A transcriptional program shared across lineages underlies cell differentiation during metazoan development

[1+*]Marina Ruiz-Romero, [1,4+]Cecilia C. Klein, [1]Sílvia Pérez-Lluch, [1,$]Amaya Abad, [1,#]Alessandra Breschi, [12*]Roderic Guigó

[1] Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology (BIST), Dr. Aiguader 88, Barcelona 08003, Catalonia, Spain

[2] Universitat Pompeu Fabra (UPF), Barcelona, Catalonia, Spain

[4] Departament de Genètica, Microbiologia i Estadística and Institute of Biomedicine (IBUB), Universitat de Barcelona. Barcelona, Catalonia, Spain

[#] Current address: Tempus Labs, Inc.

$ Current address: Center for Applied Medical Research (CIMA), University of Navarra and Institute of Health Research of Navarra (IdiSNA), Pamplona, Spain.

+: these authors have contributed equally to this work.

* Corresponding authors: marina.ruiz@crg.eu, roderic.guigo@crg.eu

Abstract

Background

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.

Results

To characterize the dynamic transcriptional profiles during cellular differentiation, we tracked down the transcriptome of committed cells in different *Drosophila melanogaster* tissues and compartments at a number of developmental stages. We found that during fly development, temporal transcriptional changes shared across lineages are much larger than spatial lineage-specific transcriptional changes, and that cellular differentiation is dominated by a transcriptional program, common to multiple lineages, that governs the transition from undifferentiated to fully differentiated cells independently from the differentiation end point. The program is under weak epigenetic regulation, and it is characterized by downregulation of genes associated with cell cycle, and concomitant activation of genes involved in oxidative metabolism. Largely orthogonal to this program, tissue specific transcriptional programs, defined by a comparatively small number of genes are responsible for lineage specification. Transcriptome comparisons with worm, mouse and human, reveal that this transcriptional differentiation program is broadly conserved within metazoans.

Conclusions

Our data provides a novel perspective to metazoan development, and strongly suggest a model, in which the main transcriptional drive during cell type and tissue differentiation is the transition from precursor undifferentiated to terminally differentiated cells, irrespective of cell type.

Keywords

Tissue differentiation, development, gene regulation, evolutionary conservation

Background

All pluricellular organisms develop from a single totipotent cell. In the course of development, cells proliferate and commit to distinct cell fates to ultimately, through cell differentiation, produce a plethora of cell types that combine in specialized tissues and organs. Such a diversity of cell types, all sharing the same genome sequence, is the consequence of differential expression of specific genes, which is driven by complex transcriptional and epigenetic regulatory networks. The conventional view of differentiation explains cell fate commitment as a linear and progressively restricted path that is distinctive for each specific cell type (based on Waddington's diagram of epigenetic landscape (1)). In this model, the transition from a proliferative state to a 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 have slightly changed this view, by showing that adult cells retain a certain plasticity and that the differentiation process is reversible to varying degrees, both *in vitro*, for example in the case of reprogramming inducible pluripotent stem cells (iPS cells) and *in vivo*, for example in the case of dedifferentiation and transdifferentiation after injury (reviewed in (2)).

In the past two decades, research into the regulatory mechanisms underneath cell fate and tissue differentiation has been enormously facilitated by next generation sequencing (NGS) technologies. Recent single cell sequencing technologies, in particular, have led to the identification of specific gene expression profiles associated with tissues or cell types in adult organisms or after *in vitro* differentiation (3–9).

Here, we analyze the development of *Drosophila melanogaster* --an experimentally manageable model within metazoans-- to characterize the temporal and spatial transcriptional programs that underlie tissue differentiation during animal development. In contrast to previously published fly development transcriptional studies, here we specifically, label primordial cells from imaginal discs --internal epithelial sacs in larvae that are committed to give rise to specific tissues in adults (10)--, isolate them using fluorescence-activating cell sorting (FACS), and profile their transcriptional state with RNA-seq at different developmental stages. Differential gene expression analysis reveals that, in contrast to the prevalent view, a transcriptional program common across cell lineages governs the transition from undifferentiated to fully differentiated cells, dominating tissue specific programs, which are

