A conserved transcriptional regulatory program fuels tissue differentiation during development

Fly ERC team *

2020

Abstract

During development, most cells undergo striking changes in order to develop into functional tissues. All along this process, the identity of each tissue arises from the particular combination of regulatory transcription factors that specifically control the expression of relevant genes for growth, pattern formation and differentiation. In this scenario, regulation of gene expression turns out to be essential to determine cell fate and tissue specificity. Although many studies aim to decipher tissue signature through the analysis of their transcriptome profiles, most lack temporal information during development and, in consequence, many differentiation events are poorly understood. To characterize dynamic transcriptional profiles during differentiation, we tracked down the transcriptome of committed cells throughout differentiation of eye, leg and wing of Drosophila melanogaster. Our analyses indicate that transcriptome profiles of different tissues at a specific stage of differentiation are more similar to each other, than to their own lineage in the following stage. We identified a gene regulatory program shared among tissues, which suggests a tight and coordinated regulation during differentiation. Moreover, we deeply analyzed the regulatory features associated with differentiation genes and generated a gene regulatory network to model the transcriptional program of tissue differentiation in flies. Comparative genomic analysis across metazoans indicated that tissue transition from precursor undifferentiated to fully mature state in other species, such as mouse, human and worm, follow a similar regulatory program than in fly. In

*funded by ..

conclusion, our data supports a novel view of differentiation model, in which different cell types and tissues share transcriptional changes to accomplish differentiation.

Background

Background

All animals developed from a single totipotent unique cell. Thus, in the course of development cells need to proliferate and get committed to distinct cell fates in order to ultimately, through cell differentiation, form specialized organs and tissues composed of a wide variety of cell types. The arrangement of such a plethora of diverse cell lineages, that share the same genome, can only be achieved by differential expression of specific genes, mediated by intriguing and complex transcriptional and epigenetic regulatory networks. Conventional view of differentiation uses Waddington's diagram of epigenetic landscape [**?**] to explain cell fate commitment during development. Waddington envisioned tissue development as a linear and progressive restriction path that is particular and distinctive for each specific tissue or organ. In this model the transition from a proliferative state to a definitive differentiated quiescent state is achieved through several cell fate decisions driven by precise epigenetic regulatory programs. However, studies from the last decades in cell reprogramming, transdifferentiation and regeneration slightly changed this perspective, introducing the idea that developmental and adult cells have certain plasticity and that the differentiation process is, totally or partially, reversible, both, in vitro (ie. reprogramming of inducible pluripotent stem cells (iPS cells)) or in vivo (ie. dedifferentiation or transdifferentiation after injury), review in [**?**]. The medical potential of such discoveries has encouraged in the last few years many investigators to try to decipher the regulatory mechanisms underneath cell fate and tissue differentiation. In that direction, many studies took advantage of next generation sequencing (NGS) technologies to try to characterize gene regulation in tissue differentiation by analyzing the transcriptome of different organs, tissues or cell types. Most of these investigations provided lists of genes specifically associated with certain tissues or cell types in adult organisms or in vitro [**???????**], but, unfortunately, many of them lack temporal dynamics during the process of differentiation and most don't take into account developmental context. In this study we took advantage of Drosophila melanogaster development to understand the mechanisms underlying cell differentiation and organ formation during development. Using RNA-Seq we

follow the transcriptional dynamics of cells committed to give rise to specific cell types from the eye, leg and wing, through development. These organs arise from Drosophila imaginal discs, that are epithelial sacs first specified in embryos that grow and acquire specific cell fates during larval stages [**?**]. This valuable data set allows us to characterize a Differentiation gene regulatory program shared among diverse tissues in fly development and to identify distinct functions and gene features associated with un-differentiated proliferative and fully differentiated quiescent state, in flies. To document deeply the regulatory programs underlying differentiation, we studied the dynamics of chromatin accessibility and H3K4me3 mark landscape surrounding transcriptional changes, as well as, the differential expression of associated Transcription factors (TFs). Altogether, these results allow us to generate a Gene regulatory network of eye, leg and wing differentiation to give insights into the models of gene regulation through eye,leg and wing differentiation. Interestingly, other tissues in Drosophila, for instance the brain and the gut, experience close transcriptional changes in the course of their differentiation. Besides, the orthologs of the Differentiation genes follow similar expression dynamics in highly specialized organ differentiation, through metazoans. Thus, the Differentiation regulatory mechanism we characterized in this manuscript is conserved among different cell types and species. In conclusion, we propose that beyond transcriptional divergence throughout tissues maturation, a common and coordinated mechanism would fuel progression through tissue and cell differentiation.

