

RESEARCH ARTICLE

Expanding the fly eye gene regulatory network: From *Drosophila* to the hoverfly *Episyrphus balteatus*

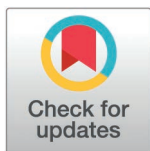
Tomás Navarro¹, Jesús Torres¹, Reinaldo Sáez-Moreno¹, Genoveva Guerrero¹, Isabel Almudí^{1a}, Antonella Iannini¹, Javier Figueras^{1ab}, Karl Wotton², Stein Aerts³, Fernando Casares^{1*}

1 CABD, CSIC/Universidad Pablo de Olavide/Junta de Andalucía, Seville, Spain, **2** Centre for Ecology and Conservation, University of Exeter, Penryn, United Kingdom, **3** VIB and Center for Human Genetics, KULeuven, Leuven, Belgium

^a Present address: Dept. Genetics, Microbiology and Statistics, School of Biology, and Institut de Recerca de La Biodiversitat (IRBio), Universitat de Barcelona. Barcelona, SPAIN.

^{ab} Present address: Department of Biosciences, Durham University, Durham, UK.

* fcasfer@upo.es



OPEN ACCESS

Citation: Navarro T, Torres J, Sáez-Moreno R, Guerrero G, Almudí I, Iannini A, et al. (2026) Expanding the fly eye gene regulatory network: From *Drosophila* to the hoverfly *Episyrphus balteatus*. PLoS Genet 22(2): e1011864. <https://doi.org/10.1371/journal.pgen.1011864>

Editor: Ville Hietakangas, University of Helsinki, Helsingin Yliopisto, FINLAND

Received: August 28, 2025

Accepted: February 17, 2026

Published: February 26, 2026

Copyright: © 2026 Navarro et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All relevant data are within the manuscript and its Supplementary Material files, which can be accessed at <https://doi.org/10.5281/zenodo.16389990>.

Funding: FC was funded by grants PID2021-122671NB-I00 and PID2024-161591NB-I00

Abstract

Comparing organ-specific gene regulatory networks (GRNs) across large evolutionary distances remains a major challenge, particularly when the species under study differ in data resolution. This applies to the GRN controlling compound eye development in insects, which is well characterized in *Drosophila melanogaster* but less well understood in other lineages. Here, we introduce the marmalade fly *Episyrphus balteatus* (Syrphidae), which diverged from *Drosophila* approximately 90 million years ago, as a comparative model to study eye development in Diptera. Using RNA-seq and ATAC-seq datasets, we reconstruct the first eye GRN in *Episyrphus*. Many genes involved in early *Drosophila* eye specification and differentiation are also active in *Episyrphus*. Among these, both species share a set of 22 transcription factors (TFs). The GRN built from these TFs and their DNA-binding motifs displays a high degree of internal connectivity. Link conservation analysis, followed by experimental testing in *Drosophila*, further identifies the AML1/Runx transcription factor *lozenge* (*lz*) as a negative regulator of the retinal determination gene *dachshund* (*dac*). Within the eye GRN, the degree of regulatory link conservation varies among genes, and both the number and position of regulatory regions often differ across orthologous loci. These results expand the eye GRN of Diptera and suggest extensive network rewiring over the evolutionary span separating *Episyrphus* and *Drosophila*.

Author summary

The gene regulatory network (GRN) controlling the specification and early patterning of the *Drosophila* eye is often used as a reference for studying insect

(by MICIU/AEI/10.13039/501100011033 and “ERDF/EU”, Spain). TN was a recipient of doctoral contract BES-2016-079458. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

compound eyes. However, *Drosophila* belongs to a derived lineage of flies, and it remains unclear which aspects of its eye GRN are broadly conserved and which are lineage-specific. To identify genes and regulatory interactions conserved beyond *Drosophila melanogaster* and its close relatives, we generated the eye GRN of *Episyrphus balteatus*, a hoverfly that diverged from *Drosophila* approximately 90 million years ago, and compared it with the *Drosophila* eye GRN using RNA-seq and ATAC-seq data. This comparison reveals a core set of conserved genes, including 22 shared transcription factors. The resulting networks show a high degree of connectivity, although the conservation of individual regulatory links varies among genes. We further tested the predictive power of conserved links by examining the relationship between the AML-family transcription factor Lozenge (*Lz*) and the retinal gene *dachshund* (*dac*), and found that *dac* is indeed regulated by *Lz*. Overall, our work expands the Dipteran eye GRN and suggests that network connectivity is evolutionarily fluid, allowing extensive rewiring over long evolutionary timescales.

Introduction

Despite their very different architectures, eye development across animals relies on a conserved genetic toolkit composed of transcription factors (TFs) and signaling molecules. This fact suggests that all animal eyes, despite having evolved multiple times independently [1], may share a deep developmental homology [2,3]. However, the development of the different eye types, being these single lens camera type or compound (comprising many individual eyes), is also remarkably different. To understand to what extent the processes governed by the set of conserved TFs and signaling pathways are conserved or divergent, two further levels of description are required. A first level is the identification of the genes lying downstream of these conserved TFs and signaling pathways. A second level is the integration of these genetic elements into a gene regulatory network (GRN)-level description. A comprehensive understanding of the genetic control of eye development with this resolution is lacking for most eye types. An exception is the eye of *Drosophila melanogaster*, a small fly whose compound eye has long served as a paradigm for this very common eye type. Developmental genetics studies have already uncovered many of the genes essential for eye development and established some of their mutual cross-regulations. This work was initially complemented with the analysis of *cis*-regulatory elements (mostly activating elements called “enhancers”) to establish the skeleton of an eye GRN (the so-called Retinal Determination Gene Regulatory Network [4]). Further research, including transcriptomics, epigenomics and genome-wide TF binding, and more recently single-cell omics, has resulted in more complete eye GRN models [5–7]. The ultimate test for these models would be the correct prediction of eye phenotypic changes when specific elements in the GRN are perturbed, something that is still pending. Globally considered, the *Drosophila* compound eye represents one of the most comprehensively studied examples of visual organ development [6,8–12].

The eye GRN is deployed in the posterior region of the eye-antennal imaginal disc, which gives rise to most structures of the adult head. The first action of the GRN is the specification of this posterior region as the eye progenitor domain. Then, signals derived from the disc's posterior margin, mostly the *hedgehog* morphogen (Hh), start the recruitment of eye progenitors into retinal precursors, which then undergo further differentiation. As these early photoreceptors also express Hh, the differentiation of the retina proceeds like a wave [13]. The differentiation wavefront is characterized by an indentation in the disc's epithelium, called the Morphogenetic Furrow (MF), which is used as a morphological reference of the differentiation process. Posterior to the MF, differentiating cells either exit the cell cycle or undergo a last round of mitosis (the so-called “second mitotic wave”) before differentiating. The products of this differentiation process are clusters of photoreceptor cells, surrounded by lens-secreting and pigment-producing cells. Each of these cartridges is a unit eye, or ommatidium, which is identifiable at the eye surface by its covering lens (or facet) [8,13].

Even within insects, the level of resolution of the eye GRN is far from that of *Drosophila* [14–16]. Work in beetles (*Tribolium*; [17,18]), honeybees (*Apis*; [19]), crickets (*Grillus*, [20]) and grasshoppers (*Schistocerca*, [21]), among others, has focused on specific elements of the GRN, finding that the genes known to be important for eye development in *Drosophila* are generally also relevant for eye development in these species, although sometimes the discovered function does not fully correspond to that previously identified in *Drosophila* (see, for example, the analysis of the role of the *eyg* TF in *Tribolium* eye development [22]). This is not totally unexpected, because *Drosophila* development is quite derived within insects. While in most insects eyes develop progressively in the cephalic region of the larval ectoderm, in flies eye development takes place inside the larva as an imaginal disc, and happens relatively rapidly, spanning the last phase of the larval life [23]. The sparsity of the sampling means that it is even unclear to what extent eye development in *Drosophila* is typical within Diptera, an order where eyes exhibit great morphological and functional diversification [24]. To tackle this issue, we have developed the first eye GRN model for the hoverfly *Episyrphus balteatus* and compared it to that of *Drosophila melanogaster*.

The evolutionary divergence between *Episyrphus* and *Drosophila* is approximately 90 million years (Myrs), placing it between the Mediterranean fruit fly, *Ceratitidis capitata* (140 Myrs), and the tsetse fly, *Glossina morsitans* (65 Myrs) [25]. *Episyrphus balteatus* (De Geer, 1776; Diptera: Syrphidae), commonly known as the marmalade hoverfly, is a cosmopolitan species widespread across the Palearctic region. It plays important ecological roles, both through its aphidophagous larval stage and as a generalist pollinator during its adult stage. Adult *Episyrphus* flies have relatively large compound eyes, a trait likely associated with their flight dexterity, that includes hovering, aerial mating, and long-distance migration [26,27]. In addition, annotated high-quality versions of its genome have been published [28,29]. The *Episyrphus* eye GRN and its comparison with that of *Drosophila* should shed light on a number of still unresolved issues. First, what is the minimal complement of genes that constitute the GRN? Second, are the transcription factors governing eye specification, growth, and early differentiation in *Drosophila* conserved between these two distantly related flies, and can this repertoire be expanded? Third, is the regulatory code controlling the expression of shared genes itself conserved—that is, to what extent has the network been rewired during evolution? Finally, are conserved regulatory links predictive of *functional* regulatory interactions, at least when tested in *Drosophila*?

Here we show, through the comparative analysis of the eye transcriptomes, that very similar functions and processes are activated during eye development in both species. The *Drosophila* and *Episyrphus* eye transcriptomes share a set of genes that can be used as a reference for the further characterization of eye development in other species. We reconstructed the eye GRN connected by linking expressed TFs to target genes through potential eye enhancers. These enhancers were identified using ATAC-seq as chromatin regions with high accessibility in eye discs but not in wing discs (which we call “eye differentially accessible regions”, or “eye DARs”). In order to validate the regulatory predictions of our GRN, we tested a predicted link between two key nodes, *lozenge* (*lz*) and *dachshund* (*dac*), using functional assays in *Drosophila*, and showed that *lz* acts as a *dac* repressor, lending support to our GRN model. The global view of the resulting GRN models identified shared regulatory links, but also links predicted to be species-specific. The core GRN

comprises 106 conserved genes, and includes a set of 22 TFs (“kernel” TFs), which according to the GRN model, are heavily interconnected. The degree to which connectivity was conserved, though, varied across genes.

Although our *Episyrphus/Drosophila* conserved GRN model is surely incomplete, among other things because we still lack DNA binding motifs for a few conserved TFs, it represents the largest expansion of the eye GRN from *Drosophila* to another dipteran beyond drosophilids. In this way, this GRN could be considered a tentative minimal eye GRN model within dipterans, to be used as a reference for comparative studies in other Dipterans and beyond.

Results

Episyrphus eye disc organization resembles that of *Drosophila*

The compound eyes of adult *Episyrphus* are larger than those of *Drosophila*. In comparison, while the *Drosophila* eye is composed of approximately 800 ommatidia, the *Episyrphus* eye contains more than 3,500 (Fig 1A and 1B). Additionally, the eyes are sexually dimorphic: in males, they extend all the way to the dorsal midline, a trait known as “holoptic eyes.” This dimorphism, however, was not analyzed further. To begin investigating *Episyrphus* eye development, we dissected L3 larvae at different time points and fixed the eye-antennal imaginal primordium (or “disc”) without discriminating the sex of the larva (Fig 1C). This disc is an epithelial sac that gives rise to most structures of the adult head [30]. In L3, it is formed by two lobes: the antennal lobe anteriorly and the eye lobe posteriorly. The eye lobe is kidney-shaped, with the concave region pointing towards the posterior pole. A furrow, similar to the morphogenetic furrow in *Drosophila*, is observed, separating anterior and posterior regions. This latter region shows clustering of cells into ommatidia. As development proceeds, the area behind the MF increases, indicating a wave-like progression of the differentiation process. In *Drosophila*, mitotic cells are detected in two regions of the disc: anterior to the MF, where proliferative progenitors lie, and in a band of mitotic cells posterior to the MF, the so-called “second mitotic wave” [31]. In *Episyrphus*, mitotic cells (detected with an antibody against phosphorylated histone H3 (PH3)) are found both, anterior and posterior to the MF. However, mitotic cells posterior to the MF are not especially concentrated, but scattered through the differentiating region. Overall, eye development in *Episyrphus* is similar to that in *Drosophila*, progressing in a wave-like fashion across the growing disc epithelium, although some aspects, such as the curvature of the MF or the pattern of proliferation behind it differ. Next, we describe the steps towards the generation of its GRN.

Identification of the nodes in the eye GRN

In a GRN model, the expression of any gene in the network is the result of the activating or repressing regulatory links connecting its *cis*-regulatory elements (CREs) to upstream TF regulators. CREs (most common CREs are positive regulators of transcription, or *enhancer* elements) can be identified as chromatin regions with increased accessibility relative to the surrounding chromatin. This can be assessed by a number of techniques [32], including Assay for Transposase-Accessible Chromatin and sequencing (ATAC-seq [33]). A link between TF-A and a gene *B* is predicted to exist if A and *B* are co-expressed and there is an “accessible region” (i.e., an ATAC-peak) in the vicinity of *B* containing a transcription factor binding site for A. Then, *B* is said to be a putative regulatory “target” of TF-A. These predicted links could be further experimentally validated by biochemical and genetic techniques. Therefore, and as a first approximation, three types of data must be gathered in order to knit a GRN model: gene transcription (expression), accessible regions (ATAC-peaks) and TF binding sites. To build a first eye GRN model for *Episyrphus*, we obtained bulk RNA-seq and ATAC-seq data for eye and wing discs from mid-third instar larvae. To derive a comparable eye GRN in *Drosophila*, equivalent sets of data were obtained in this fly (see Materials and Methods/ RNA Sequencing and Bioinformatic Processing). For all samples, a mix of male and female discs was used.

The first step in our GRN building strategy was to identify the sets of genes specifically involved in eye development. By comparing “eye-antennal” against “wing” disc transcriptomes we obtained the set of differentially expressed genes

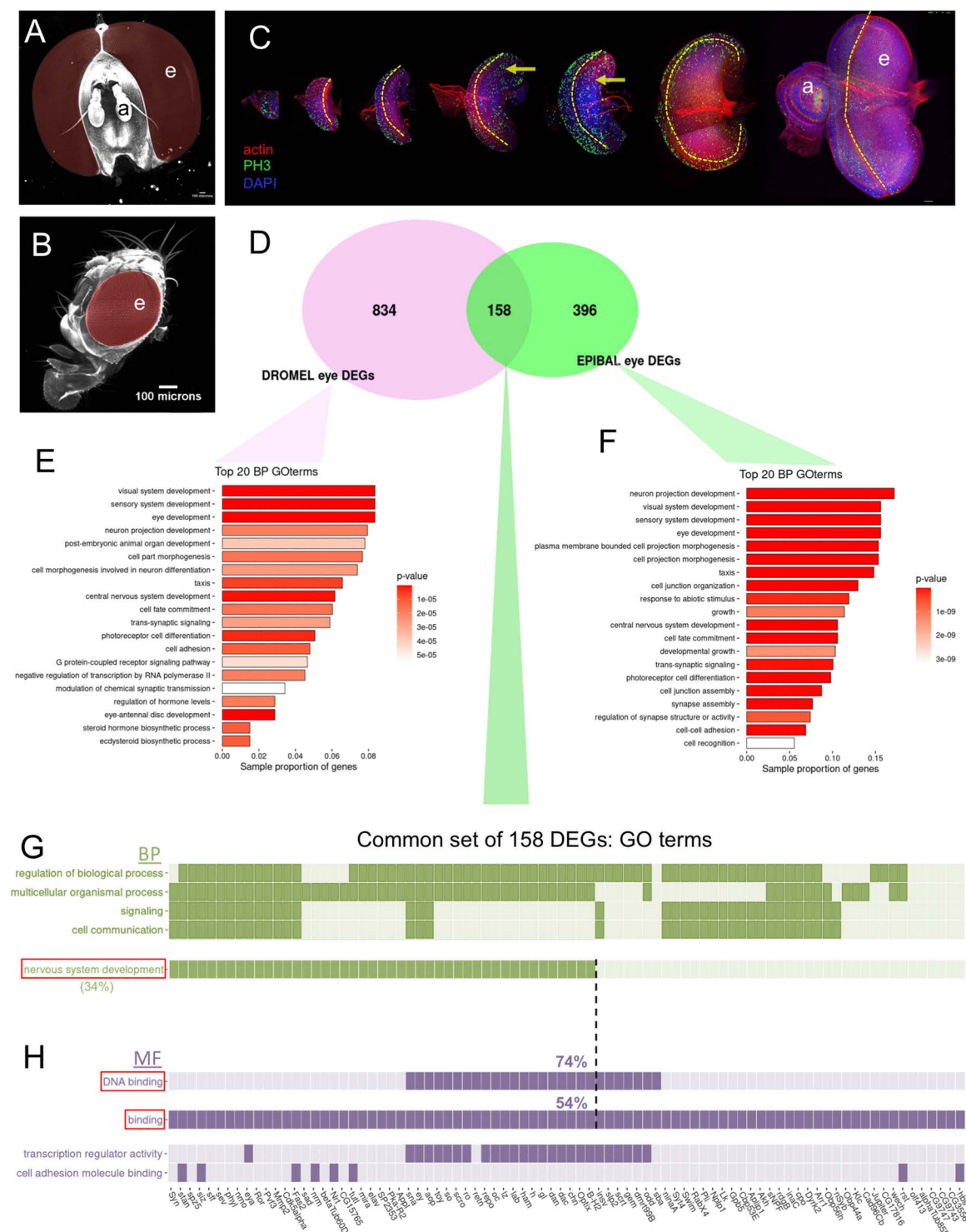


Fig 1. The *Episyrphus* and *Drosophila* eyes and their comparative transcriptomics. (A,B) Light sheet images of the adult heads of *Episyrphus balteatus* (A, male) and *Drosophila melanogaster* (B) (scale bar = 100 micrometers). Eyes are pseudocolored in red. (C) Third larval stage (L3) *Episyrphus* eye discs at progressive stages of their development, stained for actin (rhodamine-phalloidin), DNA (DAPI) and the mitotic marker phosphorylated Histone H3 (PH3). The morphogenetic furrow (MF, marked by the dashed yellow lines) is not straight, but convex. Mitotic cells are abundant anterior to the MF. In addition, frequent mitotic cells are detected scattered throughout the differentiating area of the disc, posterior to the MF (indicated by the yellow arrows). "e" and "a": eye and antenna, respectively. (D) Venn diagram indicating the number of differentially expressed genes between eye and

wing discs (“eye DEGs”) in *Drosophila* (“DROMEL”, pink) and *Episyrphus* (“EPIBAL”, green). 158 eye DEGs are shared between the two species. (E, F) Top 20 Biological Process (BP) gene ontology terms for the *Drosophila* (E) and *Episyrphus* (F) eye DEGs. (G, H) eye DEGs associated with the top five GO terms in the Biological Process (G) or Molecular Function (H), where only 4 terms were enriched among the shared 158 eye DEGs. A dark cell indicates that a gene (column) is associated with the GO term in the respective row. 34% of the genes belong to the BP term *nervous system development*. Within the set of genes associated with the MF term *binding*, the proportion of genes associated with *nervous system development* rises to 54%, further increasing to 74% when considering only those associated with *DNA-binding*.

