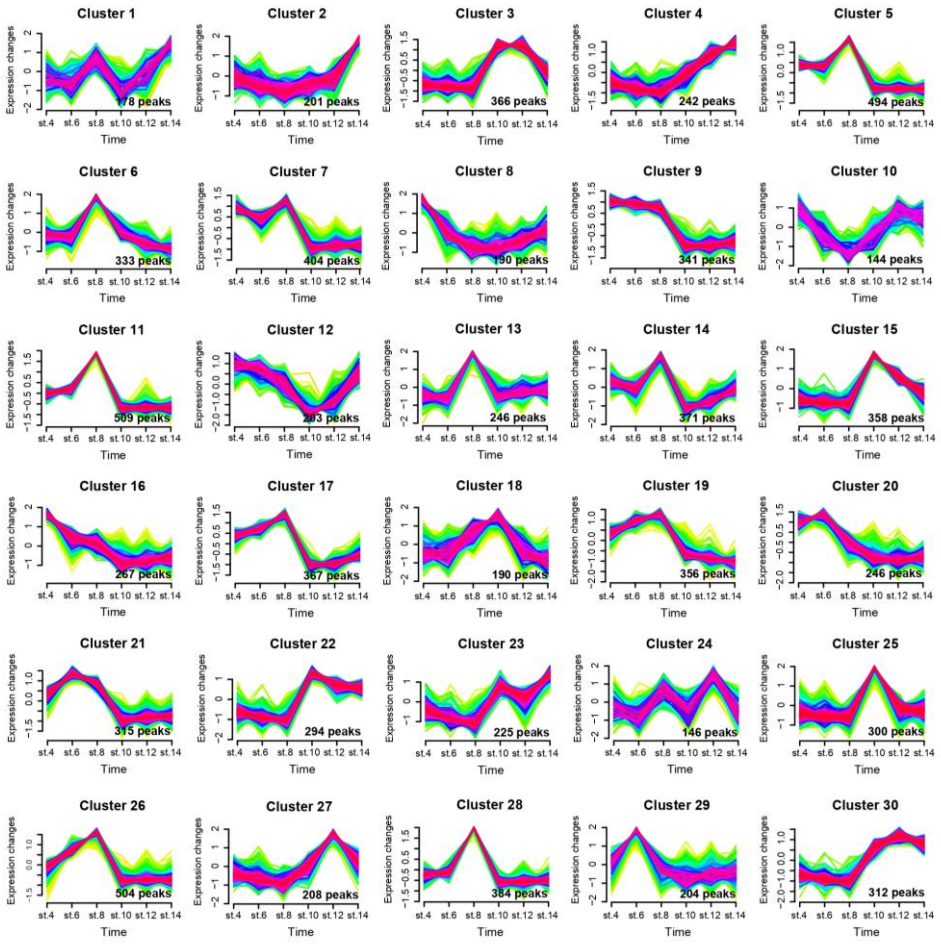


Figure S5. Mfuzz clustering. Patterns of chromatin accessibility across different developmental stages.

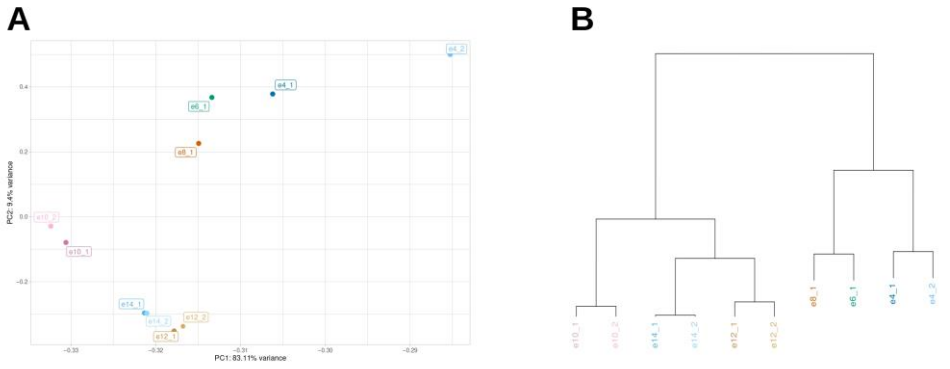
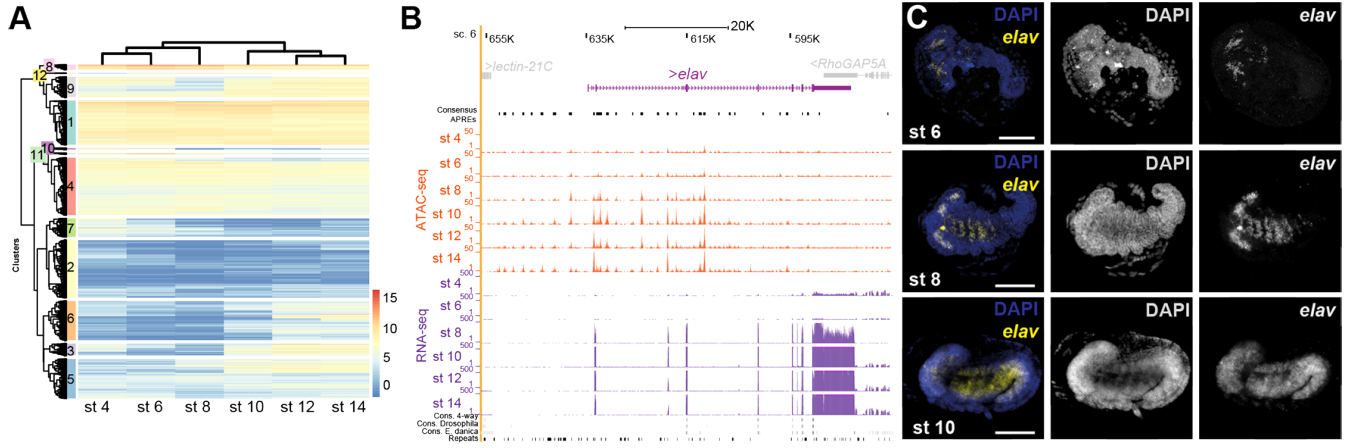