93 defined by a comparatively low number of genes, and tend to be activated late during
94 development. Comparative transcriptomics analyses show that this program is substantially
95 conserved across metazoans.
96
97
98 # Results
99
100 **Spatial and temporal transcriptional analysis of imaginal discs during *Drosophila***
101 ***melanogaster* development**
102
103 During fly metamorphosis, imaginal tissues undergo cell differentiation and morphogenetic
104 rearrangement to give rise to adult functional appendages (**Fig. 1A)**. To investigate the
105 molecular basis of this process, we interrogated the transcriptome of imaginal tissues at
106 different stages during terminal fly development. Within each tissue and developmental stage,
107 we selected the precursor cells that differentiate into the adult tissue. To track down precursor
108 cells, we used GFP reporter lines, in which GFP is driven by the promoter of genes specifically
109 expressed in a particular region of each tissue. Briefly, imaginal tissues were manually
110 dissected and disaggregated by trypsin treatment. After that, cells were collected by
111 fluorescence-activating cell sorting (FACS), RNA was extracted and processed for NGS (see
112 Methods and **Fig. 1B** and **Supplementary Fig. 1A,B**). Overall we generated RNA-Seq data
113 from eye, leg and wing from three different stages: third instar larvae (L3, around 110 h of
114 development), when cells are predetermined and committed to specific cell types in adult, but
115 they are still undifferentiated and keep proliferative capacity(10)), early pupa (EP, around 120 h
116 of development), immediately after entering pupariation and coinciding with Ecdysone hormone
117 signaling peak, and late pupa (LP, around 192 h of development, corresponding to 72 h after
118 pupa formation), when tissues are fully differentiated and almost functional. In addition, we also
119 generated RNA-seq data for antenna and genitalia discs (male and female), for L3 and EP.
120 Finally, we produced RNA-seq for four wing compartments (anterior, posterior, ventral and
121 dorsal) for the three developmental stages. In total, considering two replicates per condition, we
122 generated RNA-Seq data for 54 samples. In addition, we generated H3K4me3 ChIP-Seq for
123 eye, leg and wing at these three developmental time points (two replicates per condition, 18
124 ChIP-Seq samples, in total).
125

From the RNA-Seq data, we estimated expression values for 17,158 annotated genes (FlyBase gene annotation r6.05, summary statistics of RNA-seq samples in Methods**,** all data available at https://rnamaps.crg.es). Genes with known tissue specific (11–15), or developmental (16–18) transcriptional patterns behaved as expected (**Fig. 1C** and **Supplementary Table 1**).

A common transcriptional differentiation program is shared across tissues during fly development

Principal component analysis (PCA) and hierarchical clustering (**Fig.1D**, **Supplementary Fig. 2A, B**) show that samples cluster preferentially by developmental time than by tissue lineage. That is, the transcriptomic profile of a given tissue at a specific stage of development is more similar to the profile of other tissues at that stage, than to the same tissue in other developmental stages. Using linear models (see Methods (19)), we estimated that the proportion of the gene expression variance explained by the developmental time (41% on average) is indeed much larger than that explained by the tissue (22%) (**Fig. 2A, B**). During fly tissue differentiation, therefore, temporal changes of gene expression dominate over spatial changes.

From a set of 9,334 genes that are expressed at least 5 TPMs in at least two samples, we identified a set of 2,034 genes that change expression across tissues (eye, leg and wing) and/or time points (Developmentally Dynamic Genes, DDGs, see Methods(20)). We classified these genes in three categories: differentially expressed across developmental stages (stage genes, SGs, 1,445 genes, consistently with the result above, the largest category), across tissues (tissue genes, TGs, 345) and across both tissues and stages (tissue-stage genes, TSGs, 255), **Supplementary Table 2**). Within SGs, we further classified genes as downregulated or early differentiation genes (822 genes, 56%, preferentially expressed in L3 and EP), upregulated or late differentiation genes (571 genes, 39.5%, preferentially expressed in LP), and peaking or metamorphosis entrance genes (52, 4.5%), which are upregulated at EP in all tissues (**Fig. 2C** and **Supplementary Fig. 3A,B**). Early genes are associated with cell cycle, gene regulation, RNA processing and translation processes; metamorphosis entrance genes to endoplasmic reticulum localization and apoptosis signaling, and late genes are related to cuticle formation and chitin metabolism (**Supplementary Fig. 3C**). Among the late genes, a

large fraction (324, 57%) are poorly characterized genes with no associated functions, compared with 42% of metamorphosis entrance genes, and 27% of early genes. This could be ascertainment bias, as most functional characterization studies in flies are performed in developing, not in fully differentiated, animals.