Results

Transcriptional profile through tissue differentiation show strong contribution of time

During fly metamorphosis, imaginal tissues undergo cell differentiation and morphogenetic rearrangement to give rise to adult functional appendages (Fig. 1a). To characterize tissue differentiation during this process we interrogated the transcriptome of imaginal tissues in Drosophila along terminal development in the fruit fly . We tracked down precursor cells along differentiation using GFP reporter lines, in which transcription of GFP was induced under the promoter of genes specifically expressed in a particular tissue (see Methods and Supplementary Fig. 1a and 1c). Briefly, imaginal tissues were manually dissected and disaggregated by trypsin treatment. After that, cells were collected by fluorescence-activating cell sorting (FACS) (Supplementary Fig. 1b), RNA was extracted and processed for NGS. We generated

RNA-Seq data from eye, leg and wing from three different stages: Larva3 (L3, around 110h of development), when cells are predetermined and committed to give rise to specific cell type in adult but still undifferentiated and keep proliferation capacity **?**), White pupa (WP), immediately after entering pupariation and coinciding with Ecdysone hormone signaling peak (around 120h), and Late pupa (LP, around 192h , corresponding to 72h after pupa formation), when cells from all tissues are fully differentiated and almost functional. In addition, we also generated RNA-seq data fWe expression values for 17,158 annotated genes (FlyBase gene annotation r6.05, summary statistics of RNA-seq samples in Supplementary Fig 2a,b).As a control, we investigated the expression of well characterized genes expressed in eye, leg and wing **??????**, as well as genes reported to be transcriptionally changing during fly development **???** (Fig. 1b). In all cases, genes behaved as previously reported in the literature. We used hierarchical clustering and principal component analysis (PCA) to analyze the gene expression data. Interestingly, the results show that, first, transcriptomes appear to be more similar when the tissues are undifferentiated and diverge while they become differentiated, and second samples cluster preferentially by early or late stage of differentiation, followed by separation of neural (eye) and non-neural tissues (leg and wing), followed by neural late stage and non-neural late stage (Fig.1c, Supplementary Fig. 2c). These findings indicate that, unexpectedly, time may have a strong contribution to the global differences in gene expression during tissue differentiation, and suggest that transcriptome profiles of different tissues at an early stage of differentiation are more similar to each other than to their own lineage when cells are differentiated.