<https://doi.org/10.1371/journal.pgen.1011864.g001>

(DEGs) in the eye-antennal imaginal disc in each species, *Episyrphus* or *Drosophila*. Although we note that antennal genes will be included in this set as well (see below), we will call this gene set “eyeDEGs” for simplicity. These first gene sets of eyeDEGs, one per species, served to check whether the eye minus wing data sets captured eye development-specific processes. *Drosophila* and *Episyrphus* eye discs have 992 and 554 eyeDEGs, respectively (Fig 1D). We attribute the difference in numbers to the fact that the annotation of *Episyrphus*’ genome is less complete than that of *Drosophila*, thus reducing the total number of genes with identifiable homologs. Gene ontology (GO) analysis of these eyeDEGs identified very similar overrepresented terms in both flies, including “visual system/sensory system/eye development” and “neuron projection development”, in agreement with the major cell populations of the eye discs at the stage studied: eye progenitors and early differentiating photoreceptor neurons (Fig 1E and 1F). The overlap between the two sets is of 158 genes (Fig 1D). This 158 gene set is further enriched (relative to the total of eye DEGs) in terms such as “signaling” and “cell communication”, and molecularly in “binding/DNA binding” (Fig 1G and 1H), which includes transcription factors. Within both sets, the enrichment of genes belonging to the *nervous system development* term is even greater (Fig 1G and 1H), indicating that a key part of the common (conserved) DEGs set is composed of genes that regulate differentiation process.

Genes with eye-specific accessible regions as nodes of the eye GRN

The involvement of a gene in the eye GRN can be inferred, even in the absence of significant differential expression, if it is associated with regulatory regions that are specifically active in the eye—that is, ATAC-peaks which are detected in eye, but not in wing discs, called eye “Differentially Accessible Regions” (eye “DARs”). One example helps clarify this point. The T-box *bifid* (*bilomb*) gene is known to be expressed in both *Drosophila* eye and wing discs and to be required for the development of some of their adult derivatives [34,35]. Although our RNA-seq analysis did not identify *bi* as an “eye DEG”, when the ATAC-seq analysis was performed, the *bi* locus showed eye and wing-specific DARs (S1 Fig). This indicated that *bi* is indeed a node controlled by two separate networks, those of the eye and the wing. Its eye and wing-specific DARs are the connections to their respective GRNs. Therefore, having an eye-specific DAR is a definitive criterion for an expressed gene to be included as node in the eye GRN. Considering this, we decided to build the eye GRN with all the genes expressed in the eye (even if not *differentially* so), which are linked to one or more eye-specific DAR. DARs were linked to target genes using a proximity criterion to the nearest expressed gene (see Materials and Methods/DARs annotation). We called these genes “eye GDARs” (eye-expressed genes with linked eye-specific DARs) and became the nodes of our network models.

We detected 533 eye GDARs in *Episyrphus* which were GO-characterized, similarly to the eye DEGs, by terms such as “eye/visual/sensory system development”, “neuron projection development” and “photoreceptor cell differentiation”. In addition, this gene set is characterized by terms such as “wing disc development” that may point to those genes which, in addition to their eye expression, have known functions in wing development (as is the case for *bilomb*, described above), or “head development” in agreement with the fact that the eye disc also contributes to the development of other head structures, including the head capsule and the antenna (S2A Fig, right). The 624 *Drosophila* eye GDARs were also characterized by the same terms as the *Episyrphus* equivalent set, although some additional general terms pertaining to development, such as “epithelial tube morphogenesis” and “post-embryonic animal organ development”, appeared as

more significant (S3B Fig, right). However, and in general, the set of eye-expressed genes linked to eye-specific DARs (eye GDARs) pointed to very similar biological functions in both species. This similarity in GO-characterization extended further when only eye GDARs of eye DEGs were considered, with “visual system/sensory system/eye development” being the top terms (S2A and S2B Fig; left). This GO enrichment analysis supports the notion that GDARs can be considered the nodes of the GRN models.

A core of 106 eye genes

Of the 533 and 624 eye GDARs that we identified in *Drosophila* and *Episyrphus*, respectively, 106 genes (which we call “core” genes) are shared between the two species. This core gene set showed GO enrichment in terms such as “visual system development” or “photoreceptor cell differentiation” relative to the union of 1157 *Episyrphus* and *Drosophila* eye GDARs (Fig 2A–2C). This was remarkable taking into account that the *Episyrphus* and *Drosophila* eye GDAR sets were already significantly enriched in these terms. We checked the expression of these genes in the single cell atlas dataset generated by Bravo Gonzalez-Blas and co-workers. [5]. All genes showed expression in some of the cell types of the eye disc epithelium but for *stumps*, which showed expression in glial cells. Among the core gene set, 29 genes have been associated to one or several signaling pathways, including the Hh, wg/Wnt and TGF- β , with components of the MAPK pathway being particularly enriched (Fig 3).

Of this core set of 106 genes, 22 were TFs (Fig 2A, 2D, and 2E). These TFs, which we call “kernel” TFs, include all retinal determination gene network genes previously known to be responsible for the early specification and growth of the *Drosophila* eye primordium (*hth*, *toy*, *ey*, *Optix*, *so*, *tsh*, *tio* and *dac*—*eya* was the only gene of this early acting network not included because it is not a TF, but a transcriptional cofactor of *So*) [4,8,12]. *Slp-2* is necessary for the growth of the eye primordium [36], *gl* (*glass*), *peb*, *lilli* and *luna* are required for the differentiation of retinal cells [37–40], *B-H1* and *Iz* are required for the maintenance of the progenitor cells that feed the retinal differentiation process [41,42], and *oc*, *lab*, *Lim1* and *disco* are expressed in prospective head/antennal cells adjacent or surrounding the eye [43–46] (Fig 2E). Finally, an involvement of *ovo*, *snail* or *E5* in the development of the eye disc has not yet been reported. Of these, sc-RNAseq data confirms that *ovo* and *E5* are expressed in eye discs (S3 and S8 Figs) and [47]. None of these datasets confirmed *snail* transcription. We decided to analyze *snail* expression *in situ* by using a *sna:GFP* transgenic strain (see Materials and Methods). Eye discs showed expression in cells posterior to the MF (S3C and S3C' Fig). This expression seems specific for the eye, as antennal, leg and wing discs were negative (S3C–S3E Fig). Therefore, *Sna* was considered part of the bona fide kernel of eye TFs. Thus, the identified kernel TF set includes all the known *Drosophila* retinal determination gene network components plus several other genes known to be essential for eye development in *Drosophila*, or at least expressed in the eye primordium and awaiting functional studies. *Episyrphus* shares these genes, which would be expected to contribute functionally in the same way to eye development in this species. Next, we used STRING [48] to build a protein interaction network for the *Drosophila* kernel TF proteins, where the links are physical and functional interactions described for those *Drosophila* proteins. The resulting network showed high connectivity, confirming a strong functional relationship among kernel TFs (Fig 2D).

Our results indicated that the *Episyrphus* eye GRN shares many of the known nodes of the *Drosophila* eye GRN. Next, we aimed at finding out how the eye GRN is weaved in both species by predicting regulatory links between TFs and core eye GDARs based on TF binding motifs presence. Of the regulatory kernel of 22 TFs, only 3 lack associated motifs in our collection (those for *bi* (*omb*), *dac*, and *tio*). Therefore, the GRN built with the kernel TFs should be very complete.

DARs of eye genes are enhancers

We recruited a gene as a node to the GRN when an eye-specific DAR is linked to it, since sites of increased chromatin accessibility are good predictors of enhancers [49–52]. We confirmed that this was the case for a small sample of *Drosophila* loci encoding TFs in our kernel set for which enhancers had been determined by the Flylight project

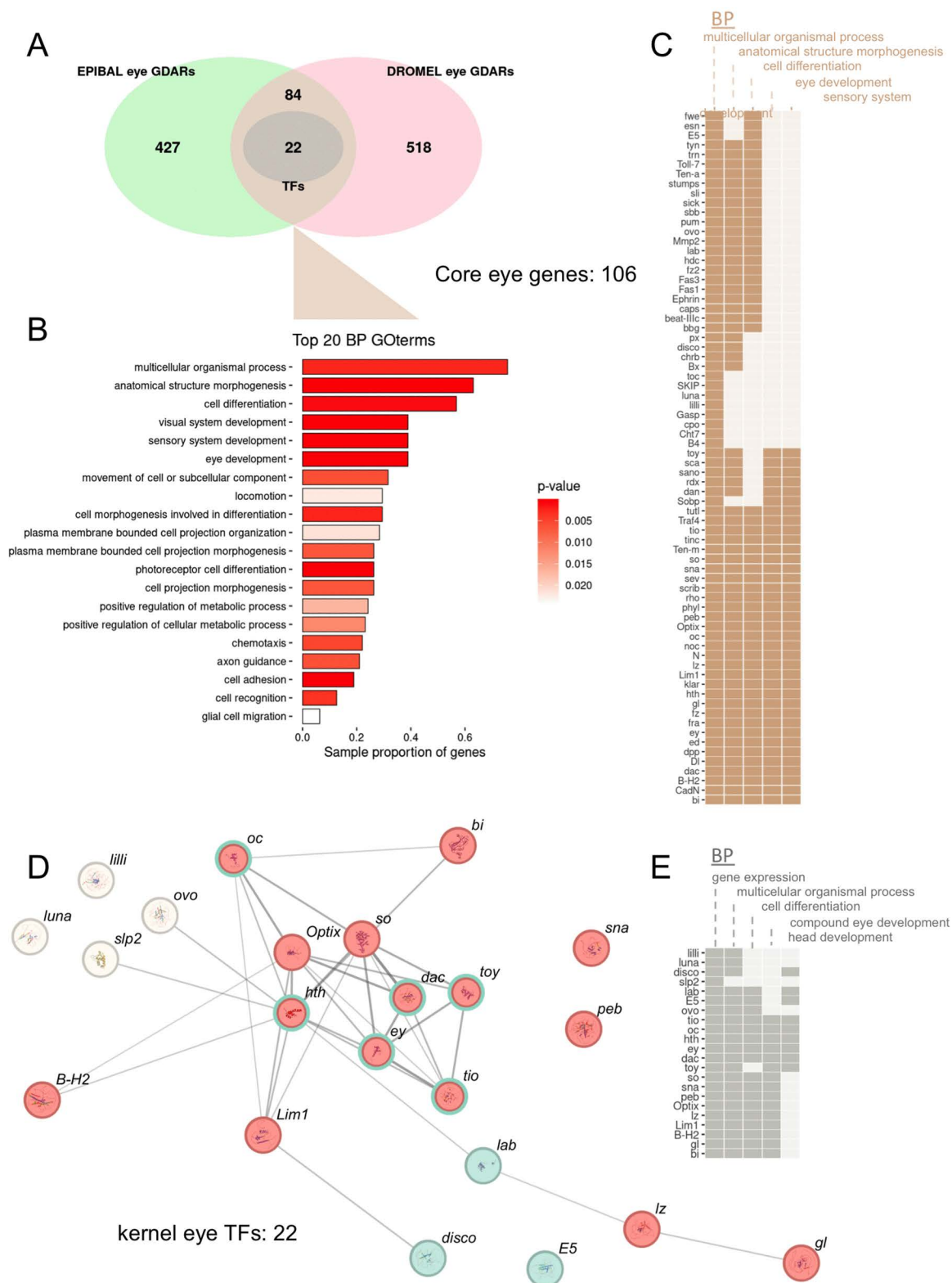


Fig 2. Analysis of the *Episyrphus* and *Drosophila* eye GDARs. (A) Venn diagram of the intersections between the two gene sets. Among the shared 106 genes ("core" gene set), there are 22 common TFs, the "kernel" TF set. (B) GO analysis of the core 106 gene set, showing the 20 top Biological Process (BP) terms. (C) List of genes, among the 106 set, with annotated terms from the 5 top GO terms. Dark cells indicates that a gene (row) is

associated with the GO term in the respective column. **(D)** STRING interaction networks of the TF kernel set using *Drosophila*-defined protein interactions. Increasing link thickness indicates higher confidence. In red, the TFs associated with eye development; in cyan, the TFs associated with head development (TFs associated with both processes share both colours); and in white, the TFs not associated with neither of the two processes. The network is highly connected. **(E)** Kernel TF set and the assignment of the relevant (non-redundant) five GO terms to each of them. As above, dark cells indicate annotation of the gene (row) with the specific GO term (columns). Note that the automatic GO annotation of *lilli* does not retrieve, for example, “compound eye development” even though it has been reported that *lilli* interacts with and regulates *ato* (and *Da*) thus affecting eye development (see main text for references). Therefore, some GO annotations may be incomplete.

<https://doi.org/10.1371/journal.pgen.1011864.g002>



Fig 3. KEGG pathway association of core genes. The KEGG pathway denomination and associated genes are indicated.

<https://doi.org/10.1371/journal.pgen.1011864.g003>

(<https://flweb.janelia.org/cgi-bin/flew.cgi>; [53]). The *Optix* locus contained four DARs, of which three are eye DARs and one is a wing DAR (S4A Fig). Eye-DAR 1 and wing-DAR fell within fragment R30B08, which showed expression in the eye and the wing. Eye-DARs 2,3 fell within fragment R30D11, which showed enhancer activity in two domains in the eye disc: the ocellar region and the compound eye. Likely, each of the two DARs in this fragment is responsible for one of these two activity domains. Therefore, all DARs detected were associated with enhancer activity and in their respective discs. Another example is the intergenic region between paralogs *B-H2* and *B-H1* (S4B Fig). In it, we detected 12 eye DARs. Nine out of twelve DARs fell within fragments with eye enhancer activity. Of the remaining three, two fell in regions that had not been tested for enhancer activity, and the third DAR was located in a region with no imaginal disc enhancer activity reported, but overlapping an annotated long non-coding RNA. The transcription of

this long non-coding RNA could explain the chromatin accessibility (S4B Fig). Again, this showed an excellent, almost one-to-one, correspondence between DARs and enhancer activity in *Drosophila*. Together with previous results, this analysis reinforces the idea that DARs generally exhibit enhancer activity, an activity that we assume can be transferred to *Episyrphus* DARs. Thereby, these DARs serve as directional anchor points between the regulators (TFs) and the network nodes (GDARs).

Prediction of TF binding motifs and the eye GRN

Establishing a regulatory link between a TF and a target gene depends on the finding of the TF motif in the DARs of the prospective target gene. Therefore, once the collection of eye GDARs had been identified, the next step was the identification of these motifs within DARs. As a method to detect TFBSs in our DARs collection, we used HOMER (<http://homer.ucsd.edu/homer/motif/>; [54]. See S1 File Methods for a comparison of motif search methods). In brief, HOMER uses Positional Weight Matrices (PWMs) to represent motifs, with a preset threshold (score_0) that the target sequence must exceed (Max. score) for the associated TF to be considered as binding. The HOMER motif collection with known score_0 and directly usable by the search algorithm included only 1005 PWMs (at the time this paper was written), and many of these (636) were discarded because they belonged to RNA-binding proteins or to TFs without an ortholog in *Drosophila*. To broaden HOMER's motif search capabilities, we integrated various PWM collections, totalling 9860 motifs (S2 File Methods). All new PWMs lacked a preset score_0 , though, so these had to be estimated in order to determine the binding criteria for their associated TFs (S3 File Methods). Using the collection of PWMs from HOMER with a known score_0 , a predictive model was trained (elastic net linear regression) which, with acceptable predictive capacity ($\sim 60\%$ R^2), later enabled the prediction of this threshold for the expanded collection of PWMs. These were the “estimable” set of motifs. However, some of these PWMs had characteristics that differed significantly from those in HOMER, which were used in the model fitting, leading to extreme score_0 estimates. These PWMs were classified as “non-estimable” (3087). Rather than discarding them, we set their score_0 —95% of the theoretical maximum score to make the TF binding predictions more conservative (S5 Fig).