TGs are enriched for the expected tissue-specific cell fate functional categories (**Supplementary Fig. 3D**). Some of these genes are known to be essential to regulate cell determination and tissue formation during development (e.g. *ey* and *gl* in eye). TGs correspond broadly to genes that are already differentially expressed at L3, when cells within imaginal tissues are undifferentiated, and remain differentially expressed all through development. TSGs, in contrast, are genes activated in a tissue specific manner only at specific developmental time points. We found that most TSGs are specifically activated in the transition from EP to LP, driven mostly by an expansion of eye specific genes. This results in a larger number of tissue specific genes at the terminal stage of differentiation associated to each tissue function (**Fig. 2D and Supplementary Fig. 3E**)**.** Transcriptional differences between tissues, therefore, increase with developmental time (**Fig. 1D, 2D**), suggesting that an expansion of tissue regulatory programs is needed for terminal tissue differentiation.

Overall, our results strongly suggest that during fly development, there is a temporal transcriptional program common to all tissues that dominates over tissue specific transcriptional programs, which are defined by a comparatively small number of genes. This program is of fractal nature, as it can be observed at different organizational scales. Indeed, we produced and analyzed expression data from the distinct compartments within wing imaginal discs. As with tissues, variation of gene expression is much larger across developmental time than across compartments (**Supplementary Fig. 4A,B,C**) and the transcriptional behavior of DDGs within the wing imaginal discs replicates the behavior observed among imaginal discs during development (**Supplementary Fig. 4D**).

We have investigated the epigenetic features underlying the fly temporal differentiation program. We generated H3K4me3 ChIP-Seq profiles for eye, leg and wing differentiating tissues and analyzed FAIRE-Seq data available for these tissues to assess chromatin accessibility (21). We focused on the core promoters (+- 250bp from the transcription start sites, TSS). Overall, we found most DDGs either in closed conformation and/or unmarked, or unspecifically open and/or marked (**Fig. 3A, Supplementary Fig. 5A-D**), Chromatin

accessibility, and H3K4me3 marking in particular, mostly reflect, actually, the breadth of gene expression. Genes with restricted expression (expression restricted to a single stage or/and tissue) are in closed chromatin conformation or/and unmarked more often than genes with widespread expression (**Fig. 3A, Supplementary Fig. 5A-E)**. This is consistent with previous reports that show absence of marking by canonically activating histone modifications in genes regulated during fly development (15,22,23). The exception are early genes with restricted expression patterns. These tend to be marked at early developmental stages and remain marked in late differentiation, even when not being expressed.

These results suggest that the fly transcriptional developmental program is, broadly, under weak promoter epigenetic regulation. We did find, however, a strong enrichment of transcription factors (TFs) within TGs (20%) compared to SGs (6%) or TSGs (5%, Fisher's Exact Test on Count Data, $p<0.001$) (**Fig. 3B,C),** suggesting that TFs play a comparatively more important role in spatial than in temporal differentiation. Within SGs the number of TFs is higher in early than in late genes, in concordance with previous observations during mammalian development (24).

A developmental gene regulatory network in fly differentiation

Gene regulatory networks (GRNs) integrate information from transcriptional regulators and their targets to provide a holistic view of the regulatory program of a particular biological process. Previous studies have successfully used transcriptional-based GRN to model gene regulation along developmental processes (12,25,26). Here, we used Weighted Correlation Network Analysis (WGCNA) (27) to construct a differentiation co-expression network connecting Developmentally Dynamic genes (DDGs) with their putative regulatory TFs. To identify reliable TF-target pairs, we scanned for conserved TF binding motifs occurring in open chromatin regions within the promoter regions of the target genes (see Methods for details, **Supplementary Fig. 6A**).