Figure S6. PCA analysis of RNA-seq Libraries. (A) PCA displaying the distribution of RNA samples across the first two principal components. **(B)** Hierarchical clustering of RNA samples.



D

Brain

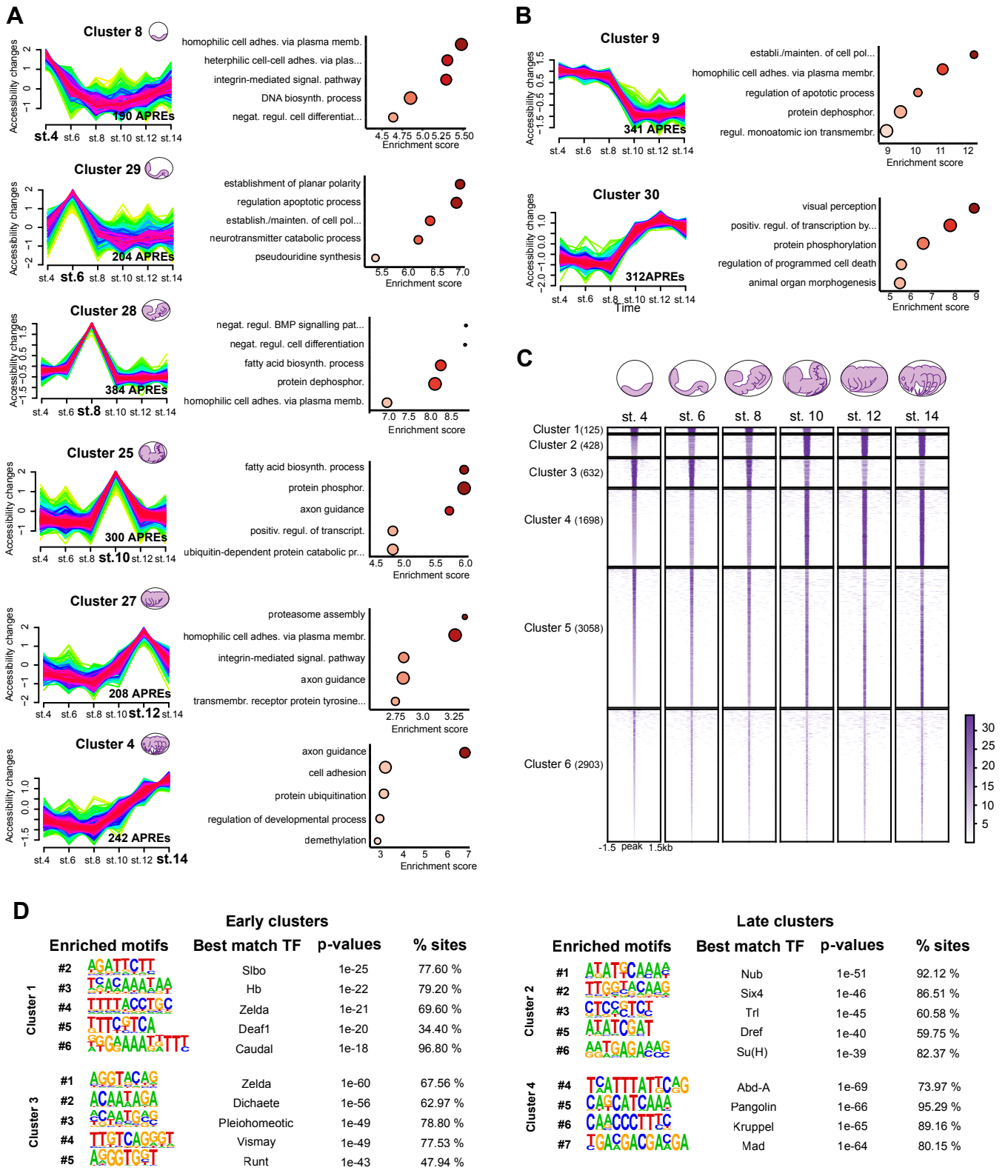
Enriched motifs	Best match TF	p-values
#2	Su(H)	1e-19
#4	dl	1e-15
#6	retn	1e-13
#8	Cut	1e-12