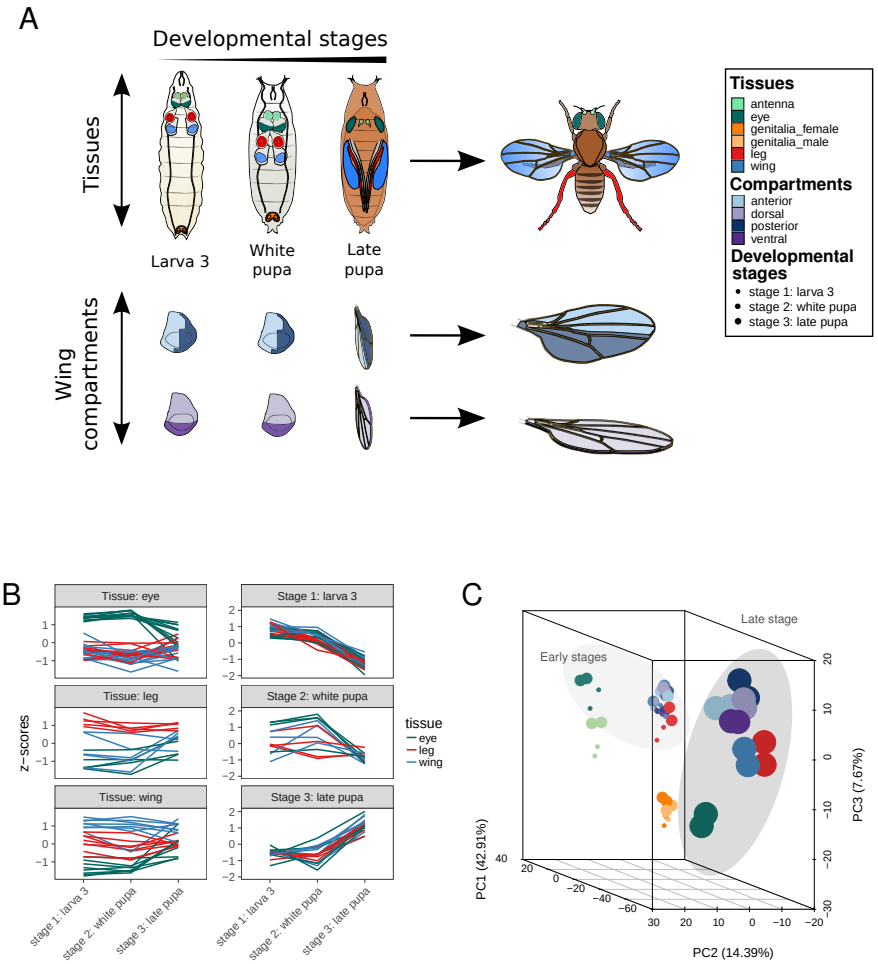


Fig. 1 Transcriptional profiling of *Drosophila* imaginal dics. (**a**) Overview of sequenced tissues and wing compartments and respective developmental stages. (**b**) Control genes for each tissue and for each developmental stage described in literature. (**c**) Principal component analysis based on the expression of 1000 most variable genes in tissues and developmental stages. PC1 indicates time, mainly separating the early stages and the late stage. PC2 separates neural and non neural tissues. PC3 separates eye late stage to the other samples. Selected genes show expression levels of at least 5 TPMs in at least 2 samples. Gene expression is computed as log10-normalized TPM with pseudocount of 0.01.

Dynamic changes in gene regulation occurs in a coordinated manner along tissue development

To quantify the relative contribution of developmental time and tissue in gene regulation and to further characterize which genes are essential to establish tissue specificities during development, we investigate spatio-temporal dynamics of eye, leg and wing transcriptomes. We analyzed the total variance of the expression levels of a given gene using linear models (see Methods) **?**. More in detail, we defined a linear model for each gene, in which the expression level is a function of developmental time (the developmental stages), the tissue, the interaction between both factors, and an additional residual error (Figure 2a). Notably the time contribution to transcriptome variation is much larger than the others (Fig. 2b), suggesting that all imaginal tissues, independently of their fate, experience similar changes in transcriptome while they undergo differentiation. The transcriptome of Drosophila tissues suffer strong changes during metamorphosis but there are no huge discrepancies in gene expression between different tissues. According to our results it seems that is the expression of few genes what makes the distinction between tissues. To deeper investigate the role of a differentiation program in tissue development and to characterize tissue regulatory programs, we identified differentially expressed genes (DEGs) across tissues and developmental stages based on a concordant gene set of a compendium of linear models and gene profilers (see Methods) **??**. This analysis allowed us to classify regulated genes in three categories: differentially expressed genes across differentiation (referred to as stage genes (SGs)), differentially expressed genes across eye, leg and wing (referred to as tissue genes (TGs)) and differentially expressed genes specifically in a subset of tissue(s) and stage(s) (referred to as tissue-stage genes (TSGs)). In concordance with our previous results, a larger amount of genes, 1445 genes, were classified as SGs, 334 as TGs, and 255 as TSGs (Figure 2d). As expected SGs mainly correspond to the genes showing high variation across time while TGs show high variation across tissue (Fig. 3c). Later, we classified DEGs in minor groups based on upregulation in precise tissues/stages (Fig. 2e and Supplementary Fig. 3b). To further study the mechanisms coordinating differentiation of imaginal tissues, we inspected DEGs that behave similarly in eye, leg and wing during differentiation (SGs). Our analysis showed that genes downregulated during differentiation, referred as early differentiation genes (822 genes) represent around 56Tissue comparison shows higher divergence in gene regulation in the eye than in leg and wing. Tissue-specific cell fate GO terms are enriched for each minor group (Supplementary Fig. 3d). We also found terms related to regulation of transcription, which agrees with the enrichment of transcrip-