With this new collection, we used HOMER to search for motifs of expressed TFs within each eye-specific DAR annotated to nearest gene (GDARs) from both species. This was made under the assumption that binding sites are generally conserved for orthologous TFs across long evolutionary distances [55]. Thus, links were established between TFs and target genes. As an overall quality check of the predicted connections, we performed two tests (see S4 File Methods). First, we asked whether motifs for TFs expressed exclusively in the eye were enriched in eye-specific DARs. Indeed, we found this to be the case for some, although not all motifs (S6A Fig). Notably, motifs directly associated with *Drosophila* TFs (that is, not derived from human or mouse orthology) and those deemed “estimable” showed eye/wing enrichment. Thus, eye/wing enrichment could be used as a measure of confidence in the links established through these motifs. As a second test, we asked if these links complied with the expectation that developmental TFs in the GRN should be, on average, more connected than other genes in the network [56,57]. We used two measures of connectivity. The number of DARs per gene and the number of unique motifs found in all DARs linked to each gene (a sort of upstream regulatory “diversity”). We observed that indeed TFs had more linked DARs as well as a larger number of upstream TF regulators, both in *Drosophila* and in *Episyrphus* (S6B and S6C Fig).

After these verifications, the eye GRN built with the 106 core genes included 700 shared edges (that is, a motif for the same TF was present in the DAR regions of the orthologous gene in both species), 384 that were *Episyrphus* specific (the motif was present only in the DARs of the *Episyrphus* ortholog) and 390 were *Drosophila* specific. The edges linking a TF to a specific motif in an eye GDAR contained information on: (1) whether the edge had been identified in either of the species or in both; (2) whether the motif derived from *Drosophila* or from human/mouse orthologs; (3) whether the motif was enriched in eye DARs relative to wing DARs, and (4) whether the motif was estimable or non-estimable (see above). With

all this information, we generated a GRN model comprising the 106 core eye genes. The GRN for the 22 kernel TFs (19 of which have an associated motif) is presented in [S7 Fig](#).

One first observation, both in the *core genes* network and the *kernel TF* network was that the degree of conservation of regulatory links varies between pairs of orthologs. To compute the different degrees of shared connectivity observed, we compared the set of TFs (motifs collapsed by TF) potentially binding the regulatory landscape (sum of associated DARs) of each orthologous gene pair. Using the Jaccard distance as a measure of dissimilarity -that is, two orthologs were deemed regulatory identical ("least dissimilar") if their DARs had the same collection of binding TFs- we could sort the core genes by degree of shared connectivity ([S5 Table](#)). On the contrary, they would be most dissimilar if their TF collections were non-overlapping.

Considering just the collection of potential upstream TFs does not reflect the fact that the expression pattern of a gene is usually the result of the piecemeal action of several enhancers, each often active in distinct spatial and temporal subdomains (see, for example, the cases in [S4A](#) and [S4B Fig](#), above). Therefore, comparison of the TF-binding composition of individual enhancers might reveal different degrees of similarity among enhancers, with those most similar being responsible for shared aspects of the gene's expression pattern, while those most dissimilar conferring species-specific regulation. Thus, we decided to cluster (using hierarchical clustering) the DARs of each pair of the 106 orthologs of the core set, based on their TF-binding composition (using Jaccard distances between pairs; [S5 Table](#))

Identification of genes with potentially highly conserved regulation based on DAR distribution, TF motif composition and expression pattern: *glass (gl)* as an example

One example of a gene showing a high degree of predicted regulatory similarity is the TF *gl (glass)*, which is expressed and required for the differentiation of the neural and non-neural cell types of the retina [[58–61](#)]. *gl* ranks 100 out of 106 genes in the list of increasing connectivity similarity ([S5 Table](#)), and it is the most similar among the kernel TFs factors ([S6 Table](#)). The two orthologous *gl* shared most of their predicted TF regulators ([Fig 3A](#); shared kernel subset). When the two loci were compared, both showed two eye DARs, one overlapping the promoter and another immediately 5' to it ([Fig 3B](#)). Clustering analysis based on predicted TF binding indicated that *Drosophila* eye DAR-1 and *Episyrphus* eye DAR-1 and 2 were the most similar, sharing binding sites for Optix, Hth, Ovo, So, E5, Limb1, Luna and Lab ([Fig 3C](#) and [3D](#)). Analysis of gene transcription in single cell data shows *gl* coexpression with *so*, *B-H2*, *luna*, *ovo* and *hth* in the differentiating retina population. *Optix* is expressed in eye progenitors and precursors, and is mutually exclusive to *gl*, suggesting a repressive interaction ([S8 Fig](#)). *E5* and *lab* are expressed in prospective head cells, questioning their relevance as *gl* upstream regulators ([S8 Fig](#)). These results suggested a very similar regulation of *gl* in *Episyrphus* and *Drosophila*. To test this, we analyzed *gl* expression in *Episyrphus* eye discs using *in situ* hybridization (ISH, [Fig 3E](#)). *gl* transcription is detected posterior to the MF, with the highest levels immediately after it, then weakening in more differentiated retinal cells, exactly as has been described for *Drosophila gl* using also ISH [[61](#)]. Therefore, the similarity we detected in *gl* TF regulation and DAR similarity correlated with the high similarity in the expression pattern of the *gl* orthologs ([Fig 4](#)).

Variation of DAR number and location in orthologous loci: *dachshund (dac)* as an example

We noted, however, that in general, neither the number of eye DARs nor their genomic positions were conserved between orthologous genes. Furthermore, when TF motif composition similarity was assessed using our clustering approach, the degree of similarity among DARs linked to orthologous genes varied widely. Nevertheless, as a general trend, the degree of DAR similarity was low (See [S9 Fig](#) for some examples of orthologs). Among these, only E5 DARs showed a high degree of similarity ([S9 Fig](#)). These observations suggest that either equivalent regulatory information may be distributed differently across *cis*-regulatory regions, or it may have diverged evolutionarily. Such differences may or may not lead to changes in gene expression patterns. In the absence of a comprehensive characterization of the expression patterns of the kernel TFs in *Episyrphus*, we began characterizing *dachshund (dac)* as an example of divergence in the regulatory

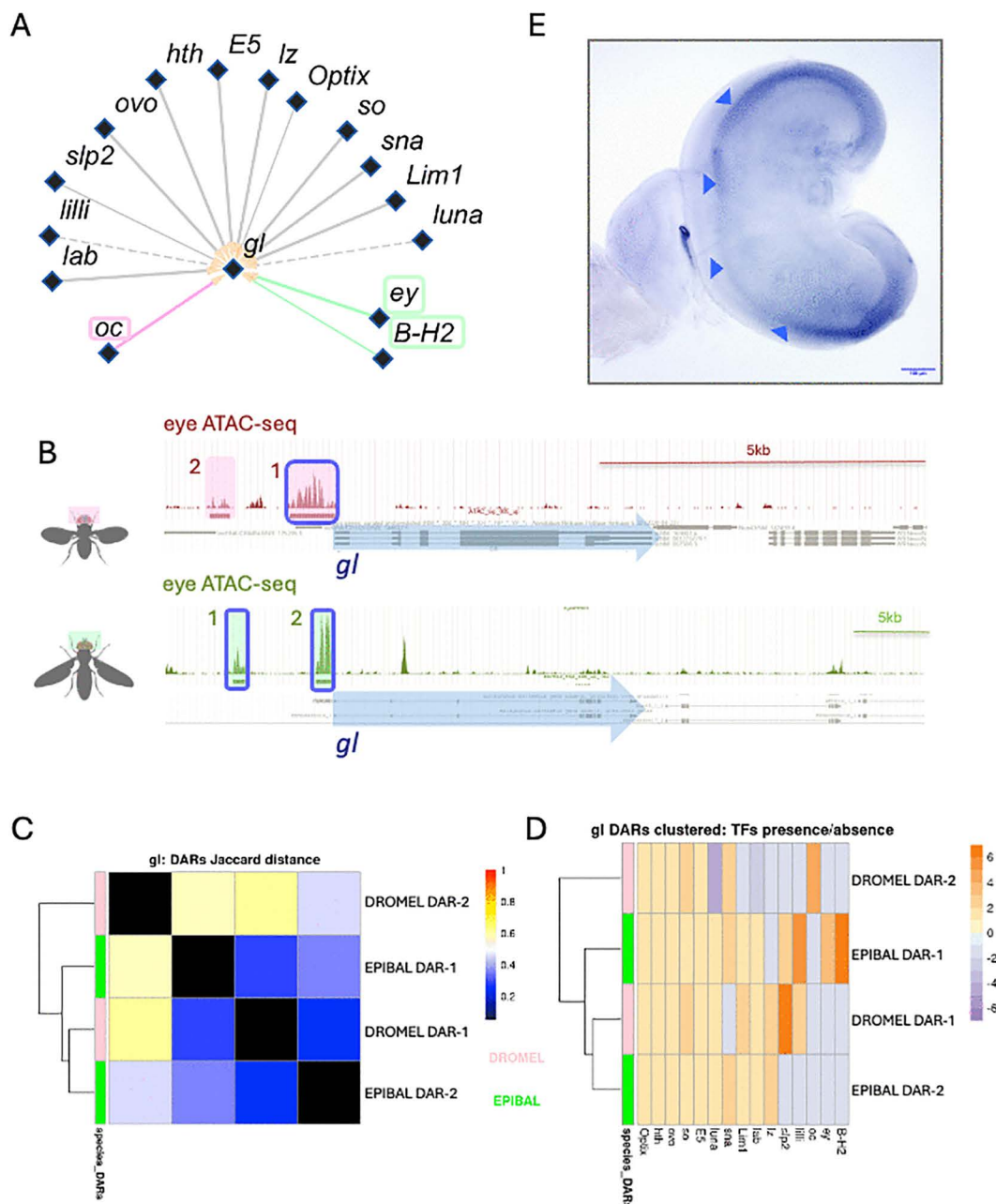


Fig 4. Comparative GRN exemplified by the *glass* (*gl*) locus. (A) *gl* regulatory network considering the kernel of TFs. Links found in *Drosophila* (“DROMEL”) and *Episyrrhus* (“EPIBAL”) are represented in pink and green, respectively. Most links are found in both species (grey), except for *Oc/Otd* (*Drosophila*-specific) and *Ey* and *B-H2* (*Episyrrhus*-specific). Line type: *solid*: DROMEL motifs; *dashed*: human and mouse motifs; Increasing Line Transparency: estimable to non-estimable; Increasing Line Width: non enriched in eye versus wing to enriched (see S3 File Methods for details). (B) Genome browser view of the *gl* locus in *Drosophila* and *Episyrrhus*. The transcription unit is marked by the blue arrow pointing to 3'. DARs are shaded and numbered. (C) Clustering of *gl* DARs according to their motif composition similarity (kernel TFs), with colder colors indicating increasing similarity. *Drosophila* and *Episyrrhus* DARs are in pink and green, respectively. The most similar DARs are DROMEL DAR-1 and EPIBAL DAR-1,2. (D) Same clustering showing distribution of kernel TF presence. The color warmth is inversely proportional to the frequency of finding the motifs for a given TF in the set of eye-GDARs for each species, and can be interpreted as an indication of how much more atypical a motif is compared to a ubiquitous one. DROMEL DAR-1, and EPIBAL DAR-1 and 2 all contain motifs for *Optix*, *Hth*, *Ovo*, *So*, *E5*, *Lim1*, *Luna* and *Lab*. (E) *gl* expression in *Episyrrhus* eye discs, detected by mRNA *in situ* hybridization (ISH). *gl* expression starts posterior to the MF (arrowheads). Immediately after the MF *gl* expression is highest, with the signal becoming weaker more posteriorly. Scale bar: 100 microns.

<https://doi.org/10.1371/journal.pgen.1011864.g004>

predictions of the GRNs. *dac* is a central gene within the *Drosophila* retinal determination gene network: its expression is kept low in eye progenitors but becomes transcriptionally upregulated by the protein complex formed by So and Eya in eye precursors just anterior to the differentiation wave (morphogenetic furrow, MF). *dac* expression remains high straddling the MF and just posterior to it, but then decays as retinal differentiation progresses. In addition, *dac* is expressed in a ring in the adjacent antennal primordium [62,63]. When we examined *dac* expression in *Episyrphus*, using a cross-reactive antibody, we detected *dac* expression in both the antennal ring and the eye (Fig 5A). In the eye, *dac* was expressed in all cells anterior to the MF, with expression decreasing posterior to it (Fig 5A). Therefore, *dac* expression is globally similar in *Drosophila* and *Episyrphus*. Next, we looked among the kernel set of TFs for potential *dac* regulators in our GRN model (Fig 5B). Of these, a subset comprised regulators shared by both species, another subset was *Episyrphus*-specific and finally, *lilli* and *gl* were predicted to be *Drosophila*-specific, although these were low confidence motifs (coming from human/mouse homologs, not eye-enriched and non-estimable). The common predicted regulatory inputs on *dac* were *hth*, *ey*, *optix*, *so*, plus *Iz* (*lozenge*), *E5*, *sna* (*snail*), *ovo* and *luna* (although this latter is also a low confidence motif). *Hth*, *Ey* and *So* had already been shown to be *dac* regulators [62,64], and *Optix* was a likely prediction, considering its role in eye development [65] and its capacity to induce eye development when ectopically expressed [66]. The fact that our GRN model included all known *dac* regulators increased our confidence in other predicted links. Thus, we decided to test one of those novel links: *Iz*.

Iz encodes a TF of the AML1/Runx gene family. In the *Drosophila* eye, *Iz* is required posterior to the MF for the differentiation of most retinal cell types, including photoreceptors, lens-secreting and pigment cells [41,67]. Previous work had shown that *Dac* and *Lz* expression domains were complementary [68], something we confirmed using single-cell transcriptomic data (Fig 5C), suggesting already a repressive interaction between *Iz* and *dac*. We carried out three experiments in *Drosophila* to test this regulatory relationship. First, we checked the *in situ* expression of *dac* relative to *Iz* detecting *Dac* in a strain in which *Iz* had been GFP-tagged (*Iz:GFP*). *Lz* is expressed in cells whose nuclei lie mostly basally [67]. *Dac* expression decays posterior to the MF coinciding with the expression of *Iz:GFP* (Fig 5D and 5D'), confirming previous results [68]. Second, we analyzed *dac* expression in an eye-null *Iz* allele [67]. In *Iz77a7* primordia, the reduction of *Dac* expression in basal nuclei is less marked (compare Fig 5D and 5E), consistent with the quantitative analysis (Fig 5F), indicating that *Iz* is required for *dac* down-regulation. Finally, we ectopically drove *Lz* anterior to the MF, where *dac* is expressed [69]. In *optix2/3-GAL4; UAS-Iz* ("*optix>Iz*") eye discs, *dac* expression was dramatically down-regulated (Fig 5G and 5G'). Therefore, our experiments showed that *Iz* is necessary and sufficient for *dac* down-regulation in *Drosophila*, thus supporting the regulatory link from *Iz* to *dac* derived from our GRN analysis—a link that should exist in both species. Still, definitive proof for *direct* regulation would require biochemical evidence of *Lz* binding to the *dac*'s regulatory regions.