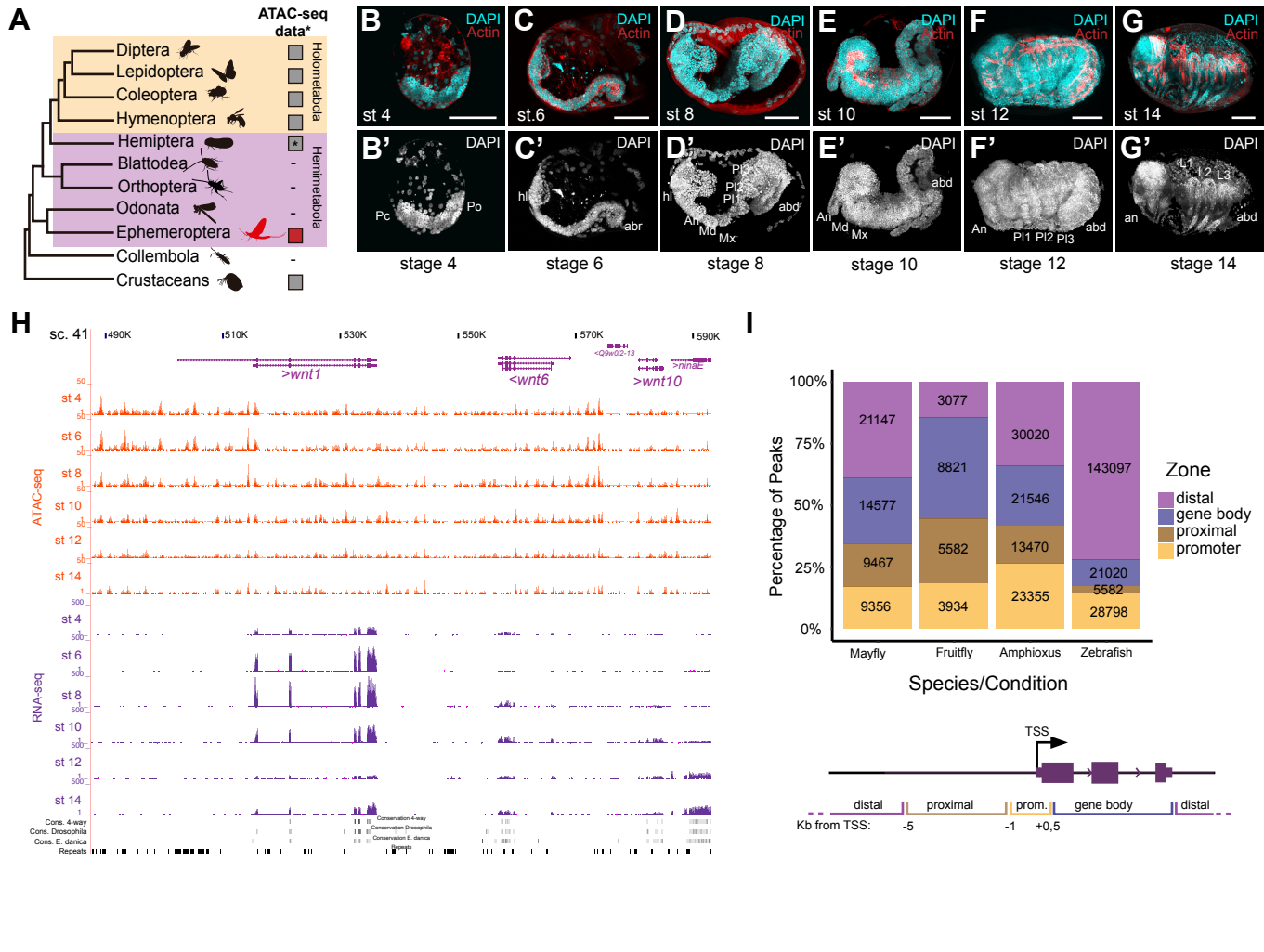
Muscle

Enriched motifs	Best match TF	p-values
#1	Hsf	1e-20
#2	bap	1e-20
#4	twi	1e-18
#5	Deaf1	1e-18