The resulting GRN includes 1,656 nodes (1,485 DDGs and 229 TFs), and 14,039 edges (**Fig. 4A**, **Supplementary Fig. 6B-F, Supplementary Fig. 7 A,B** and **Supplementary Table 3**). The WGNA identified 15 regulatory clusters (**Fig. 4B, Supplementary Fig. 6G-I**), to which we associated functional categories by GO enrichment analysis (**Fig. 4C**, **Supplementary Fig. 8**

and **Supplementary Table 3)**. Clusters 1 to 3 form a super-cluster mostly composed of early genes. Cluster 1 is functionally associated with regulation of gene expression (**Fig. 4C**, **Supplementary Fig. 8**) and, as expected is the cluster with the highest number of interactions (8,486, **Fig. 4D**). The TFs in the cluster include general transcriptional regulators, known to have functions on chromatin structure and regulation of gene expression like BEAF-32, Dref, Z, Dalao, BAP170 or Br; developmental chromatin remodelers such Trl, Pcl, Pho and Phol; insulators like Cp190 and SuHw; TFs associated to imaginal disc morphogenesis, like Hth, Exd, Sd, Da, Mad, Ets21C, Rn, Myc or Max; as well as TFs related to metabolic regulation, like Bigmax, Foxo and ATF-2 (**Supplementary Fig. 7C**). Although these TFs bind many early genes, interactions with late genes and TGs were also predicted. Clusters 2 and 3 are functionally associated with cell cycle and translation, respectively. Clusters 4 and 5, corresponding to very few peaking genes (**Supplementary Fig. 6H**) could also be included in this super-cluster (**Fig. 4A**).

Clusters 6 to 9 form a second super-cluster mostly composed of late genes. Only one of these, cluster 6, is preferentially regulating late genes, while the TFs within cluster 7 and 9 are predicted to have high number of interactions also with early genes, suggesting potential negative regulation (**Fig. 4C,D)**. Among these TFs, many are activated through stress, immune and hormonal signaling pathways, like dl, Gce, Hr4, Kay, Rel, Eip74EF, Eip75B and Eip93F; and many are known or predicted to have repressor capacity. This hints at a possible link between whole animal signaling and SGs. Signaling cascades, likely associated with metamorphosis, could induce expression of TFs that may repress early genes and activate late genes. Additionally, cluster 8 includes ten TFs: Abd-B, Awh, D, Gsb, Gsb-n, Lim1, Odd, Opa, Retn, Sob, all known to play a role in development and patterning, that predominantly interact with cell cycle genes (cluster 2, **Fig. 4D**).

Clusters 10 to 13 form a third super-cluster mainly composed of eye genes and eye late genes. They are functionally associated with eye morphogenesis and neural fate. Among them, cluster 12 is associated specifically with eye development, and includes eye fate regulators and neurogenesis inductors predicted to preferentially bind eye genes (**Fig. 4D**). These include well characterized genes like: *ey, gl, mirr, toy, oc, pnt, ro, scrt, ttk, lola* and *so* (**Supplementary Fig. 7D**). Some TFs inside these clusters, which are not necessarily expressed in a restricted manner in the eye, are predicted to also bind several early genes. The same is observed for TFs in cluster 14 (leg) and 15 (leg and/or wing, **Fig. 4D**), which form the fourth super-cluster.

These results point to a possible crosstalk in which tissue fate regulators mediate down-regulation of early genes when tissues differentiate.

TF-target interactions tend to take place within super-clusters that are associated to different temporal (early, late) and spatial (eye, leg-wing) niches (**Fig. 4D**), suggesting that the temporal expression program is largely orthogonal, with some, but little, cross-talk with tissue specific regulatory programs (**Fig. 4A,C,D**). Further plotting the network according to betweenness centrality, in which highly connected nodes are placed in proximity irrespective of the direction of the correlation (**Supplementary Fig.7A,B**), revealed that early genes are highly connected with TFs regulating a large number of genes, while late genes are more sparsely connected to their regulators, consistent with our observation that tissue specific expression programs unfold late during development (**Fig. 1C, 2C,D**).

The GRN helps to functionally characterize the temporal differentiation program shared across tissues. In this program, the transition from precursor to differentiated states is driven, independent of cell type, by downregulation of genes associated with cell cycle (cluster 2), with regulation of gene expression (cluster 1) and with translation (cluster 3), and by a concomitant activation of genes involved in oxidative metabolism and other metabolic pathways (cluster 7), and in (terminal) differentiation (cluster 6 and 8).