Variation of TF binding motif composition in DARs of orthologous *dac*

Having identified potential TFs regulating *dac* expression, we next explored how their motifs were distributed in the *dac*-associated DARs in *Drosophila* and *Episyrphus*. To this, we mapped the identified DARs on the orthologous loci. *Drosophila dac* showed two DARs, one intragenic and another immediately 3' to the end of the transcription unit, while *Episyrphus dac* had four, all of them 3' to the end of the transcription unit (Fig 6A). The *Drosophila dac* DAR1 and *dac* DAR2 mapped within the sole *dac* eye enhancers known to date [64] and Fig 6) confirming the sensitivity of our ATAC-seq data. When we clustered all six DARs, *Episyrphus* DARs 1, 2 and 4 clustered closer to *Drosophila* DAR1, while *Episyrphus* DAR3 was the most dissimilar among DARs (Fig 6B) although the level of similarity was, in general, weak. In particular, *Episyrphus* DAR1 shared with *Drosophila* DAR1 motifs for *hth*, *Optix*, *so* and *Iz*, among others (Fig 5C). *Drosophila* DAR1 (3EE [64]) has been shown to be expressed in the developing eye [64] and in the developing lamina of the optic lobe, where a gene regulatory network similar to the retina is at place [70]. We decided to test whether these shared motifs were predictive of eye and optic lobe-specific activity. We cloned the *Episyrphus* DAR1 upstream of GAL4 and examined its enhancer activity in *Drosophila* (*EPIBAL dac_DAR1-GAL4; UAS-GFP*). Indeed, this enhancer showed

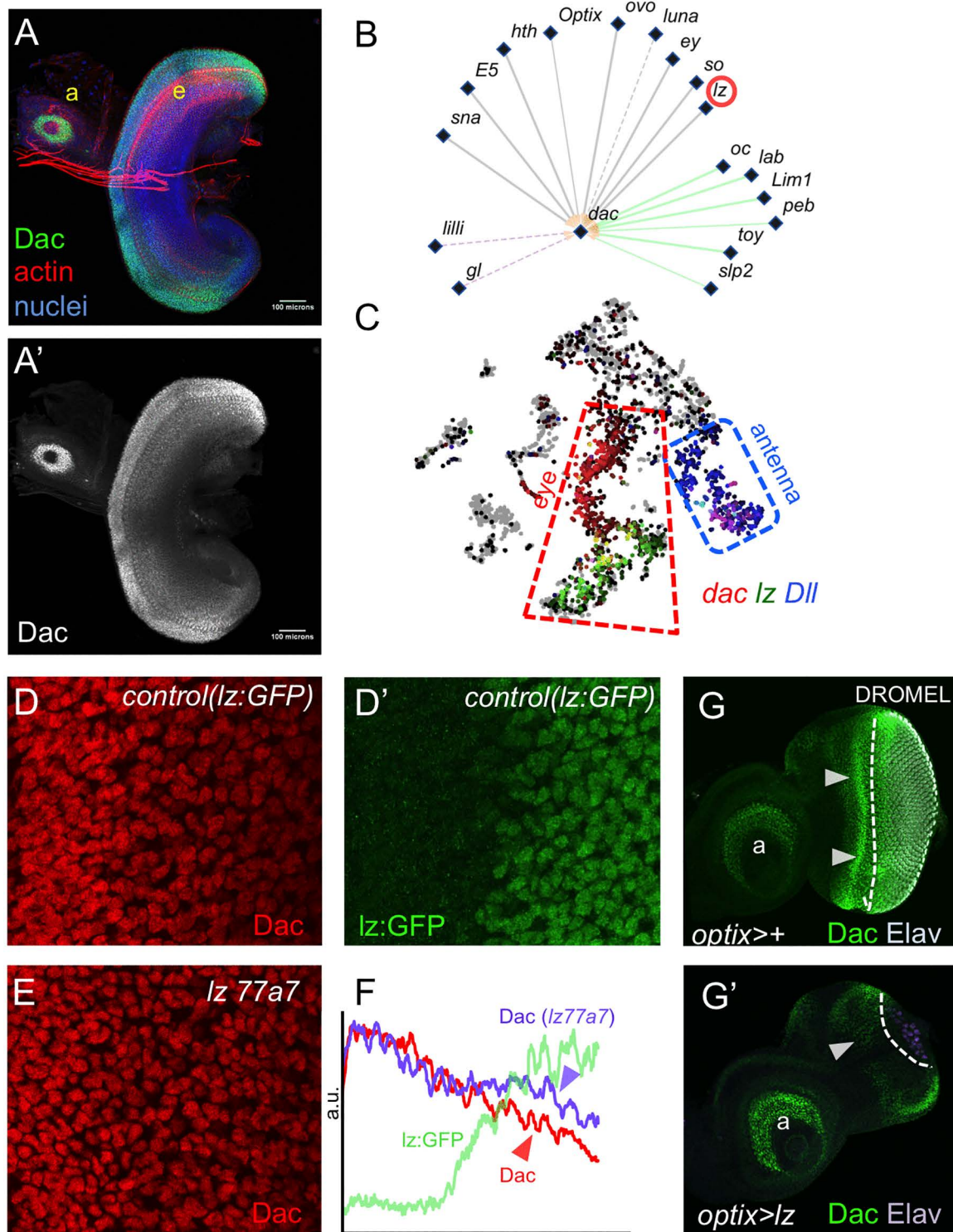


Fig 5. *dac* expression and its upstream regulation. (A,A') *dac* expression in an L3 *Episyrrhus* primordium stained, in addition, for nuclei (DAPI, blue) and actin (phalloidin, red). The *dac* signal is shown in (A'). The antennal and eye regions are marked by "a" and "e", respectively. (B) Predicted regulatory inputs on *dac* from the kernel set of TFs. Shared inputs are in grey, *Episyrrhus*-specific are in green. *Drosophila*-specific links, which are low confidence, are shown in pink. The TF *lozenge* (*Iz*) is highlighted, as its predicted regulatory action on *dac* is tested in this paper. (C) SCOPE single cell atlas

showing the complementary expression of *dac* and *Iz*. The eye cells are boxed. The expression of *Dll* is also included to mark the antennal cells (where *Iz* is expressed as well). *Drosophila* eye primordium single cell data is from 5. (D,E) *dac* expression in basal cells in *Iz:GFP* (Control) (D) and *Iz* mutant (*Iz77a7*) in representative *Drosophila* eye primordia. In controls, *dac* expression (Dac, D) decreases as *Iz:GFP* expression (*Iz:GFP*, D') increases. (E) This decrease is less marked in *Iz77a7* mutants (F) Signal quantification of (D, D' and E). The Dac profile in control and *Iz77a7* eyes is represented in red and purple, respectively. The *Iz:GFP* profile (from D) is shown in green. (G,G') *dac* expression in control (*optix>+*) and in *Iz*-overexpressing eye primordia (*optix>Iz*). The arrows point to *dac* expression in the *optix* domain, anterior to the MF (dashed line). In *optix>Iz* (G') *dac* expression is almost completely repressed, compared to the controls (G). Retinal differentiation, marked with the photoreceptor marker *Elav*, is also impaired in *optix>Iz* primordia. "a", antenna.

<https://doi.org/10.1371/journal.pgen.1011864.g005>

specific expression in both domains (Fig 6D–6D"). Detailed analysis showed that *EPIBAL dac_DAR1 > GFP* overlaps the *dac* domain, although it starts posterior to the anterior-most limit of *dac* expression (Fig 6E). This result indicates that the *Episyrphus* sequence can read the regulatory state of the *Drosophila* eye and optic lobe almost correctly, suggesting that a similar regulatory state (that is, a similar set of TFs and active signaling pathways) is present in *Episyrphus*. As we noted above, the motif composition is not particularly similar between *Drosophila* *dac-DAR1* and *Episyrphus* *dac DAR1*, and yet, the expression pattern driven by the latter in *Drosophila* recapitulates the expression domains where the former is active. This may indicate that the combination of the shared TFs is sufficiently specific to activate *Episyrphus* *DAR1* in the correct territories. Other TF motifs may play secondary or refining roles. Another, non-mutually exclusive explanation, might be that both *Drosophila* *dac_DAR1* and *Episyrphus* *dac_DAR1* shared additional sequences not included in our motif collection, or some sequence "grammar" we have not considered—that is, these DARs would actually be more similar than our method could estimate.

Discussion

The comparison of *Episyrphus* and *Drosophila* eye transcriptomes extends the set of eye genes

Most comparative analyses of compound eye development have been carried out either among closely related *Drosophila* species [71] or between *Drosophila melanogaster* and distant species within insects, or even further with chelicerates and myriapods [18, 21, 72–77]. In general, long-range evolutionary comparisons involve a small set of genes selected based on prior knowledge derived in *Drosophila*. Here we have aimed at starting to bridge this evolutionary gap by constructing an eye GRN model in the hoverfly *Episyrphus balteatus* and comparing it to a *Drosophila* eye GRN built using an identical approach.

The first result was the global characterization of genes differentially expressed in the developing eye disc compared to the wing disc. Here, the eye gene sets ("eye DEGs") point to the same functions in the two species and match perfectly the developing eye processes (Fig 1). However, even though both eye gene sets are characterized by similar functional and molecular terms, there is only limited overlap between them (158 shared eye DEGs). Many of the *Episyrphus*-specific eye genes are annotated as "eye-related" because their *Drosophila* orthologs carry the same annotation. The fact that these genes were not detected as eye-specific in *Drosophila* likely reflects differences in their relative expression levels between the eye and wing discs of the two species (as in the example of the *bi* gene (see above), the levels of which are too similar between eye and wing discs to be detected as differentially expressed). This could explain the limited overlap in eye-expressed genes. Since the process of photoreceptor differentiation is exclusive to the eye, we expected this 158 gene set to be particularly enriched in genes expressed specifically in retinal cells. This was indeed the case: 34% of them were associated to Nervous System Development (Fig 1G and 1H), a process that, in the eye-antennal primordium mostly represents retina development and differentiation. Therefore, the set of 158 shared eye DEGs (Fig 1) should be considered a conservative list of genes specifically associated to the specification and early differentiation of the retina across a broad swath of the Diptera tree of life.

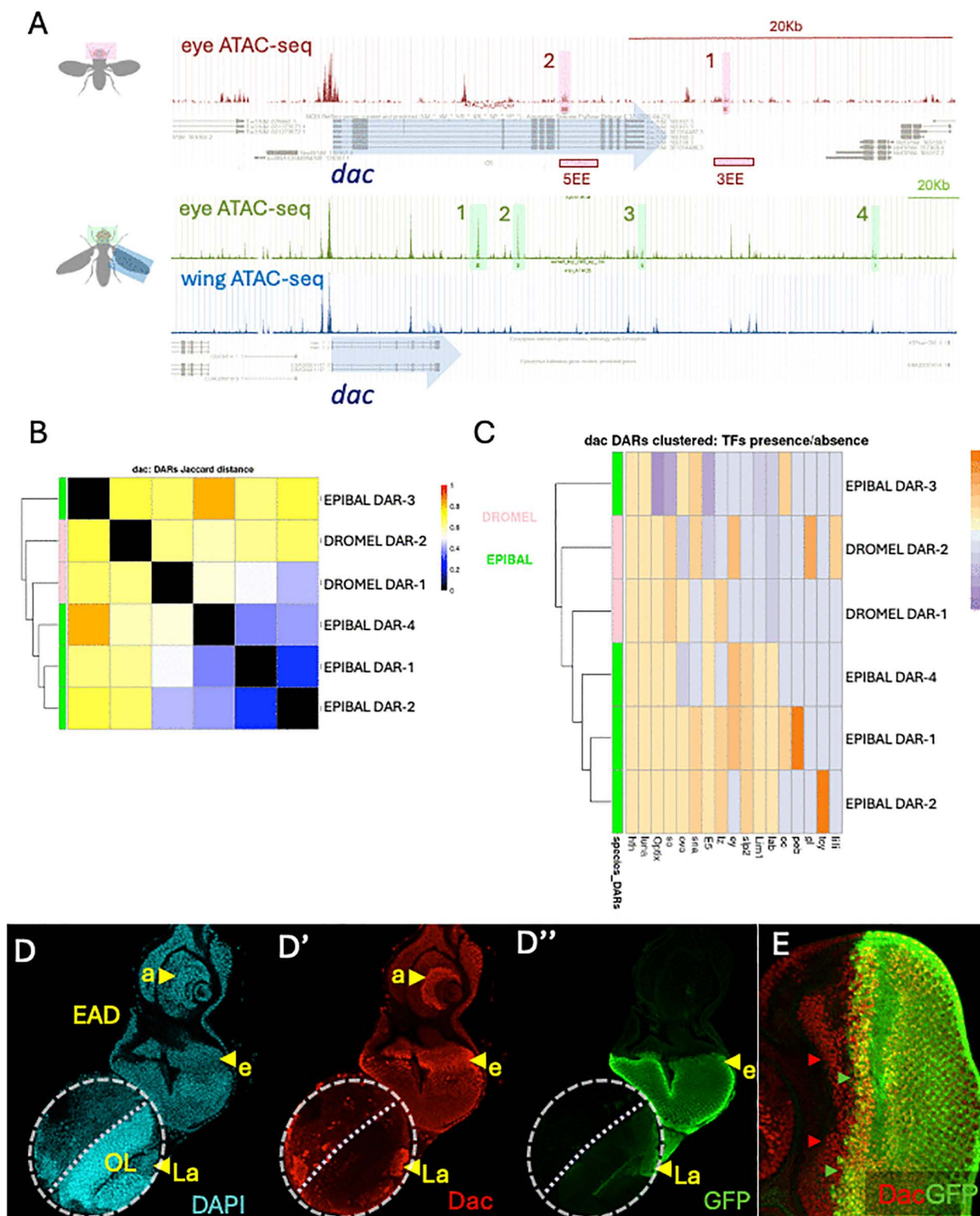


Fig 6. The *dac* regulatory landscape and similarity-based clustering of its potential enhancers. (A) *dac* locus in *Drosophila* (top) and *Episyrphus* (bottom). Genome browser eye tracks (one out of two replicates) are shown. One wing track for *Episyrphus* is also included, to show that these DARs are eye specific. The transcription unit is marked by the blue arrow pointing to 3'. DARs are shaded and numbered. The two *Drosophila* eye DARs

correspond to the already characterized 3EE and 5EE enhancers by Pappu et al., 2004 (boxes below the tracks). The ATAC-seq data also identifies clear eye-specific DARs in the *Episyrphus* genome. However, instead of two there are four and, in any case, located in non-syntenic positions relative to *Drosophila*, except for, perhaps, DARs 1 in *Drosophila* and *Episyrphus*, which are located immediately downstream of the transcription unit. **(B)** Clustering of *dac* DARs according to their motif composition similarity, with colder colors indicating increasing similarity. *Drosophila* and *Episyrphus* DARs are in pink and green, respectively. The collection of motifs used was that for TFs shared by both species ("kernel" set). **(C)** Same clustering showing the distribution of kernel TF motifs. Color warmth inversely proportional to TF frequency in the whole DAR set (as in Fig 4). Shared kernel TFs between DROMEL *dac*_DAR1 and EPIBAL *dac*_DAR1 are: *hth*, *Optix*, *so*, *Iz*, *luna*, *ovo* and *E5*. **(D,E)** Activity of the EPIBAL *dac*_DAR1 enhancer in the *Drosophila* larval visual system. **(D-D')** Visual system of a EPIBAL *dac*_DAR1>GFP L3 larva, stained for DAPI (nuclei, **D**), *Dac* (**D'**) and GFP (**D''**). "EAD": eye antennal disc. The central brain (outlined) includes the Optic Lobe ("OL"). "a": antenna; "e": eye; "La": lamina of the optic lobe. EPIBAL *dac*_DAR1 drives GFP expression in the eye and lamina. **(E)** Higher magnification view of an L3 eye disc of the same genotype, stained for *Dac* and GFP. The anterior border of the GFP signal (green arrowheads) is detected posterior to *Dac*'s anterior border (red arrowheads).

<https://doi.org/10.1371/journal.pgen.1011864.g006>

Some considerations on our GRN-building strategy

A critical step for GRN modeling was the identification of eye-specific CREs and their linked genes, as these CREs would be the primary sites of GRN control. As we described above, we identified eye DARs (that is, ATAC peaks that were detected in eye but not in wing discs) and linked them to the closest eye-expressed gene. Indeed, DARs have been shown to act as enhancers through *in vivo* transgenic, enhancer-reporter assays in the *Drosophila* embryo [49], in the *Drosophila* eye [5], and in the *Drosophila* brain [50]. More recently, through an international challenge the same principle was shown in the mouse brain, where differential ATAC-seq signal outperformed any other method to predict cell type specific enhancers [51]. Nevertheless, this approach has a number of caveats. A first caveat is that of the limits of a regulatory locus. While the association between an intragenic DAR and the hosting locus is usually safe, the association of an intergenic DAR to any gene upstream or downstream to it may depend on where the chromatin boundaries lie [78], so that in addition to proximity, the association should take into account boundaries, something we did not do, as this type of data was not available for *Episyrphus*. However, GO analysis of the genes linked to eye-specific DARs indicated that, globally, the associations should be correct, since the enriched terms clearly pointed to eye development in both *Drosophila* and, especially, in *Episyrphus*. In addition, we found enrichment of motifs for TFs expressed in the eye in eyeDARs, when compared to wing DARs, especially for motifs identified for *Drosophila* TFs, globally supporting the connectivity of our network model.

A second caveat affects the sensitivity of DAR detection. Although the probability of missing a DAR in a bulk experiment increases with the rarity of the cell state within the tissue in which the enhancer is active, we were aiming at the gene network responsible for the specification and growth of the eye and the initiation of its differentiation. During these processes, the relative proportion of cells in the same state is expected to be large and therefore, it is also very likely that many cells are activating the same CREs (that is, opening their regulated DARs), making them readily detectable using ATAC-seq. This is why we expected that, by and large, we should not be missing major enhancers in neither of the two species. The *Drosophila* gene sample for which we studied the correspondence between functional enhancers and DARs, although very limited, showed almost one-to-one correspondence, supporting the idea that our ATAC-seq data was not missing major enhancers in *Drosophila*, which led us to assume that likely neither it did in *Episyrphus*.

A third caveat is that DAR-based enhancer prediction may miss enhancers that are constitutively accessible, but not differentially accessible. This may occur when a pioneer TF is bound to an enhancer, such as Zelda-bound regions in the *Drosophila* embryo showing broad accessibility, yet specific activity; or Grainyhead-bound regions in epithelial cells of the eye-antennal disc [79]. Given that the Grainyhead pioneering function is conserved from flies to human [79], we assume that this binding mechanism is also conserved between *Drosophila* and *Episyrphus*, and hence does not interfere with our strategy of finding species-specific DARs. The identification of species-specific constitutive ATAC peaks may be an interesting avenue for the future, but requires additional data modalities or sequence models to distinguish enhancers from other constitutive peaks.

A final caveat is that our enhancer predictions do not use the sequence under the ATAC-seq peak for the initial selection, we rather used motif discovery a posteriori. For initial selection, we deliberately avoided the use of PWM-scanning approaches, because these yield notoriously high false positive predictions (see [80] and references therein). In future studies, sequence-to-function deep learning models may be used jointly with DARs, to improve enhancer prediction [51,81]. Such models however would require large single-cell ATAC-seq atlases of both species, to be used as training data.

In conclusion, for our data set, a DAR-based strategy was considered the optimal approach to predict candidate enhancers, even though we cannot assure 100% recall of eye-specific enhancers.