Gills

Enriched motifs	Best match TF	p-values
#1	dve	1e-15
#2	gcm	1e-15
#3	Ap-1	1e-14
#4	dl	1e-13


A
Cluster 8
190 APREs
homophilic cell adhes. via plasma memb.
heterphilic cell-cell adhes. via plas...
integrin-mediated signal. pathway
DNA biosynth. process
negat. regul. cell differentiat...
Enrichment score
Cluster 29
204 APREs
establishment of planar polarity
regulation apoptotic process
establish./mainten. of cell pol...
neurotransmitter catabolic process
pseudouridine synthesis
Enrichment score
Cluster 28
384 APREs
negat. regul. BMP signalling pat...
negat. regul. cell differentiation
fatty acid biosynth. process
protein dephosphor.
homophilic cell adhes. via plasma memb.
Enrichment score
Cluster 25
300 APREs
fatty acid biosynth. process
protein phosphor.
axon guidance
positiv. regul. of transcript.
ubiquitin-dependent protein catabolic pr...
Enrichment score
Cluster 27
208 APREs
proteasome assembly
homophilic cell adhes. via plasma membr.
integrin-mediated signal. pathway
axon guidance
transmembr. receptor protein tyrosine...
Enrichment score
Cluster 4
242 APREs
axon guidance
cell adhesion
protein ubiquitination
regulation of developmental process
demethylation
Enrichment score
Accessibility changes
st.4 st.6 st.8 st.10 st.12 st.14
B
Cluster 9
341 APREs
establi./mainten. of cell pol...
homophilic cell adhes. via plasma membr.
regulation of apototic process
protein dephosphor.
regul. monoatomic ion transmembr.
Enrichment score
Cluster 30
312APREs
visual perception
positiv. regul. of transcription by...
protein phosphorylation
regulation of programmed cell death
animal organ morphogenesis
Enrichment score
Time
C
st. 4
st. 6
st. 8
st. 10
st. 12
st. 14
Cluster 1 (125)
Cluster 2 (428)
Cluster 3 (632)
Cluster 4 (1698)
Cluster 5 (3058)
Cluster 6 (2903)
-1.5 peak 1.5kb
30
25
20
15
10
5
D
Early clusters
Late clusters
Enriched motifs
Best match TF
p-values
% sites
Cluster 1
#2 Slbo 1e-25 77.60 %
#3 Hb 1e-22 79.20 %
#4 Zelda 1e-21 69.60 %
#5 Deaf1 1e-20 34.40 %
#6 Caudal 1e-18 96.80 %
Cluster 3
#1 Zelda 1e-60 67.56 %
#2 Dichaete 1e-56 62.97 %
#3 Pleiohomeotic 1e-49 78.80 %
#4 Vismay 1e-49 77.53 %
#5 Runt 1e-43 47.94 %
Cluster 2
#1 Nub 1e-51 92.12 %
#2 Six4 1e-46 86.51 %
#3 Trl 1e-45 60.58 %
#5 Dref 1e-40 59.75 %
#6 Su(H) 1e-39 82.37 %
Cluster 4
#4 Abd-A 1e-69 73.97 %
#5 Pangolin 1e-66 95.29 %
#6 Kruppel 1e-65 89.16 %
#7 Mad 1e-64 80.15 %


A
ATAC-seq data*
Diptera
Lepidoptera
Coleoptera
Hymenoptera
Hemiptera
Blattodea
Orthoptera
Odonata
Ephemeroptera
Collembola
Crustaceans
Holometabola
Hemimetabola
B DAPI Actin st 4
C DAPI Actin st.6
D DAPI Actin st 8
E DAPI Actin st 10
F DAPI Actin st 12
G DAPI Actin st 14
B' DAPI Pc Po
C' DAPI hl abr
D' DAPI Pl1 Pl2 Pl3 hl An Md Mx abd
E' DAPI abd An Md Mx
F' DAPI An Pl1 Pl2 Pl3 abd
G' DAPI L1 L2 L3 an abd
stage 4
stage 6
stage 8
stage 10
stage 12
stage 14
H
sc. 41
490K
510K
530K
550K
570K
590K
>wnt1
<wnt6
>wnt10
>ninaE
ATAC-seq
RNA-seq
st 4
st 6
st 8
st 10
st 12
st 14
Cons. 4-way
Cons. Drosophila
Cons. E. danica
Repeats
I
Percentage of Peaks
100%
75%
50%
25%
0%
21147
14577
9467
9356
3077
8821
5582
3934
30020
21546
13470
23355
143097
21020
5582
28798
Mayfly
Fruitfly
Amphioxus
Zebrafish
Species/Condition
Zone
distal
gene body
proximal
promoter
TSS
distal
proximal
prom.
gene body
distal
Kb from TSS:
-5
-1
+0,5