tion factors (TFs) found in TGs compared to the other major DEG groups (Fisher's Exact Test on Count Data, p¡0.001, Fig. 2d and Supplementary Fig. 3b). We also classified our genes regarding changes in expression from less than 5TPM in all other DEG groups but the one of interest (restricted genes) in comparison to widespread genes (Fig. 2e and Supplementary Fig. 3f,g). Most of TFs in TGs are tissue restricted, i.e. only expressed in the tissue where it is differentially expressed (Supplementary Fig. 3h). In addition to that, the TG TFs show higher fold change of gene expression compared to nonTF TGs, while other DEG groups do not show such trend (Supplementary Fig. 3i). Overall, our results indicate that most genes associated with tissue specificity are expressed in a restricted manner already within undifferentiated imaginal tissues in L3 (Fig. 2f). Some of these genes, for instance, are well known to be essential to regulate cell determination and tissue formation during development (ie. eyeless (ey) and glass (gl) in eye). This evidence suggests that tissue regulators are not enough to drive tissue differentiation alone during development. Finally, most changes in time and tissues occur, as expected, in late differentiation (Fig. 2e and Fig. 1e). Altogether, our analysis supports the idea of a novel temporal programme that could lead differentiation in imaginal tissues through tightly regulated changes in transcription that may alter cell cycle, gene regulation, translation and metabolism. We hypothesize that this temporal differentiation mechanism in combination with particular tissue genes could be pivotal for organ differentiation.


https://public-docs.crg.es/rg/Data/cklein/publications/flyAtlas/figures/ma


https://public-docs.crg.es/rg/Data/cklein/publications/flyAtlas/figures/supp

Gene regulatory features in tissue differentiation

It is thought that during development and differentiation sequence-specific DNA-binding Transcription Factors (TFs) are ones of the major players in establishing tissue specificities and ultimately cell fate (review in **??**). To understand how changes in transcription during differentiation are orchestrated, we first explored the presence of TFs among the set of genes changing across imaginal tissue differentiation. We found a strong enrichment of TFs inside TGs (around 20Presence of H3k4me3 mark around the transcription start site (TSS) has been associated with active gene regulation (review in **??**). Using chromatin Immunoprecipitation assays we generated H3K4me3 profiles for eye, leg and wing differentiating tissues. Analysis of H3K4me3 peak presence/absence in TSSs of Differentiating genes through stages and

tissues reveals that widespread genes promoters are marked with H3K4me3 in a general manner, correlating with their wide pattern of expression (Fig. 3b and Supplementary Fig.4a). On the contrary and in agreement with previous results, showing that expression can occur in absence of H3K4me3 in development genes **???**, restricted genes show mostly unmarked promoters at any tissue or stage. However, this is not the case for Early and Peaking restricted genes, that remain marked even if they are not expressed when tissues aredifferentiated. In conclusion, only widespread genes display concordance between presence of H3Kme3 in TSS and their expression. Although few restricted genes show specific changes in H3K4me3 marking according to changes in expression, many of these genes are unmarked, suggesting that their tight regulation is not a direct consequence of H3K4me3 presence in their promoters.