A conserved eye GRN integrated in the global regulatory network of the developing head

With all these considerations in mind, we built the GRN model with the 106 eye GDARs. In Fig 7 we summarized a large part of this network. In it, we show the eye GRN comprising the core genes expressed in the developing eye (as in the single cell atlas by [47]). These genes (nodes) were located more or less centrally depending on their degree of connectivity, with their size representing their degree of regulatory similarity (with larger nodes being more similar). In addition, we include temporal information on their expression based on the single cell atlas of *Drosophila* eye disc produced by the Mardon laboratory [47]. We further assumed, as a first approximation, that the same temporal dynamics applied to the *Episyrphus* orthologs.

When kernel TFs were considered, the resulting network was highly interconnected, in agreement with the retinal GRN being subject to many regulatory feedbacks. Of all kernel TFs, only *sna* and *ovo* lack evidence of being involved in eye disc development. Our network includes all the known *Drosophila* retinal determination gene network (RDGN) members (*ey*, *toy*, *so/Six2*, *Optix/Six3* and *dac*; reviewed in [8]) and extends these cross-regulations to include further TFs, involved in stimulating the growth of eye progenitors (*hth*, *tio* and *tsh*) [82–85] or in driving retinal differentiation (*gl*, *peb*, *lilli*, *luna*, *disco*, *Iz*, *B-H1*, *B-H2*) [37–42,86]. Some of the kernel TFs have been previously involved in the development of disc derivatives other than eye, like the head (*oc*, *lab*, *Lim1*, *slp2*) [36,43–46]. Since the starting material was the eye-antennal disc, this was expected. The fact that we predicted them connected within the network might indicate regulatory relationships between the genetic programs controlling the eye and other regions of the disc, like the mutual repression between *ey* and *cut*, which has been proposed previously to be responsible for the partitioning of the disc into eye and antennal fields, respectively [87,88]. Therefore, the eye GRN should be considered a part of a global gene regulatory network controlling the development of all the head structures from *Episyrphus* to *Drosophila* at least. In addition, we identified four cases, in which the kernel TF (*tio*, *B-H2*, *E5* and *slp-2*) was part of a linked paralogous gene couple: *tio/tsh*, *B-H2/B-H1*, *E5/ems*, and *slp-2/slp-1*. It has been previously shown for these pairs that paralogs have very similar, if not identical expression patterns [36,82,89–91]. In the case of *E5* and *ems* single cell RNAseq data showed also coincidental expression (S3A and S3B Fig). Therefore, the corresponding paralogs may well be part of the conserved eye GRN.

Evolution of CREs: Conservation and divergence in orthologous loci of developmental genes and developmental system drift

The comparative analysis of CREs between *Drosophila* and *Episyrphus* reveals significant variability in the conservation and organization in the orthologous loci of developmental genes. While some genes, such as *gl*, exhibit highly shared connectivity, others, such as *dac*, show greater divergence, suggesting varying degrees of evolutionary flexibility in gene regulation. This variability in shared connectivity implies that not all genes in the GRN are equally free to evolve, which could influence the stability or plasticity of their expression patterns. However, we did not identify any rule that could help us predict the shared connectivity observed. This metric (shared connectivity) showed no correlation with the degree of connectivity and remained invariant across gene ontologies.

A particular point of interest is the *Drosophila* enhancer *dac* 3EE and its potential counterpart in *Episyrphus* (DAR-1). Despite the low similarity in motif composition, DAR-1 exhibited regulatory activity in the eye and optic lobe lamina of *Drosophila*, similar to that of 3EE, suggesting that only a subset of the TFs predicted as regulators might be truly

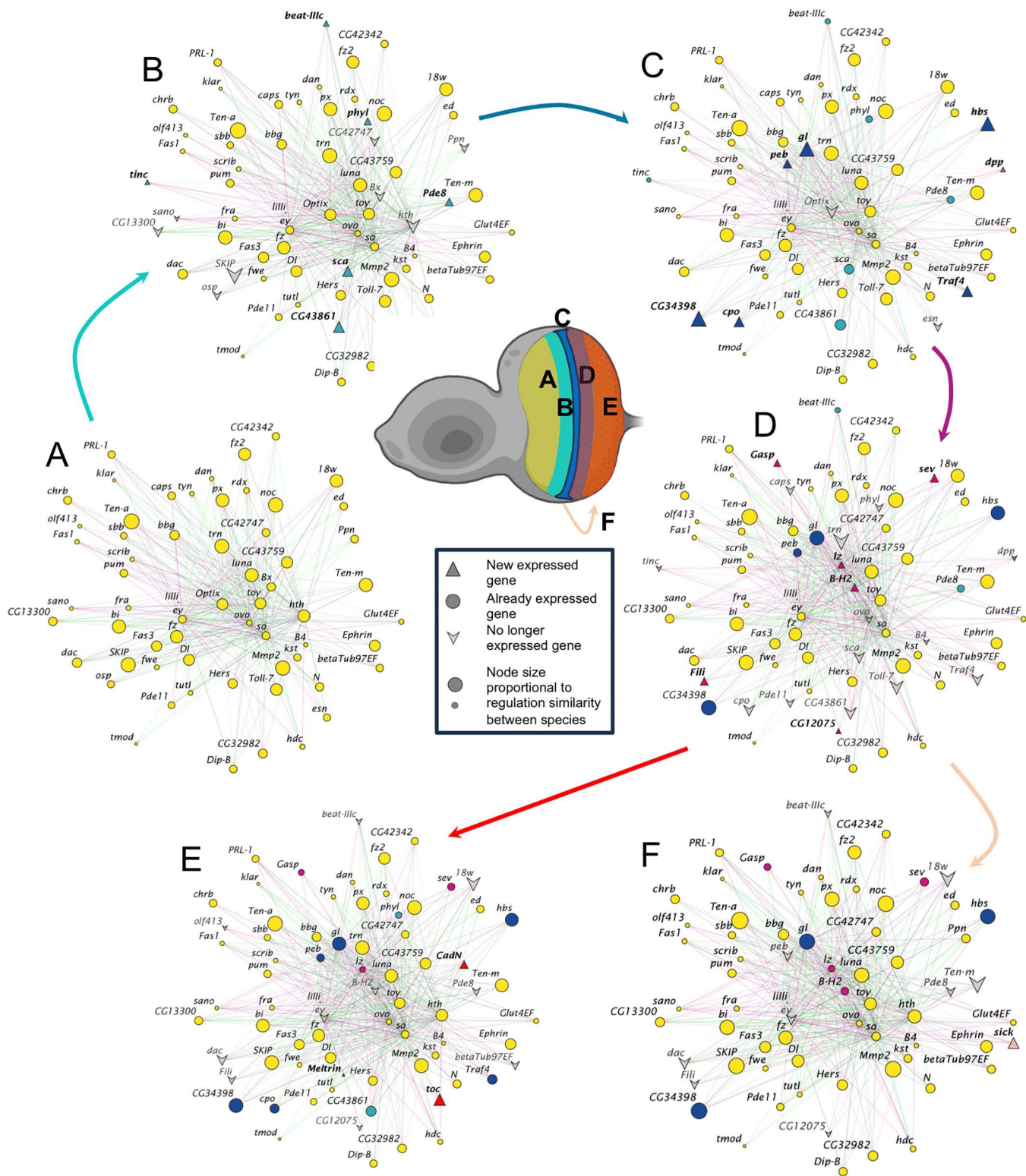


Fig 7. The collapsed eye GRN containing the core gene set and its predicted regulatory interactions. The network represents the changes in gene expression as they occur during *Drosophila* eye development (see Materials and Methods/Identification of core genes expressed in each eye

primordium cell population using single-cell RNA-seq). (A) Eye progenitor; (B) Precursor; (C) morphogenetic furrow region; (D) early photoreceptor; (E) late photoreceptors and accessory cell types; (F) basal retinal progenitors. Color and line type of links as in Fig 4. The colors of the symbols and the arrows connecting the networks represent the cell populations of the disc as represented in the center of the figure.

<https://doi.org/10.1371/journal.pgen.1011864.g007>

functional, with the rest being redundant or subtle modulators of expression. However, since this assay was conducted in *Drosophila*, the result depends on the TFs expressed in the developing eye of this species. To confirm that DAR-1 in *Episyrphus* exhibits equivalent activity, functional assays in its native context would be required. In any case, the fact that an *Episyrphus* enhancer shows an activation pattern consistent with its associated gene, in this case *dac*, indicates that the *trans-cis* code of the regulatory regions (that is, the correspondence between TFs and their target DNA binding motifs) is conserved between *Drosophila* and *Episyrphus*, allowing *Episyrphus*' enhancers to "interpret" the regulatory state of *Drosophila* cells driving a coherent pattern. A recent example of this "coherence" has been reported for *Aedes* enhancers in *Drosophila* transgenics [92].

More broadly, the comparison of orthologous loci suggests that the distribution of CREs has shifted over the 90 million years of evolution separating the two species. The lack of conservation in sequence and arrangement of the DARs supports the idea that regulatory information is highly dynamic and can be rearranged without necessarily altering the expression patterns they control during evolution, which has been called "developmental system drift" [93]. One way in which this developmental drift may be implemented at the transcriptional control level is in the "fluidity" of enhancer grammar. One example of this fluidity is the *even-skipped* stripe 2 enhancer in *Drosophila*, where the precise number and distribution of TF binding sites within the enhancer changes over relatively short periods of time without affecting the enhancer's activity [94–96]. This fast turnover of TF binding sites is a general feature of *Drosophila* embryonic enhancers [94,97]. In our analysis, though, since we have not taken into account numbers or relative distributions of TF binding motifs, but their presence/absence, differences in CREs reflect differences in motif composition and potentially differences in their regulatory links. Still, if we compared how orthologous gene pairs cluster according to the similarity in motif composition of their DARs relative to the clustering of orthologs with a random assignment of DARs, more orthologous pairs remained clustered than expected by chance (S10 Fig), indicating that, at least for *Drosophila* and *Episyrphus*, a certain degree of shared motif composition exists, either deriving from a common ancestral CRE or from functional constraints demanding CREs with specific regulatory activities.

Another way in which regulatory flexibility may arise is the emergence of novel enhancers with similar or partly overlapping regulatory activity, called "shadow" enhancers [98]. Recent work shows that a relatively small number of mutations can turn a random sequence with no enhancer activity into a cell type-specific enhancer, suggesting that enhancer function is easily evolvable [99]. The presence of two (or more) partly redundant enhancers might thus lead to the redistribution of regulatory information among the enhancers or even the new enhancer taking over the ancestral one, something that could lead to the degeneration of the latter. This phenomenon would result in the loss of synteny between synonymous (or functionally equivalent) enhancers in orthologous genes. This might be a rather general feature, at least among developmental genes. This phenomenon of "enhancer shuffling" may require sufficient evolutionary time, though. In a recent analysis of the *Wnt5A* locus, involved in setting the pigmentation pattern of butterfly wings, most CREs showed sequence conservation and synteny among a sample of four nymphalid butterflies. However, the number of non-conserved, non-syntenic CREs increased dramatically in *D. plexippus*, which belongs to the Danainae, the sister group of nymphalids –and therefore, more distant in the evolutionary tree of butterflies [100]. In the gene sample analyzed here, which comprised in addition to *gl* and *dac*, the TFs *ey*, *Optix*, *bi (omb)* and *Iz*, as well as the BMP signal *dpp* (Figs 4, 6 and S9), the number and location of eye DARs varied between most orthologous pairs, suggesting that *Drosophila melanogaster* and *Episyrphus balteatus* have undergone sufficient time of separate evolution to have allowed for a significant redistribution of regulatory information. All this evidence reinforces the notion that regulatory evolution is rather fluid, generating regulatory diversity without compromising the functionality of developmental programs.

The extended, conserved GRN model allows making testable predictions about gene regulation

The final GRN model included a set of shared 106 eye genes connected by conserved and species-specific regulatory links. Our criteria for calling the presence of a TF motif were relaxed (for example, we did not include any motif “grammar” as criterion). However, we observed an enrichment in TF motifs for eye-specific TFs especially when the motifs were derived from *Drosophila*, supporting the specificity of the predictions. An additional criterion strengthening the confidence on predicted links could be conservation, especially since we found that, as a general rule, motif composition was not particularly conserved when DARs from paralogous genes were compared. We tested this hypothesis for a new predicted regulatory link between *dac* and *lz*. Our results, including loss and gain-of-function experiments, showed that indeed *lz* was a *dac* regulator in *Drosophila*, confirming the predicted regulatory relationship. The fact that *Lz* binding sites were found on *dac* DARs suggested that this regulation is direct. We think it likely that this regulatory link also exists in *Episyrphus*. But how generalizable are these GRN predictions? Testing this sort of hypotheses (that is, whether predicted regulatory links and indeed functional) will require further genetic tool-development in *Episyrphus*—as these should allow manipulating gene function post-embryonically. Even in *Drosophila*, a systematic testing of these hypothetical links, being these tests biochemical (e.g., chromatin immunoprecipitation) or genetic (e.g., mutational analysis, RNA interference), would not be straightforward. In other words, a systematic functional analysis of a GRN might not ever be practical and always remain a prediction, at least partially. It is likely that deep learning methods using large collections of data may be used to increase the probability of predicting true positive links. However, we foresee that GRN models will most likely be tested not at the level of individual links, but by comparing the global effects of specific perturbations *in vivo* and *in silico*, such that good agreement between biological and computational predictions would validate the GRN model as a whole.

Materials and methods

Episyrphus balteatus culture and ommatidial number estimate

Episyrphus larvae and adults are reared in different rearing cages. To feed the larvae, aphids are cultured on broad bean plants in an independent cage. All these cultures were maintained at a temperature of 22 ± 1 °C.

Adults are fed ad libitum with two types of food simultaneously: pollen crushed in a kitchen grinder, spread on paper; and cotton wool soaked in honey dissolved in water. Then there is the egg and larval cage. Larvae feed mainly on aphid-infested bean plants. The broad beans are sown every 2 days, and the aphids from the older plants are then transferred to the newly sprouted plants, which are most easily colonized by aphids. For egg-laying, we place one of these plants in the adults' cage. After a short time, females will start laying eggs in the places where aphids are concentrated. After a variable time, the plant is removed from the adult cage and placed in a separate one, which is the specific larval cage. This is the time 0 used to measure age, in hours after egg-laying (hAEL). At 36–48 hPEL, the eggs hatch and the larvae start feeding on the aphids. Then, every 48 hours, the larvae are transferred to new, aphid-infested plants. After about 260 hours (10–11 days), the larvae have reached the end of the larval stage and form pupae, usually outside the feeding area. The pupae are collected and placed in a petri dish, which in turn is placed in the adult cage.

To estimate the approximate number of ommatidia of the *Episyrphus* eye, we measured the area occupied by 30 ommatidia in a central region of the eye and then extrapolated the number of ommatidia to the total area, measured on a picture from a lateral view of a male eye, using the Fiji “polygon selection” tool. This figure, 3660 ommatidia, is likely an underestimation, due to the eye's curvature. In addition, it did not take into account potential ommatidial size variation in different regions of the eye.

Genomes

Drosophila melanogaster genome version dm6 [101]. *Episyrphus balteatus* genome and annotation [28].

RNA sequencing and bioinformatic processing

Experimental design. RNA sequencing (RNA-seq) experiments were conducted on *Drosophila melanogaster* (DROMEL) and *Episyrphus balteatus* (EPIBAL) to investigate gene expression differences in eye and wing imaginal discs. EPIBAL cultures were raised as indicated above and DROMEL at 25 °C under standard conditions. To stage the discs such that tissues were at an equivalent developmental stage in both species, we used the morphogenetic furrow (MF) position as a marker of differentiation progression. A stage in which the MF had progressed midway along the anterior-posterior axis was chosen, as in these discs an equivalent representation of progenitor/precursor cells (anterior to the MF) and differentiating retina (posterior to the MF) was included for both species. This developmental stage is reached at 96 hours after laying (hAEL) in DROMEL and at 120 hAEL in EPIBAL in our culture conditions.

In DROMEL, samples were collected at two developmental time points: 96 and 114 hAEL, with two and three biological replicates per tissue, respectively. In EPIBAL, two biological replicates per tissue were collected at 120 hAEL. RNA sequencing was performed using the Illumina NovaSeq 6000 platform with a read length of 51 base pairs (bp) in a single-end mode. Libraries were prepared using the Illumina TruSeq RNA Sample Preparation Kit.

Sample collection and RNA extraction. For each condition, 50 eye or wing imaginal discs were dissected from *Drosophila melanogaster* and *Episyrphus balteatus* larvae at the specified developmental time points. Total RNA was extracted using the RNeasy Kit (Qiagen) following the manufacturer's instructions. RNA quality and concentration were assessed using a Bioanalyzer (Agilent Technologies) before proceeding with library preparation. No discrimination between male and female larvae was carried out.

Library preparation and sequencing. RNA libraries were prepared according to the Illumina TruSeq RNA protocol. The resulting libraries were sequenced on the Illumina NovaSeq 6000 system in a single-end mode with a read length of 51 bp. The sequencing output was quality-checked using FastQC.

Bioinformatic processing. Raw sequencing reads were processed for quality control and filtering using Trimmomatic and Fastp to remove low-quality bases and adapter sequences. Filtered reads were aligned to the reference genome using HISAT2. The resulting alignments were used to generate read count matrices with HTSeq2. Differential gene expression analysis was conducted using DESeq2 to identify genes with significant expression differences between conditions.

Additional RNA-seq data acquisition. To complement our experimental data, additional RNA-seq datasets were retrieved from the Gene Expression Omnibus (GEO) repository under accession number GSE59707. These datasets included RNA-seq samples from eye-antennal and wing imaginal discs of wild-type *Drosophila melanogaster* strains (RAL-208 and Canton-S). Raw sequencing data were accessed from the Sequence Read Archive (SRA) under the study identifier SRP044761.

Processing of external datasets. The external RNA-seq datasets underwent the same bioinformatic pipeline as our experimental samples. Reads were filtered using Trimmomatic and Fastp, aligned to the dm6 genome with HISAT2, and read counts were generated using HTSeq2. Differential expression analysis was performed using DESeq2 to ensure consistency with our experimental analyses.

ATAC sequencing and bioinformatic processing

Experimental design. ATAC sequencing (ATAC-seq) experiments were conducted on *Episyrphus balteatus* (EPIBAL) to investigate chromatin accessibility differences in eye and wing imaginal discs. EPIBAL cultures were raised at 25 °C under standard conditions. To ensure an equivalent developmental stage, we used the morphogenetic furrow (MF) position as a marker of differentiation progress. Larvae were dissected when the MF had divided the disc into two halves, indicating that 50% of progenitors had already differentiated into photoreceptors. This developmental stage is reached at 120 hours after laying in EPIBAL.

ATAC-seq samples were collected at 120 hours after egg laying, with two biological replicates per tissue. Sequencing was performed using the Illumina NovaSeq 6000 platform in paired-end mode with read lengths of 49/25 bp. Libraries were prepared using the Nextera DNA Library Prep Kit.

Sample collection and nucleic acid extraction. For each condition, 7 eye or 21 wing imaginal discs were dissected from *Episyrphus balteatus* larvae at the specified developmental time point. Total chromatin was extracted following the standard ATAC-seq protocol.

Library preparation and sequencing. ATAC-seq libraries were prepared using the Nextera DNA Library Prep Kit. The resulting libraries were sequenced on the Illumina NovaSeq 6000 system in paired-end mode with read lengths of 49/25 bp. The sequencing output was quality-checked using FastQC.

Bioinformatic processing. Raw sequencing reads were processed for quality control and filtering using Trimmomatic and Fastp to remove low-quality bases and adapter sequences. Filtered reads were aligned to the reference genome using HISAT2. Peak calling was performed using MACS2 (Model-based Analysis of ChIP-Seq). Each tissue conservative peaks were selected using the Irreproducible Discovery Rate (IDR) method, with a global IDR threshold of ≥ 1 . Then, the tissue conservative peaks were merged. Read count matrices for chromatin accessibility were obtained using HTSeq2, and differential accessibility analysis was conducted using DESeq2.

Additional ATAC-seq data acquisition. Additionally, DROMEL ATAC-seq experimental data were retrieved from the GEO repository (GSE101827 series). *Drosophila melanogaster* cultures and sampling conditions were consistent and compatible with our omics experiments. The dataset included samples GSM2716860–GSM2716867. Raw data were retrieved from the SRA repository (SRP113497).

Processing of external datasets. The external RNA-seq datasets underwent a very similar bioinformatic pipeline as our experimental samples. Raw sequencing reads were processed for quality control and filtering using Trimmomatic and Fastp to remove low-quality bases and adapter sequences. Filtered reads were aligned to the dm6 genome using HISAT2. Peak calling was performed using MACS2 (Model-based Analysis of ChIP-Seq). Each tissue sample was merged and split into two pseudoreplicates. Subsequently, conservative peaks were selected using the Irreproducible Discovery Rate (IDR) method, applying a global IDR threshold of ≥ 1 . Then, the tissue conservative peaks were merged. Read count matrices for chromatin accessibility were obtained using HTSeq2, and differential accessibility analysis was conducted using DESeq2.

DARs annotation. We used the *annotatePeaks.pl* script from the HOMER software [54] to annotate DARs to the nearest Transcription Start Site (TSS) within a 1 Mb range. The annotations were then filtered to retain only genes with expression levels above a threshold, determined by the median FPKM values across species and tissue samples (see [S4 File Methods](#)).

Motif search. Different motif search methods were discussed ([S1 File Methods](#) before agreeing to adapt the HOMER method [54] to an expanded collection of motifs (PWMs) derived from three general databases: HOMER, JASPAR24 [102], and SCENIC+ [103] ([S2 File Methods](#)). The adapted method involves predicting the binding prediction thresholds used by HOMER (score_0) for the expanded collection ([S3 File Methods](#)). The predicted score_0 for each PWM was used as a motif detection threshold on the set of DARs from the eye and wing in both EPIBAL and DROMEL. To this end, the “*annotatePeaks.pl*” function of HOMER was executed. Finally, the HOMER motif search method applied to the expanded collection was evaluated based on the results obtained from the eye and wing DARs in both species ([S4 File Methods](#)), revealing differential connectivity of TFs in the eye networks of both species and an eye/wing enrichment of motifs corresponding to genes expressed in the eye of each species.

DARs clustering. Both the DARs and the GDARs, obtained by merging the former for each gene, were classified based on the Jaccard distance of their motif composition, following a hierarchical clustering algorithm ([S5 File Methods](#)). The classification results for each subset of motifs used are included in [S5 File Methods](#).

DAR-GAL4 construct generation and transgenesis

Genomic DNA extraction was carried out according to the protocol obtained from the Vienna Drosophila Resource Center (VDRC) at Vienna BioCenter Core Facilities (VBCF). For *E. balteatus* DNA extraction, three random female adults were picked, and after wings and abdomen removal the rest of the protocol was followed as-is. The protocol can be found at <https://shop.vbc.ac.at/media/pdfs/GoodQualityGenomicDNA.pdf>.

For cloning, we performed PCR on the gDNA synthesized using the Expand High Fidelity PCR System (Roche) and the amplified band cloned into the pBPGUw plasmid (Adgene plasmid #17575) by Gateway Cloning, following the provided protocols as-is. These DAR containing plasmid were chemically transformed into competent Match1 *E. coli* strain by standard methods and verified by Sanger sequencing. Once verified, transgenes were generated by site-directed transgenesis by FlyOrf (Zurich, Switzerland).

Primers used for the amplification of *Episyrphus dac*_DAR1:

Epi_dac_1_fw	ACTGATTTCCTTGAAGTT
Epi_dac_1_rv	GAATATTTGACATACAATGG

Cytoscape network models

To compare the EPIBAL and DROMEL RDGRNs, we used the network visualization tool Cytoscape [104]. Cytoscape generates a regulatory network by mapping all connections between the source (TFs) and the targets (any gene regulated through its CREs/DARs), while also allowing us to highlight the properties of these connections or the nodes themselves.

For a more comprehensive analysis, the species-specific networks were then superimposed, creating a new network that distinguishes between common and species-specific nodes and connections. Additional information determining the confidence of each motif supporting a connection (S4 File Methods) was also incorporated, including the origin of the motifs (DROMEL or Human/Mouse), predictability (estimable or inestimable), and tissue enrichment (eye-enriched or non-significant).

Furthermore, to synthesize the biological and regulatory information, all motifs associated with the same TF were condensed into a single representation. To retain the relevant confidence information for these connections, we adopted a criterion that always considers the ‘best’ properties of the set of motifs associated to the same TF in each species: DROMEL, estimable, and eye-enriched.

These visualizations allow us to observe the RDGRNs of both species on a broad scale or, alternatively, to explore the differential connectivity of specific genes or gene groups, providing valuable insights for generating hypotheses about eye regulation in dipterans.

Drosophila genetics

The following stocks were used: *UAS-Iz* (FBst0033836), *Iz:GFP* (FBst0043954), *Iz77a7* (FBst0003609) and 25kb *sna:GFP* (FBst0093875) are described in Flybase (<https://flybase.org/>), and *optix2/3-GAL4* in [69,85]. Larvae from the *optix2/3-GAL4* females to *UAS-Iz* males cross, raised at 25 °C, were used to analyze the effects of *Iz* overexpression on *dac* expression. The 25kb *sna:GFP* transgene rescues *snail* mutants to viability [105].

Immunostaining and imaging of *Drosophila* discs

For immunostaining and confocal microscopy, *Drosophila* eye discs were dissected from late third instar larvae in cold PBS and processed essentially as in [106]. Primary antibodies used were: mouse anti-Dac 1–1 (1/50) and rat anti-Elav 7EBA10 (1/1000), both from the Developmental Studies Hybridoma Bank developed under the auspices of the NICHD and maintained by The University of Iowa; rabbit anti-GFP (A11122, Molecular Probes) 1/1000. Appropriate fluorescent Alexa-conjugated antibodies were used at a 1/500 dilution to detect the primary antibodies. In the experiments aims at detecting *sna:GFP* expression, nuclei were counterstained with DAPI. Imaging was carried out on a Leica SPE confocal setup. Images were processed with Fiji [107].

In situ hybridization for *E. balteatus* glass

Identification of the *E. balteatus* glass (*gl*) and probe generation. The first step was to identify the orthologous *E. balteatus* glass (*gl*) gene. To do this, we ran a tblastn of the *D. melanogaster* *gl* amino acid sequence (Uniprot: P13360 - GLAS_DROME) against the *E. balteatus* genome. The result gave the highest overall score for a gene annotated as *gl_2_1*. We extracted the predicted CDS sequences of this gene, spliced them and translated them into different open reading frames (ORF) using the Expasy Translate tool. We took the ORF which was the size of all the CDSs combined together and applied Motif Scan motif prediction to this amino acid sequence. The predicted protein was very similar to *D. melanogaster* *gl*, with a series of C2H2 zinc finger motifs at its C-terminal end. This suggests that *gl_2_1* is the *E. balteatus* ortholog of *D. melanogaster* *gl*.

Next, we generated the *gl* sense and antisense probes (*gl_S* and *gl_AS*, respectively) by double PCR. For this purpose, we designed candidate primers flanking several CDS of the gene, specifically exons 3–5. These exons are:

Forward (fw): GGCCGCGGCGGCTGTCTATGTTGCCAGGGGA

Reverse (rv): GCCCCGCGCGGTTGCATGTTGAAGGAGCA

The resulting PCR product was of ~400 bp. The underlined regions correspond to a universal sequence that was used to generate a cDNA containing our probe with the T7 transcriptase binding site. The non-underlined regions are those that would align during the PCR: in a first PCR, they amplified the *gl* cDNAs of retrotranscribed cDNA from larval plus adult *E. balteatus* tissues. In a second PCR, we combined our glass primers with universal primers to generate the sequences to be transcribed into *gl_S* and *gl_AS*. These 5' and 3' universal primers contain the universal sequence complementary to our *gl* forward and reverse primers, respectively, plus a T7 RNA polymerase binding site. Finally, the product of the second PCR was transcribed using a New England biolabs T7 kit (M0251S) and DIG-labelled nucleotides, then precipitated, purified and dissolved in double-distilled water with the addition of 0.5 microlitres of RNAase inhibitor. The quality and integrity of the probes were confirmed by running them on a 2% agarose gel at 120V for 15 min, observing a single band of appropriate size and without smears, and measuring the concentrations in a Nanodrop. The products showed an absorbance curve suitable for a purified RNA. Finally, the probes were diluted in 1ml final volume of hybridization buffer (sol. A) and stored at -20°C.

In situ protocol

The *in situ* hybridization protocol used is a variant of the standard alkaline phosphatase (AP) protocol for *Drosophila* imaginal discs [108]. Unless otherwise stated, all washes were 5 min at RT. The major key modifications were the dissection procedure of *Episyrphus* larvae and during the Pre-hybridization step, which we describe in what follows:

Episyrphus dissection. Third larval stage larvae are collected in wells containing 1 × PBS on ice. After a few minutes, when completely still, they are dissected with dissecting forceps, cut just anterior to the level where the posterior fat body is seen under the cuticle. Then, this anterior “cone” is grasped by the mandibles and turned inside-out to expose the imaginal discs and other internal tissues. The posterior part of the larva may or may not be discarded, depending on whether the genital disc is required for sexing each individual. Alternatively, the larva can be partially opened, leaving the anterior and posterior parts together, and both turned inside out. In this way, the genital disc and the rest of the imaginal discs of the same individual can be kept together when sexing is required, and several larvae can be worked with in the same well. After turning inside-out, the digestive tube is removed and the larvae are fixed in 4% formaldehyde in PBS for 25 min at room temperature (RT).

Pre-hybridization. After fixation, the samples are washed in PTwx (1x PBS with Triton X-100 and 0.1% Tween 20 each) and twice in ethanol. They are then incubated in 1:1 v/v xylene/ethanol solution for 1 h at RT. This is followed by

2 washes in 100% ethanol. Then, the tissues are washed in serial dilutions of 80, 50 and 25% methanol/ddH₂O and finally washed 2 times in ddH₂O. At this point, the discs are incubated in 80% acetone in ddH₂O at -20°C for variable times depending on the age of the larvae (measured as hours (h) after egg laying, hAEL): a) <170hAEL = 10 min, b) 170–200hAEL = 15 min, c) 200–250hAEL = 20 min. After this incubation, the samples are washed 2 times in PTwx, postfixed in 4% formaldehyde for 25 min at RT and washed again 2 times in PTwx.

Supporting information

S1 Fig. Browser view of the *Drosophila bilomb* locus. Eye (A) and wing (B) ATACseq tracks are shown above and below it, respectively. Eye-wing differential DARs are indicated as boxes under their respective tracks. While *bilomb* is not detected as differentially expressed between eye and wing discs, it is linked to eye- and wing-specific DARs indicating this gene is subject to organ-specific regulation.

(TIFF)

S2 Fig. (Supplement To Fig 2). GO analysis (Biological Process, BP) of the eye GDARs in *Episyrphus* (A) and *Drosophila* (B) and of the intersection with the DEG set. Top 20 BP terms are shown.

(TIFF)

S3 Fig. (Supplement to Fig 2). SCOPE visualization of the single cell expression of the paralogous genes *ems* (A) and *E5* (B) and *in situ* expression of the *sna*:GFP protein reporter (C-E). The genes *Optix* and *gl* are included to mark the eye progenitor cells (1) and the differentiating retina (2), respectively. *ems* and *E5* are expressed in the prospective head (3) and antenna (4), which are also marked on the cell atlas. (C-E) Eye-antennal (C), leg (D) and wing (D) imaginal discs from *Drosophila* 25kb *Sna*:GFP L3 larvae, stained for anti-GFP (green). *Sna*:GFP-positive nuclei are only seen in the eye disc (C, arrows). A higher magnification view of the boxed region in (C) is shown in (C'), counterstained with DAPI to visualize the nuclei. The position of the morphogenetic furrow is marked by the dashed line. *Sna*:GFP-positive nuclei are detected posterior to the furrow. Note that this expression is transient, as more posterior cells no longer express *Sna*:GFP. The scale bar is 50µm in all panels.

(TIFF)

S4 Fig. Genome browser view of the *Drosophila Optix* (A) and (B) B-H2/B-H1 loci. (A). First two tracks, eye disc and wing disc transcription: Below, eye disc and wing disc ATAC-seq tracks. DARs are highlighted. Enhancer activity of the two Flylight lines indicated, as marked with GFP (green; marked by the arrowheads). Above, eye discs; below, wing disc. The discs are also stained for Dll (red) and counterstained with the nuclear marker DAPI (blue). (B) B-H2/B-H1 loci. Eye ATAC-seq track from genome browser. DARs are highlighted. Above, Flylight images of enhancer activity driven by different regions (green bars). Of the 12 DARs detected, 10 fall within tested regions. Except for 81D12 (highlighted in magenta), for which no imaginal disc enhancer activity is reported (and for which the ATAC peak might derived from the fact that it coincides with an annotated long non-coding RNA, lncRNA:CR43491/NR_047828), all fall within regions with eye disc enhancer activity. Two DARs lie in regions not covered by the Flylight fragments (highlighted in grey). As in (A), the antennal region of the eye-imaginal disc is marked in red. CE: compound eye; OC: ocellar region; PE: peripodial epithelial layer; WP: wing pouch.

(TIFF)

S5 Fig. Pipeline for filtering and integrating the motif collections used. The filtered STEIN collection (yellow, top left) is merged with JASPAR24 (orange) and the manually inserted PWMs (red). The HOMER collection is filtered twice (cyan, top right) before calculating the intersection with the other collections. The subsets derived from the intersection (cylinders) are classified according to the origin of the annotations, direct (DROMEL) or indirect (Hum/Mou). The common

PWMs (green cylinders) are categorized into direct annotations in the STEIN+JASPAR24 collection (SJ DROMEL), direct annotations in the HOMER collection (HOMER DROMEL), and indirect annotations in the STEIN+JASPAR24 collection (Hum/Mou).

(TIFF)

S6 Fig. (A) TF motif enrichment in *Drosophila* DARs of eye versus wing, for different subsets of motifs. *Drosophila* estimable motifs were most enriched. (B) The number of eye DARs linked to TF (dark green or dark pink) is larger than the number associated to non-TF genes (“targets”; light green or light pink) both in *Episyrphus* (EPIBAL) and *Drosophila* (DROMEL). Note, for example, that the dark bars corresponding to 2, 3 and 4 DARs are higher than the light ones in EPIBAL. (C) Similarly, the number of unique TF predicted to bind to an eye gene (within the sum of all its DARs) is also larger for TFs than for other genes in both species. This can be seen as the dark curve lies above the light one for a number of inputs higher than 100 in EPIBAL or 150 in DROMEL.