co-expression Gene regulatory network for imaginal tissues Differentiation

Gene regulatory networks (GRNs) aim to reconstruct and integrate transcriptional regulators and gene expression changes to infer how functional expression of the genome occurs in a particular process. To build a high-confidence co-expression gene regulatory network that encompasses the Differentiation genes identified in our study, as well as their putative regulators we used combined approaches centered on TF-target co-expression and motif discovery in accessible proximal promoters. First, we used FAIRE-Seq data of Drosophila imaginal tissues from L3 and LP stages **?** to identify the accessible promoter regulatory regions of Differentiation genes during development (see methods). We used motif matrices from Drosophila TFs from several databases to scan the accessible promoters with FIMO (MEME-Suite) **?** (see Supplementary Fig. 5a and methods for details). We filtered the interactions in the Differentiation network according to motif sequence conservation. Finally, we built a robust GRN, using averaged expression values between replicates, from our 27 samples, for Differentiation genes and 775 Drosophila TFs. We took advantage of the widely used Weighted Correlation Network Analysis R package (WGCNA) **?** to identify correlating TF-target pairs and built the Differentiation network. These analyses identified 14039 predicted TF-target interactions. Altogether this pipeline allowed us to reconstruct a co-expression GRN for imaginal tissue differentiation that comprises 1714 nodes (1485 predicted targets for 229 TFs) and 14039 edges (Fig. 4a, 4c, Supplementary Fig. 5b, 5c and Table S7). On one hand, as expected, results indicate that positive interactions between TFs and tar-