(TIFF)

S7 Fig. RDGRNs visualization in *Episyrphus* and *Drosophila*. (A) Heatmap representing the distribution of kernel TF motifs (columns) in the core GDARs set of both species (rows). The colour intensity in the heatmap is inversely proportional to the frequency with which each motif is found in the set of eye-GDARs for each species and can be interpreted as an indication of how much more atypical a motif is compared to a ubiquitous one. (B) Cytoscape view of the kernel TF GRN. Links are directed, with yellow arrowheads pointing to the regulated targets. DROMEL-specific links, pink; EPIBAL-specific links, green; Shared links, grey. Line type: *solid*: DROMEL motifs; *dashed*: human and mouse motifs; Increasing Line *Transparency*: estimable to non-estimable; Increasing Line *Width*: non-enriched to enriched in eye discs. (B') The shared links-only network. The degree of shared connectivity is not homogeneous.

(TIFF)

S8 Fig. (Supplement to Fig 3). SCOPE visualization of single cell mRNA abundance. (1), (2), (3) and (4) indicate eye progenitors/precursors, differentiating retina, prospective head capsule and antenna, respectively. *gl* expression is represented in green. green/red overlap: yellow. green/blue overlap: cyan. red/blue overlap: magenta. red/green/blue overlap: white.

(TIFF)

S9 Fig. A sample of paralogous loci indicating the number, location and similarity of linked DARs. In the case of *Iz*, the Flylight project identified two eye enhancer fragments (boxes) in the *Drosophila* locus. Our ATAC-seq data seems to have missed one of them (marked as “?”).

(TIFF)

S10 Fig. Conservation curve of the regulatory landscape between orthologous gene pairs. The empirical curve (dotted line) represents the decrease in the number of orthologous pairs clustered together (hierarchical clustering) as a function of the number of selected clusters. The gene set was clustered based on the distances between genes (Jaccard distances), calculated from the number of distinct motifs in the kernel of 20 TFs found in any of the DARs that make up the regulatory landscape of each gene. To establish the random distribution of conservation curves (shaded area), each gene was randomly assigned the regulatory landscape of another gene prior to clustering. This process was repeated 100 times, selecting the 95 curves (95% confidence) closest to the mean random curve (solid line). The gene core network shows a higher degree of conservation of regulation across species (i.e. the original data curve lies above the confidence area) than would be expected by chance.

(TIFF)

S1 Table. DROMEL eye GDARs. List of *Drosophila* eye genes with associated eye DARs.
(XLSX)

S2 Table. DROMEL eye DEGs. List of *Drosophila* genes differentially expressed in the eye versus the wing imaginal discs.
(XLSX)

S3 Table. EPIBAL eye GDARs. List of *Episyrphus* eye genes with associated eye DARs.
(XLSX)

S4 Table. EPIBAL eye DEGs. List of *Episyrphus* genes differentially expressed in the eye versus the wing imaginal discs.
(XLSX)

S5 Table. 106 core similarity. Jaccard and Similarity scores computed for the eyeDARs of the 106 eye genes shared between *Drosophila* and *Episyrphus* ("core").
(XLSX)

S6 Table. 22 kernel similarity. Jaccard and Similarity scores computed for the eyeDARs of the 22 transcription factor-encoding genes shared between *Drosophila* and *Episyrphus* ("kernel").
(XLSX)

S1 File. Methods. Comparison of motive search methods.
(PDF)

S2 File. Methods. Motif collection description.
(PDF)

S3 File. Methods. Inference of the PWM score thresholds (score_0) for the detection of DNA-binding of the associated TFs.
(PDF)

S4 File. Methods. Validations on inferred GRNs.
(PDF)

S5 File. Methods. Gene expression thresholding based on the distribution of median FPKMs of each gene across all samples.
(PDF)

Acknowledgments

We thank M. Irimia (GRG/ICREA Barcelona) for help with the annotation and orthology predictions of the *Episyrphus balteatus* genome version used in this paper. Also, FC acknowledges institutional funding to CABD through CSIC MaX4-SOMMA- 202620E045. Imaging was carried out at ALMIA, the CABD's Advanced Light Microscopy and Image Analysis platform. We thank Angelike Stathopoulos (CALTECH) for information on the 25kb sna:GFP transgenic line.

Author contributions

Conceptualization: Fernando Casares.

Data curation: Tomás Navarro, Karl Wotton, Fernando Casares.

Formal analysis: Tomás Navarro.

Funding acquisition: Fernando Casares.

Investigation: Tomás Navarro, Jesús Torres, Reinaldo Sáez-Moreno, Genoveva Guerrero, Isabel Almudí, Antonella Iannini, Javier Figueras, Fernando Casares.

Methodology: Tomás Navarro, Jesús Torres, Reinaldo Sáez-Moreno, Genoveva Guerrero, Isabel Almudí, Javier Figueras.

Project administration: Antonella Iannini, Fernando Casares.

Resources: Stein Aerts.

Software: Tomás Navarro.

Supervision: Stein Aerts, Fernando Casares.

Visualization: Tomás Navarro, Jesús Torres, Reinaldo Sáez-Moreno, Genoveva Guerrero, Isabel Almudí, Fernando Casares.

Writing – original draft: Tomás Navarro, Fernando Casares.

Writing – review & editing: Tomás Navarro, Karl Wotton, Stein Aerts, Fernando Casares.

References

1. Luittried S-P, Mayr E. On the Evolution of Photoreceptors and Eyes. *Evolutionary Biology*. Springer US. 1977. 207–63. https://doi.org/10.1007/978-1-4615-6953-4_4
2. Erclik T, Hartenstein V, McInnes RR, Lipshitz HD. Eye evolution at high resolution: the neuron as a unit of homology. *Dev Biol*. 2009;332(1):70–9. <https://doi.org/10.1016/j.ydbio.2009.05.565> PMID: [19467226](https://pubmed.ncbi.nlm.nih.gov/19467226/)
3. Gehring WJ, Ikeo K. Pax 6: mastering eye morphogenesis and eye evolution. *Trends Genet*. 1999;15(9):371–7. [https://doi.org/10.1016/s0168-9525\(99\)01776-x](https://doi.org/10.1016/s0168-9525(99)01776-x) PMID: [10461206](https://pubmed.ncbi.nlm.nih.gov/10461206/)
4. Kumar JP. The molecular circuitry governing retinal determination. *Biochim Biophys Acta*. 2009;1789(4):306–14. <https://doi.org/10.1016/j.bbagr.2008.10.001> PMID: [19013263](https://pubmed.ncbi.nlm.nih.gov/19013263/)
5. Bravo González-Blas C, Quan X-J, Duran-Romana R, Taskiran II, Koldere D, Davie K, et al. Identification of genomic enhancers through spatial integration of single-cell transcriptomics and epigenomics. *Mol Syst Biol*. 2020;16(5):e9438. <https://doi.org/10.1525/msb.20209438> PMID: [32431014](https://pubmed.ncbi.nlm.nih.gov/32431014/)
6. Potier D, Davie K, Hulselmans G, Naval Sanchez M, Haagen L, Huynh-Thu VA, et al. Mapping gene regulatory networks in Drosophila eye development by large-scale transcriptome perturbations and motif inference. *Cell Rep*. 2014;9(6):2290–303. <https://doi.org/10.1016/j.celrep.2014.11.038> PMID: [25533349](https://pubmed.ncbi.nlm.nih.gov/25533349/)
7. Sanchez-Aragon M, Cantisan-Gomez J, Luque CM, Bras-Pereira C, Lopes CS, Lemos Fernandez MC, et al. A Toggle-Switch and a Feed-Forward Loop Engage in the Control of the Drosophila Retinal Determination Gene Network. *Front. Ecol. Evol*. 2019.
8. Casares F, Almudí I. Fast and Furious 800. The Retinal Determination Gene Network in Drosophila. *Organogenetic Gene Networks*. Springer International Publishing. 2016. 95–124. https://doi.org/10.1007/978-3-319-42767-6_4
9. Chen Y-C, Desplan C. Gene regulatory networks during the development of the Drosophila visual system. *Curr Top Dev Biol*. 2020;139:89–125. <https://doi.org/10.1016/bs.ctdb.2020.02.010> PMID: [32450970](https://pubmed.ncbi.nlm.nih.gov/32450970/)
10. Halder G, Callaerts P, Gehring WJ. Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science*. 1995;267(5205):1788–92. <https://doi.org/10.1126/science.7892602> PMID: [7892602](https://pubmed.ncbi.nlm.nih.gov/7892602/)
11. Kumar JP. Retinal determination the beginning of eye development. *Curr Top Dev Biol*. 2010;93:1–28. <https://doi.org/10.1016/B978-0-12-385044-7.00001-1> PMID: [20959161](https://pubmed.ncbi.nlm.nih.gov/20959161/)
12. Silver SJ, Rebay I. Signaling circuitries in development: insights from the retinal determination gene network. *Development*. 2005;132(1):3–13. <https://doi.org/10.1242/dev.01539> PMID: [15590745](https://pubmed.ncbi.nlm.nih.gov/15590745/)
13. Treisman JE. Retinal differentiation in Drosophila. *Wiley interdisciplinary reviews Developmental biology*. 2013;2:545–57.
14. Bushbeck EK, Friedrich M. Evolution of insect eyes: tales of ancient heritage, deconstruction, reconstruction, remodeling, and recycling. *Evolution: Education and Outreach*. 2008;1:448–62.
15. Friedrich M. Continuity versus split and reconstitution: exploring the molecular developmental corollaries of insect eye primordium evolution. *Dev Biol*. 2006;299(2):310–29. <https://doi.org/10.1016/j.ydbio.2006.08.027> PMID: [16973149](https://pubmed.ncbi.nlm.nih.gov/16973149/)
16. Kittelmann M, McGregor AP. Looking across the gap: Understanding the evolution of eyes and vision among insects. *Bioessays*. 2024;46(5):e2300240. <https://doi.org/10.1002/bies.202300240> PMID: [38593308](https://pubmed.ncbi.nlm.nih.gov/38593308/)

17. Yang X, Weber M, Zarinkamar N, Posnien N, Friedrich F, Wigand B, et al. Probing the *Drosophila* retinal determination gene network in *Tribolium* (II): The Pax6 genes *eyeless* and *twin of eyeless*. *Dev Biol*. 2009;333(1):215–27. <https://doi.org/10.1016/j.ydbio.2009.06.013> PMID: 19527703
18. Yang X, Zarinkamar N, Bao R, Friedrich M. Probing the *Drosophila* retinal determination gene network in *Tribolium* (I): The early retinal genes *dachshund*, *eyes absent* and *sine oculis*. *Dev Biol*. 2009;333(1):202–14. <https://doi.org/10.1016/j.ydbio.2009.02.040> PMID: 19324029
19. Hu X, Cheng F, Gong Z, Qin K, Shan T, Li W, et al. Knockout of a single Pax6 gene (*toy* but not *ey*) leads to compound eye deficiency and small head in honeybees. *Commun Biol*. 2024;7(1):1319. <https://doi.org/10.1038/s42003-024-07016-5> PMID: 39402171
20. Takagi A, Kurita K, Terasawa T, Nakamura T, Bando T, Moriyama Y, et al. Functional analysis of the role of *eyes absent* and *sine oculis* in the developing eye of the cricket *Gryllus bimaculatus*. *Dev Growth Differ*. 2012;54(2):227–40. <https://doi.org/10.1111/j.1440-169X.2011.01325.x> PMID: 22348272
21. Dong Y, Friedrich M. Enforcing biphasic eye development in a directly developing insect by transient knockdown of single eye selector genes. *J Exp Zool B Mol Dev Evol*. 2010;314(2):104–14. <https://doi.org/10.1002/jez.b.21313> PMID: 19637278
22. Zarinkamar N, Yang X, Bao R, Friedrich F, Beutel R, Friedrich M. The Pax gene *eyegone* facilitates repression of eye development in *Tribolium*. *EvoDevo*. 2011;2(1):8. <https://doi.org/10.1186/2041-9139-2-8> PMID: 21463500
23. Ready DF. A multifaceted approach to neural development. *Trends Neurosci*. 1989;12(3):102–10. [https://doi.org/10.1016/0166-2236\(89\)90166-5](https://doi.org/10.1016/0166-2236(89)90166-5) PMID: 2469216
24. Casares F, McGregor AP. The evolution and development of eye size in flies. *Wiley Interdisciplinary Reviews Developmental Biology*. 2020;e380.
25. Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim J-W, Lambkin C, et al. Episodic radiations in the fly tree of life. *Proc Natl Acad Sci U S A*. 2011;108(14):5690–5. <https://doi.org/10.1073/pnas.1012675108> PMID: 21402926
26. Omkar O, Mishra G. Syrphid Flies (The Hovering Agents). *Ecofriendly Pest Management for Food Security*. Elsevier. 2016. 259–79. <https://doi.org/10.1016/b978-0-12-803265-7.00008-7>
27. Wotton KR, Gao B, Menz MHM, Morris RKA, Ball SG, Lim KS, et al. Mass Seasonal Migrations of Hoverflies Provide Extensive Pollination and Crop Protection Services. *Curr Biol*. 2019;29(13):2167–2173.e5. <https://doi.org/10.1016/j.cub.2019.05.036> PMID: 31204159
28. Doyle T, Jimenez-Guri E, Hawkes WLS, Massy R, Mantica F, Permanyer J, et al. Genome-wide transcriptomic changes reveal the genetic pathways involved in insect migration. *Mol Ecol*. 2022;31(16):4332–50. <https://doi.org/10.1111/mec.16588> PMID: 35801824
29. Hawkes WL, Sivell O, Wotton KR. The genome sequence of the Marmalade Hoverfly, *Episyrphus balteatus* (De Geer, 1776). *Wellcome Open Res*. 2023;8:106. <https://doi.org/10.12688/wellcomeopenres.19073.1>
30. Haynie JL, Bryant PJ. Development of the eye-antenna imaginal disc and morphogenesis of the adult head in *Drosophila melanogaster*. *J Exp Zool*. 1986;237(3):293–308. <https://doi.org/10.1002/jez.1402370302> PMID: 3084703
31. Wolff T, Ready DF. The beginning of pattern formation in the *Drosophila* compound eye: the morphogenetic furrow and the second mitotic wave. *Development*. 1991;113(3):841–50. <https://doi.org/10.1242/dev.113.3.841> PMID: 1726564
32. Arora I, Tollefsbol TO. Computational methods and next-generation sequencing approaches to analyze epigenetics data: Profiling of methods and applications. *Methods*. 2021;187:92–103. <https://doi.org/10.1016/j.ymeth.2020.09.008> PMID: 32941995
33. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013;10(12):1213–8. <https://doi.org/10.1038/nmeth.2688> PMID: 24097267
34. del Alamo Rodríguez D, Terriente Felix J, Díaz-Benjumea FJ. The role of the T-box gene *optomotor-blind* in patterning the *Drosophila* wing. *Dev Biol*. 2004;268(2):481–92. <https://doi.org/10.1016/j.ydbio.2004.01.005> PMID: 15063183
35. Tsai Y-C, Grimm S, Chao J-L, Wang S-C, Hofmeyer K, Shen J, et al. *Optomotor-blind* negatively regulates *Drosophila* eye development by blocking Jak/STAT signaling. *PLoS One*. 2015;10(3):e0120236. <https://doi.org/10.1371/journal.pone.0120236> PMID: 25781970
36. Sato A, Tomlinson A. Dorsal-ventral midline signaling in the developing *Drosophila* eye. *Development*. 2007;134(4):659–67. <https://doi.org/10.1242/dev.02786> PMID: 17215299
37. De Graeve F, Smaldone S, Laub F, Mlodzik M, Bhat M, Ramirez F. Identification of the *Drosophila* progenitor of mammalian Krüppel-like factors 6 and 7 and a determinant of fly development. *Gene*. 2003;314:55–62. [https://doi.org/10.1016/s0378-1119\(03\)00720-0](https://doi.org/10.1016/s0378-1119(03)00720-0) PMID: 14527717
38. Distefano GM, Gangemi AJ, Khandelwal PJ, Saunders AJ, Marendra DR. *Drosophila* *lilliputian* is required for proneural gene expression in retinal development. *Dev Dyn*. 2012;241(3):553–62. <https://doi.org/10.1002/dvdy.23738> PMID: 22275119
39. Moses K, Ellis MC, Rubin GM. The glass gene encodes a zinc-finger protein required by *Drosophila* photoreceptor cells. *Nature*. 1989;340(6234):531–6. <https://doi.org/10.1038/340531a0> PMID: 2770860
40. Pickup AT, Lamka ML, Sun Q, Yip MLR, Lipshitz HD. Control of photoreceptor cell morphology, planar polarity and epithelial integrity during *Drosophila* eye development. *Development*. 2002;129(9):2247–58. <https://doi.org/10.1242/dev.129.9.2247> PMID: 11959832
41. Daga A, Karlovich CA, Dumstrei K, Banerjee U. Patterning of cells in the *Drosophila* eye by *Lozenge*, which shares homologous domains with AML1. *Genes Dev*. 1996;10(10):1194–205. <https://doi.org/10.1101/gad.10.10.1194> PMID: 8675007
42. Higashijima S, Kojima T, Michiue T, Ishimaru S, Emori Y, Saigo K. Dual Bar homeo box genes of *Drosophila* required in two photoreceptor cells, R1 and R6, and primary pigment cells for normal eye development. *Genes Dev*. 1992;6(1):50–60. <https://doi.org/10.1101/gad.6.1.50> PMID: 1346120