gets(10072/14039 interactions, 72

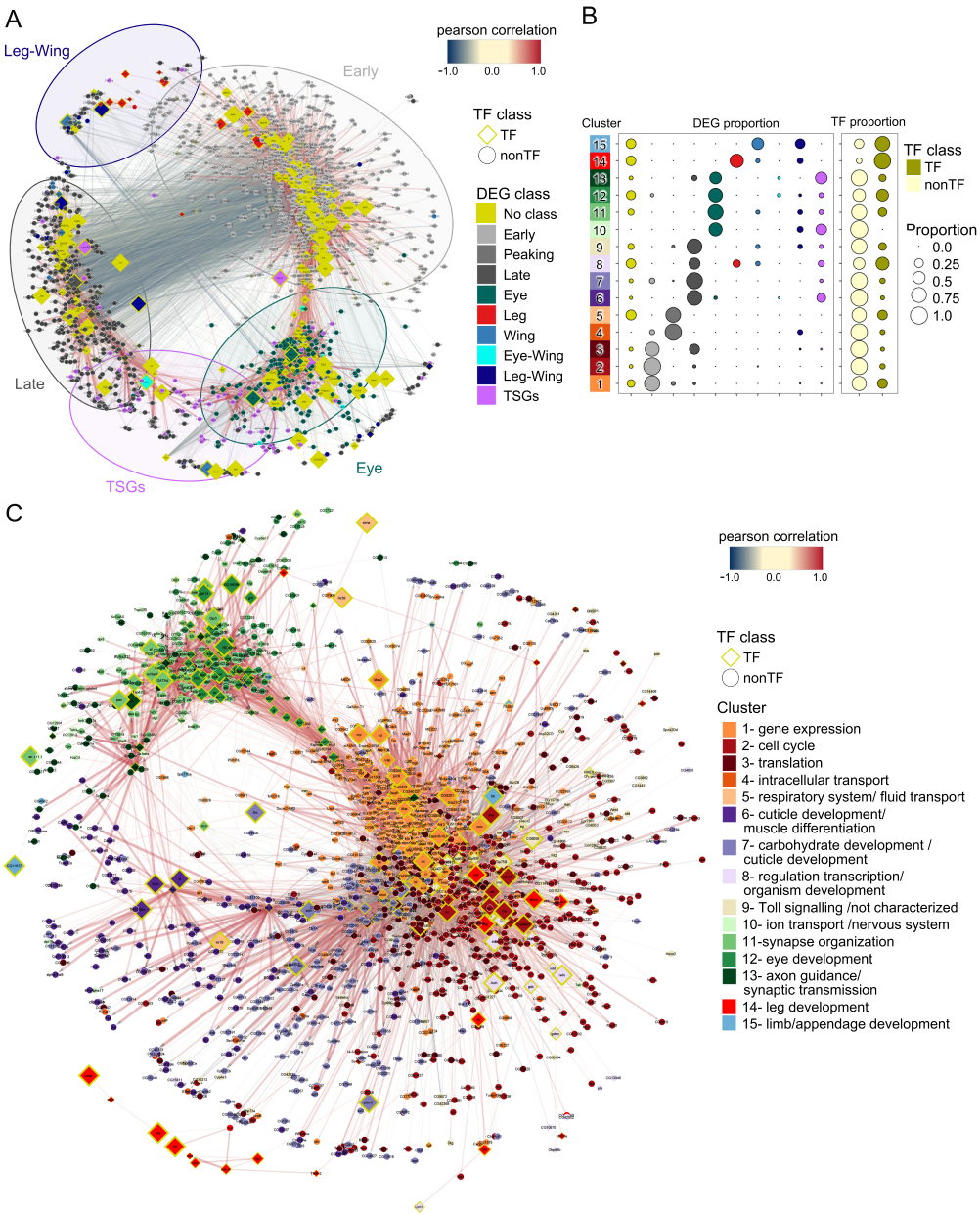


Fig. 4 Co-expression Gene Regulatory network for Differentiation genes. **(a)** GRN landscape displaying nodes according to TF-target correlation of expression. (pearson correlation). Edges were colored according to TF-target pearson correlation (red > 0.3, blue < -0.3). Positive and negative correlations between DEGs can be observed. Notice negative predictive interaction between Early and Late genes as well as among different Tissue genes. Nodes sized reflected node closness centrality.**b)** Proportion of DEGs and TFs in gene expression clusters. **(c)** GRN clusters landscape . Nodes displayed according to betweenness centrality. Cluster 1, associtaed to gene regulation, show higher centrality compare to other clusters.

Differentiation transcriptional program occurs in several tissues along differentiation in flies

Our results so far indicate that there is a coordinated response in gene expression among imaginal tissues. Such accurate and synchronized regulation suggests that this transcriptional program could occur in a systemic manner. Analysis of the expression of Time genes in RNA-seq data from whole animals at L3, EP and LP stages (modENCODE data, **??**) showed that this is not the case (Figure 5a), suggesting that the regulatory transcriptional program we identified is not followed by the entire animal during post-embryonic development. However, deeper inspection along all development, from embryo 0h to adult, again using whole body RNA-Seq data (from modENCODE) highlights an interesting dynamic expression profile during embryonic stages. Early differentiation genes are highly upregulated at the beginning of embryogenesis, coinciding with active proliferation state, and their expression decreases around mid embryogenesis, coinciding with morphogenetic arrangements and organ primordia specification. On the contrary, Late differentiation genes appear upregulated from mid to late embryogenesis, and in larval and pupa stages compared to early embryogenesis. These results are in agreement with previous reported analyses showing that early stages of embryogenesis resemble larva stages while late stages of embryogenesis mirror pupal stages **??**. These findings indicate that the genes we identified, having a dynamic pattern of expression during embryogenesis, could be characteristic features of undifferentiated and differentiated cell states. In agreement with that, inspection by a PCA using expression values for SGs from diverse tissues and cell populations of Drosophila in early and late differentiation states **???**, showed that the first component clusters apart early and late differentiation stages for all tissues and cells (Figure 5c). Altogether these results demonstrate that, indeed, the Differentiation regulatory program is characteristic of tissues and cells undergoing differentiation in flies and that the Early and Late differentiation genes are distinctive traits to commited undifferentiated cells and fully differentiated ones, respectively.

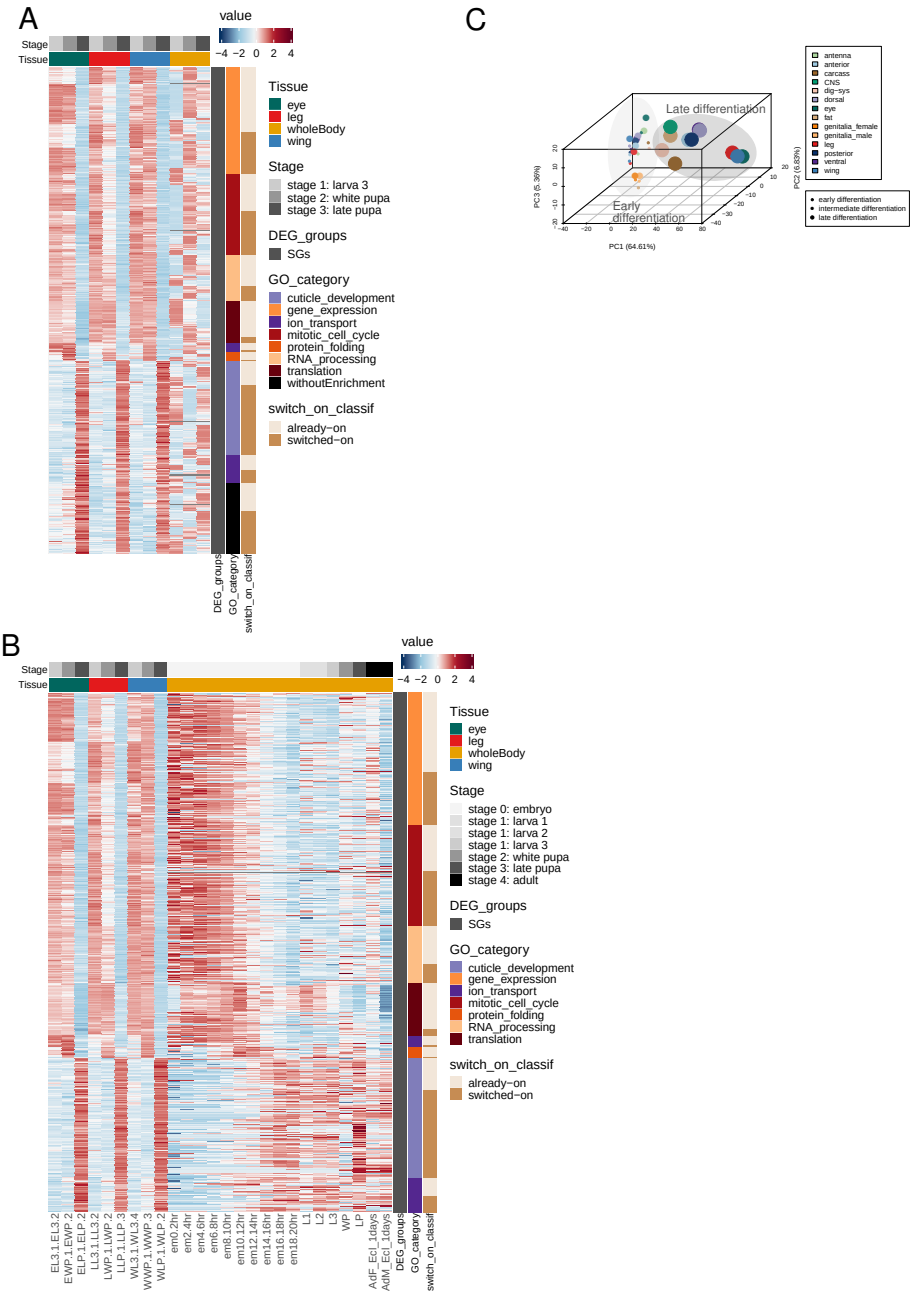
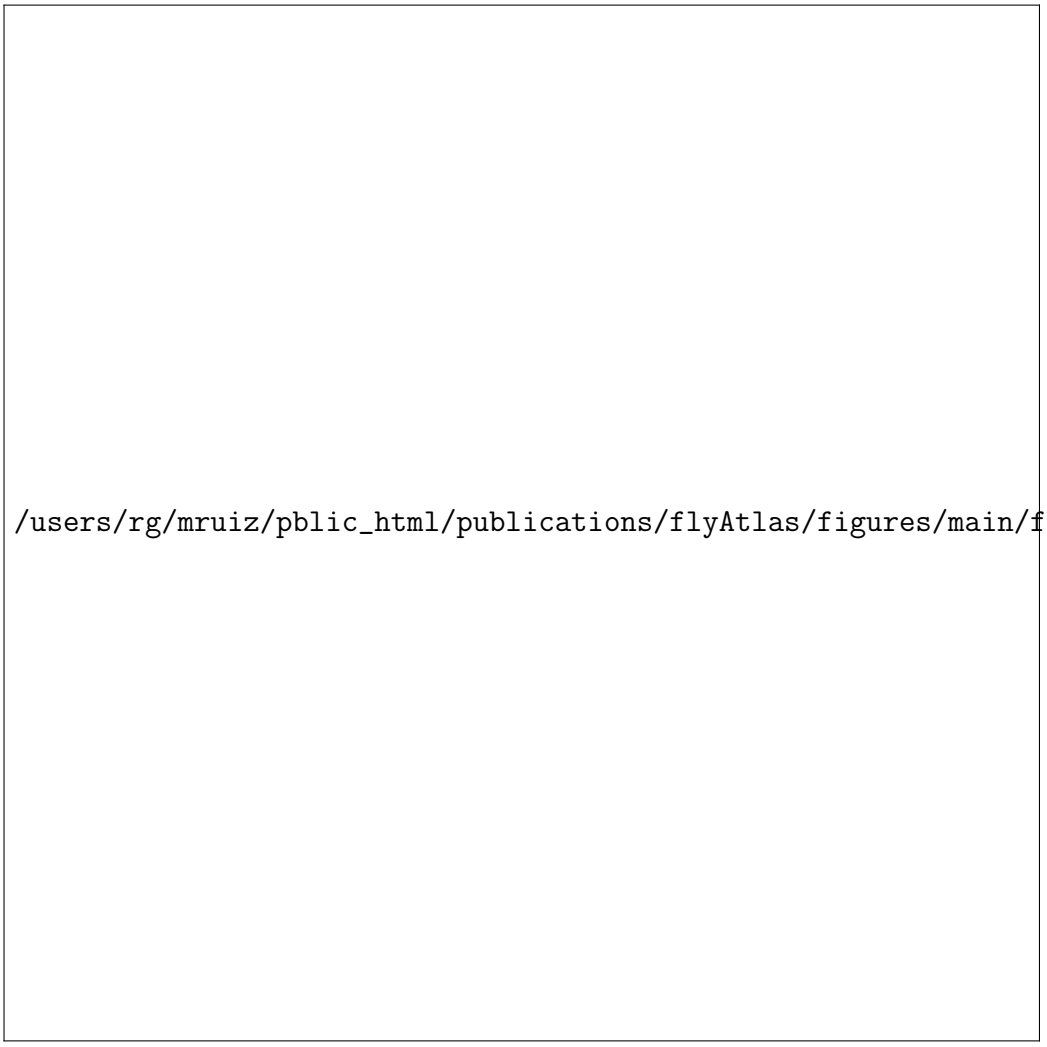