43. Diederich RJ, Pattatucci AM, Kaufman TC. Developmental and evolutionary implications of labial, Deformed and engrailed expression in the *Drosophila* head. *Development*. 1991;113(1):273–81. <https://doi.org/10.1242/dev.113.1.273> PMID: [1684933](#)
44. Lee KJ, Freeman M, Steller H. Expression of the disconnected gene during development of *Drosophila melanogaster*. *EMBO J*. 1991;10(4):817–26. <https://doi.org/10.1002/j.1460-2075.1991.tb08014.x> PMID: [1901263](#)
45. Roignant J-Y, Legent K, Janody F, Treisman JE. The transcriptional co-factor Chip acts with LIM-homeodomain proteins to set the boundary of the eye field in *Drosophila*. *Development*. 2010;137(2):273–81. <https://doi.org/10.1242/dev.041244> PMID: [20040493](#)
46. Wieschaus E, Perrimon N, Finkelstein R. orthodenticle activity is required for the development of medial structures in the larval and adult epidermis of *Drosophila*. *Development*. 1992;115(3):801–11. <https://doi.org/10.1242/dev.115.3.801> PMID: [1425355](#)
47. Bollepogu Raja KK, Yeung K, Shim Y-K, Li Y, Chen R, Mardon G. A single cell genomics atlas of the *Drosophila* larval eye reveals distinct photoreceptor developmental timelines. *Nat Commun*. 2023;14(1):7205. <https://doi.org/10.1038/s41467-023-43037-0> PMID: [37938573](#)
48. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res*. 2003;31(1):258–61. <https://doi.org/10.1093/nar/gkg034> PMID: [12519996](#)
49. Cusanovich DA, Reddington JP, Garfield DA, Daza RM, Aghamirzaie D, Marco-Ferreres R, et al. The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature*. 2018;555(7697):538–42. <https://doi.org/10.1038/nature25981> PMID: [29539636](#)
50. Janssens J, Aibar S, Taskiran II, Ismail JN, Gomez AE, Aughey G, et al. Decoding gene regulation in the fly brain. *Nature*. 2022;601(7894):630–6. <https://doi.org/10.1038/s41586-021-04262-z> PMID: [34987221](#)
51. Johansen NJ, Kempynck N, Zemke NR, Somasundaram S, Winter S, Hooper M, et al. Evaluating methods for the prediction of cell type-specific enhancers in the mammalian cortex. *bioRxiv*. 2025.
52. Thomas S, Li X-Y, Sabo PJ, Sandstrom R, Thurman RE, Canfield TK, et al. Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. *Genome Biol*. 2011;12(5):R43. <https://doi.org/10.1186/gb-2011-12-5-r43> PMID: [21569360](#)
53. Jenett A, Rubin GM, Ngo T-TB, Shepherd D, Murphy C, Dionne H, et al. A GAL4-driver line resource for *Drosophila* neurobiology. *Cell Rep*. 2012;2(4):991–1001. <https://doi.org/10.1016/j.celrep.2012.09.011> PMID: [23063364](#)
54. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576–89. <https://doi.org/10.1016/j.molcel.2010.05.004> PMID: [20513432](#)
55. Nitta KR, Jolma A, Yin Y, Morgunova E, Kivioja T, Akhtar J, et al. Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *Elife*. 2015;4:e04837. <https://doi.org/10.7554/eLife.04837> PMID: [25779349](#)
56. Marlétaz F, Firbas PN, Maeso I, Tena JJ, Bogdanovic O, Perry M, et al. Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature*. 2018;564(7734):64–70. <https://doi.org/10.1038/s41586-018-0734-6> PMID: [30464347](#)
57. Sorrells TR, Johnson AD. Making sense of transcription networks. *Cell*. 2015;161(4):714–23. <https://doi.org/10.1016/j.cell.2015.04.014> PMID: [25957680](#)
58. Bernardo-Garcia FJ, Fritsch C, Sprecher SG. The transcription factor Glass links eye field specification with photoreceptor differentiation in *Drosophila*. *Development*. 2016;143(8):1413–23. <https://doi.org/10.1242/dev.128801> PMID: [26952983](#)
59. Liang X, Mahato S, Hemmerich C, Zelhof AC. Two temporal functions of Glass: Ommatidium patterning and photoreceptor differentiation. *Dev Biol*. 2016;414(1):4–20. <https://doi.org/10.1016/j.ydbio.2016.04.012> PMID: [27105580](#)
60. Morrison CA, Chen H, Cook T, Brown S, Treisman JE. Glass promotes the differentiation of neuronal and non-neuronal cell types in the *Drosophila* eye. *PLoS Genet*. 2018;14(1):e1007173. <https://doi.org/10.1371/journal.pgen.1007173> PMID: [29324767](#)
61. Moses K, Rubin GM. Glass encodes a site-specific DNA-binding protein that is regulated in response to positional signals in the developing *Drosophila* eye. *Genes Dev*. 1991;5(4):583–93. <https://doi.org/10.1101/gad.5.4.583> PMID: [2010085](#)
62. Brás-Pereira C, Casares F, Janody F. The retinal determination gene Dachshund restricts cell proliferation by limiting the activity of the Homothorax-Yorkie complex. *Development*. 2015;142(8):1470–9. <https://doi.org/10.1242/dev.113340> PMID: [25790852](#)
63. Mardon G, Solomon NM, Rubin GM. dachshund encodes a nuclear protein required for normal eye and leg development in *Drosophila*. *Development*. 1994;120(12):3473–86. <https://doi.org/10.1242/dev.120.12.3473> PMID: [7821215](#)
64. Pappu KS, Ostrin EJ, Middlebrooks BW, Sili BT, Chen R, Atkins MR, et al. Dual regulation and redundant function of two eye-specific enhancers of the *Drosophila* retinal determination gene dachshund. *Development*. 2005;132(12):2895–905. <https://doi.org/10.1242/dev.01869> PMID: [15930118](#)
65. Li Y, Jiang Y, Chen Y, Karandikar U, Hoffman K, Chattopadhyay A, et al. optix functions as a link between the retinal determination network and the dpp pathway to control morphogenetic furrow progression in *Drosophila*. *Dev Biol*. 2013;381(1):50–61. <https://doi.org/10.1016/j.ydbio.2013.06.015> PMID: [23792115](#)
66. Seimiya M, Gehring WJ. The *Drosophila* homeobox gene optix is capable of inducing ectopic eyes by an eyeless-independent mechanism. *Development*. 2000;127(9):1879–86. <https://doi.org/10.1242/dev.127.9.1879> PMID: [10751176](#)
67. Flores GV, Daga A, Kalhor HR, Banerjee U. Lozenge is expressed in pluripotent precursor cells and patterns multiple cell types in the *Drosophila* eye through the control of cell-specific transcription factors. *Development*. 1998;125(18):3681–7. <https://doi.org/10.1242/dev.125.18.3681> PMID: [9716533](#)

68. Hayashi T, Saigo K. Diversification of cell types in the *Drosophila* eye by differential expression of prepattern genes. *Mech Dev*. 2001;108(1–2):13–27. [https://doi.org/10.1016/s0925-4773\(01\)00466-x](https://doi.org/10.1016/s0925-4773(01)00466-x) PMID: 11578858
69. Ostrin EJ, Li Y, Hoffman K, Liu J, Wang K, Zhang L, et al. Genome-wide identification of direct targets of the *Drosophila* retinal determination protein Eyeless. *Genome Res*. 2006;16(4):466–76. <https://doi.org/10.1101/gr.4673006> PMID: 16533912
70. Piñeiro C, Lopes CS, Casares F. A conserved transcriptional network regulates lamina development in the *Drosophila* visual system. *Development*. 2014;141(14):2838–47. <https://doi.org/10.1242/dev.108670> PMID: 24924198
71. Buchberger E, Bilen A, Ayaz S, Salamanca D, Matas de Las Heras C, Niksic A, et al. Variation in Pleiotropic Hub Gene Expression Is Associated with Interspecific Differences in Head Shape and Eye Size in *Drosophila*. *Mol Biol Evol*. 2021;38(5):1924–42. <https://doi.org/10.1093/molbev/msaa335> PMID: 33386848
72. Baudouin Gonzalez L, Schönauer A, Harper A, Arif S, Leite DJ, Steinhoff POM, et al. Development and patterning of a highly versatile visual system in spiders. *Proc Biol Sci*. 2025;292(2042):20242069. <https://doi.org/10.1098/rspb.2024.2069> PMID: 40068820
73. Baudouin-Gonzalez L, Harper A, McGregor AP, Sumner-Rooney L. Regulation of Eye Determination and Regionalization in the Spider *Parasteatoda tepidariorum*. *Cells*. 2022;11(4):631. <https://doi.org/10.3390/cells11040631> PMID: 35203282
74. Blackburn DC, Conley KW, Plachetzki DC, Kempler K, Battelle B-A, Brown NL. Isolation and expression of Pax6 and atonal homologues in the American horseshoe crab, *Limulus polyphemus*. *Dev Dyn*. 2008;237(8):2209–19. <https://doi.org/10.1002/dvdy.21634> PMID: 18651657
75. Prpic N-M. Duplicated Pax6 genes in *Glomeris marginata* (Myriapoda: Diplopoda), an arthropod with simple lateral eyes. *Zoology (Jena)*. 2005;108(1):47–53. <https://doi.org/10.1016/j.zool.2004.11.003> PMID: 16351954
76. Samadi L, Schmid A, Eriksson BJ. Differential expression of retinal determination genes in the principal and secondary eyes of *Cupiennius salei* Keyserling (1877). *EvoDevo*. 2015;6:16. <https://doi.org/10.1186/s13227-015-0010-x> PMID: 26034575
77. Schomburg C, Turetzek N, Schacht MI, Schneider J, Kirfel P, Prpic N-M, et al. Molecular characterization and embryonic origin of the eyes in the common house spider *Parasteatoda tepidariorum*. *EvoDevo*. 2015;6:15. <https://doi.org/10.1186/s13227-015-0011-9> PMID: 26034574
78. Oudelaar AM, Higgs DR. The relationship between genome structure and function. *Nat Rev Genet*. 2021;22(3):154–68. <https://doi.org/10.1038/s41576-020-00303-x> PMID: 33235358
79. Jacobs J, Atkins M, Davie K, Imrichova H, Romanelli L, Christiaens V, et al. The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. *Nat Genet*. 2018;50(7):1011–20. <https://doi.org/10.1038/s41588-018-0140-x> PMID: 29867222
80. Suryamohan K, Halfon MS. Identifying transcriptional cis-regulatory modules in animal genomes. *Wiley Interdiscip Rev Dev Biol*. 2015;4(2):59–84. <https://doi.org/10.1002/wdev.168> PMID: 25704908
81. De Winter S, Konstantakos V, Aerts S. Modelling and design of transcriptional enhancers. *Nat Rev Bioeng*. 2025;3(5):374–89. <https://doi.org/10.1038/s44222-025-00280-y>
82. Bessa J, Carmona L, Casares F. Zinc-finger paralogues *tsh* and *tio* are functionally equivalent during imaginal development in *Drosophila* and maintain their expression levels through auto- and cross-negative feedback loops. *Dev Dyn*. 2009;238(1):19–28. <https://doi.org/10.1002/dvdy.21808> PMID: 19097089
83. Bessa J, Gebelein B, Pichaud F, Casares F, Mann RS. Combinatorial control of *Drosophila* eye development by *eyeless*, *homothorax*, and *teashirt*. *Genes Dev*. 2002;16(18):2415–27. <https://doi.org/10.1101/gad.1009002> PMID: 12231630
84. Datta RR, Weasner BP, Kumar JP. A dissection of the *teashirt* and *tiptop* genes reveals a novel mechanism for regulating transcription factor activity. *Dev Biol*. 2011;360(2):391–402. <https://doi.org/10.1016/j.ydbio.2011.09.030> PMID: 22019301
85. Neto M, Naval-Sánchez M, Potier D, Pereira PS, Geerts D, Aerts S, et al. Nuclear receptors connect progenitor transcription factors to cell cycle control. *Sci Rep*. 2017;7(1):4845. <https://doi.org/10.1038/s41598-017-04936-7> PMID: 28687780
86. Heilig JS, Freeman M, Laverty T, Lee KJ, Campos AR, Rubin GM, et al. Isolation and characterization of the disconnected gene of *Drosophila melanogaster*. *EMBO J*. 1991;10(4):809–15. <https://doi.org/10.1002/j.1460-2075.1991.tb08013.x> PMID: 1901262
87. Ramaekers A, Claeys A, Kapun M, Mouchel-Vielh E, Potier D, Weinberger S, et al. Altering the Temporal Regulation of One Transcription Factor Drives Evolutionary Trade-Offs between Head Sensory Organs. *Dev Cell*. 2019;50(6):780–792.e7. <https://doi.org/10.1016/j.devcel.2019.07.027> PMID: 31447264
88. Wang C-W, Sun YH. Segregation of eye and antenna fates maintained by mutual antagonism in *Drosophila*. *Development*. 2012;139(18):3413–21. <https://doi.org/10.1242/dev.078857> PMID: 22912416
89. Datta RR, Lurye JM, Kumar JP. Restriction of ectopic eye formation by *Drosophila* *teashirt* and *tiptop* to the developing antenna. *Dev Dyn*. 2009.
90. Lim J, Choi K-W. Bar homeodomain proteins are anti-proneural in the *Drosophila* eye: transcriptional repression of atonal by Bar prevents ectopic retinal neurogenesis. *Development*. 2003;130(24):5965–74. <https://doi.org/10.1242/dev.00818> PMID: 14573515
91. Mahaffey JW, Griswold CM, Cao QM. The *Drosophila* genes *disconnected* and *disco-related* are redundant with respect to larval head development and accumulation of mRNAs from deformed target genes. *Genetics*. 2001;157(1):225–36. <https://doi.org/10.1093/genetics/157.1.225> PMID: 11139504
92. Schember I, Reid W, Sterling-Lentsch G, Halfon MS. Conserved and novel enhancers in the *Aedes aegypti* single-minded locus recapitulate embryonic ventral midline gene expression. *PLoS Genet*. 2024;20(4):e1010891. <https://doi.org/10.1371/journal.pgen.1010891> PMID: 38683842

93. True JR, Haag ES. Developmental system drift and flexibility in evolutionary trajectories. *Evol Dev.* 2001;3(2):109–19. <https://doi.org/10.1046/j.1525-142x.2001.003002109.x> PMID: [11341673](#)
94. Costas J, Casares F, Vieira J. Turnover of binding sites for transcription factors involved in early *Drosophila* development. *Gene.* 2003;310:215–20. [https://doi.org/10.1016/s0378-1119\(03\)00556-0](https://doi.org/10.1016/s0378-1119(03)00556-0) PMID: [12801649](#)
95. Ludwig MZ, Bergman C, Patel NH, Kreitman M. Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature.* 2000;403(6769):564–7. <https://doi.org/10.1038/35000615> PMID: [10676967](#)
96. Ludwig MZ, Patel NH, Kreitman M. Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. *Development.* 1998;125(5):949–58. <https://doi.org/10.1242/dev.125.5.949> PMID: [9449677](#)
97. Buffry AD, Mendes CC, McGregor AP. The Functionality and Evolution of Eukaryotic Transcriptional Enhancers. *Adv Genet.* 2016;96:143–206. <https://doi.org/10.1016/bs.adgen.2016.08.004> PMID: [27968730](#)
98. Hong J-W, Hendrix DA, Levine MS. Shadow enhancers as a source of evolutionary novelty. *Science.* 2008;321(5894):1314. <https://doi.org/10.1126/science.1160631> PMID: [18772429](#)
99. Taskiran II, Spanier KI, Dickmanken H, Kempynck N, Panikov A, Eki EC, et al. Cell-type-directed design of synthetic enhancers. *Nature.* 2024;626(7997):212–20. <https://doi.org/10.1038/s41586-023-06936-2> PMID: [38086419](#)
100. Mazo-Vargas A, Langmuller AM, Wilder A, van der Burg KRL, Lewis JJ, Messer PW, et al. Deep cis-regulatory homology of the butterfly wing pattern ground plan. *Science.* 2022;378(6617):304–8. <https://doi.org/10.1126/science.abi9407> PMID: [36264807](#)
101. Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, et al. The Release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Res.* 2015;25(3):445–58. <https://doi.org/10.1101/gr.185579.114> PMID: [25589440](#)
102. Rauluseviciute I, Riudavets-Puig R, Blanc-Mathieu R, Castro-Mondragon JA, Ferenc K, Kumar V, et al. JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 2024;52(D1):D174–82. <https://doi.org/10.1093/nar/gkad1059> PMID: [37962376](#)
103. Bravo Gonzalez-Blas C, De Winter S, Hulselmans G, Hecker N, Matetovici I, Christiaens V, et al. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat Methods.* 2023;20(9):1355–67. <https://doi.org/10.1038/s41592-023-01938-4> PMID: [37443338](#)
104. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504. <https://doi.org/10.1101/gr.1239303> PMID: [14597658](#)
105. Dunipace L, Ozdemir A, Stathopoulos A. Complex interactions between cis-regulatory modules in native conformation are critical for *Drosophila* snail expression. *Development.* 2011;138(18):4075–84. <https://doi.org/10.1242/dev.069146> PMID: [21813571](#)
106. Casares F, Mann RS. A dual role for homothorax in inhibiting wing blade development and specifying proximal wing identities in *Drosophila*. *Development.* 2000;127(7):1499–508. <https://doi.org/10.1242/dev.127.7.1499> PMID: [10704395](#)
107. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods.* 2012;9(7):676–82. <https://doi.org/10.1038/nmeth.2019> PMID: [22743772](#)
108. Nagaso H, Murata T, Day N, Yokoyama KK. Simultaneous detection of RNA and protein by in situ hybridization and immunological staining. *J Histochem Cytochem.* 2001;49(9):1177–82. <https://doi.org/10.1177/002215540104900911> PMID: [11511686](#)