Fig. S5 RNA dynamics through whole animal and other tissues along fly development. (**a**) Expression of SGs in modENCODE whole animals at L3, WP and LP stages. (**b**) Expression of SGs in modENCODE whole animals from embryo to adult stages. (**c**) Principal component analysis of Expression of SGs in modENCODE tissue data.

The differentiation regulatory program is conserved through metazoans

We provided evidence that the genes belonging to the differentiation regulatory program identified in imaginal tissues are potentially key indicators of the differentiation state of organs and cell types in flies. To investigate if this mechanism is conserved outside insects we compared fly differentiation to mouse, human, and worm using RNA-Seq data from diverse organs at different developmental time points **??**. We identified the orthologs of differentiation genes in every species and analyzed their expression along organ development (Figure 6 and Supplementary Fig. 7). We used self-organizing maps (SOM) to cluster the orthologs of the fly Early and Late genes in the different species. Mouse orthologs were clustered based on the expression profile in four tissues and 10 time points from embryo 10.5 days through 3 days post-natal stage. Early orthologs and Late ones are clearly clustering appart (Figure 6b). In every tissue higher expression of Early orthologs in the first time points is followed by the transition of the higher expression levels to the Late orthologs towards the last time points (Figure 6cb). While Early orthologs when more expressed are evenly distributed, Late ones show a certain tissue-specific tendency of expression. More than 80Finally we examined the correlation of the expression values between Drosophila TF-target pair orthologs in each mouse differentiating organ. Surprisingly, considering the high genome diversity in mouse compared to fly, considerably strong, negative and positive, correlations between pairs were observed during Brain and Kidney development (Supplementary Fig. 7a and 7b). These results suggest that the regulatory interactions we characterized for tissue differentiation in fly, by GRN analysis, might be conserved through metazoans, at least in certain organs development.


/users/rg/mruiz/pblic_html/publications/flyAtlas/figures/main/figure6.pdf