The fly transcriptional differentiation program captures the transition from undifferentiated to terminally differentiated cells, irrespective of the differentiation end point

To further corroborate our results, we have analyzed other fly developmental transcriptome data, which is not from tissue isolated cells, but from the whole body and from specific tissues. First, when analyzing data from carcass, central nervous system (CNS) and fat body available for L3 and LP or adult(22,28), we found that the expression of SGs clearly differentiates early from late stages (**Fig. 5A, Supplementary Fig. 9A**). This was true in particular for carcass and CNS that, as imaginal discs, contain mostly undifferentiated cells in L3 and experience differentiation during metamorphosis, while the fat body is already differentiated at L3 (review in (29)).

Next, we have analyzed whole body RNA-seq data from the modENCODE project which is available at much higher temporal resolution (28,30)) (**Figure 5B**). We found that 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. As expected, when measured on the whole organism, containing heterogeneous cell types in different states of differentiation, the shift between early and late gene expression can not be detected comparing L3 and LP stages (**Fig. 5B** and **Supplementary Fig 9B**). This suggests that the fly developmental transcriptional program is actually associated with cell differentiation, and that the endpoints of this program (early and late genes) correspond to undifferentiated and fully differentiated cells, rather than to specific chronological differentiation time points. In additional support of this, we have analyzed RNA-Seq data available for different cell types from the *Drosophila* adult midgut (31). We found that in undifferentiated or primordia cells (intestinal stem cells and enteroblasts, respectively) early genes are up-regulated compared to differentiated cells (enterocytes and enteroendocrine cells), in which late genes are up-regulated, instead (**Fig. Supplementary 9C**).

The fly transcriptional differentiation program is conserved in metazoans

To investigate whether the fly temporal differentiation program is conserved outside from insects, we analyzed RNA-Seq data from diverse organs at different developmental time points available for mouse, human, and worm (24,32). We identified the 1-to-many orthologs of the set of fly early and late genes (**Supplementary Table 4**) in each of these species. In the case of mouse, more than 80% of orthologs of fly early and late genes were classified identically by Cardoso-Moreira et al. (24) in at least one of the mouse tissues (**Fig. 6A**). Consistent with tissue specialization during development, also observed in *Drosophila*, while 38% of early orthologs are downregulated through differentiation in all four tissues (325 out of 850), only 11% of late orthologs are upregulated in the four tissues (42 out of 372, **Fig. 6A**). In agreement with the metabolic changes observed during fly development, early and late mouse orthologs are functionally associated to different metabolic pathways, including nucleic acid metabolism for early genes orthologs and ion transport and lipid metabolism for late ones (**Fig. 6B**).

We used self-organizing maps (SOM) to cluster the orthologous genes in each species based on the developmental gene expression data in that species (**Fig. 6C-E and Supplementary Fig. 10**). In the case of mouse, orthologs of fly early and late genes (corresponding to 850 and 372 fly genes, respectively) clearly cluster apart (**Fig. 6D**). In every tissue analyzed, the gene expression trajectory during differentiation followed a similar path, replicating that observed in the fly, with higher expression of early orthologs in the first time points of development gradually transitioning to higher expression of late orthologs in the last time points (**Fig. 6E**). While early orthologs have similar widely distributed expression patterns in early development, late orthologs show tissue-specific specialization late in development.

Next, we further investigated whether the specific associations TF-target detected in the fly GRN were conserved in the mouse. We computed the correlation of expression between orthologous TFs and orthologous targets across mouse samples for each tissue separately. Since for each fly TF-target pair there may be multiple mouse orthologs TF-target pairs, we selected the mouse TF-target pair with the closest correlation to the fly TF-target pair, irrespective of the direction of the correlation. In all tissues, we found the TF-target associations (direction and strength) strongly correlated between fly and mouse (**Fig. 6F**). We believe that the assumption that the best correlated pair is the one most likely to have kept the fly function in the mouse after the subfunctionalization and neofunctionalization, expected to occur following gene duplication, is the most sensible one. However, it may also lead to inflated correlation values. Thus, we have recomputed the correlations when considering all orthologous pairs for each fly TF-target pair. While the correlations are, as expected, weaker, they are still highly significant (**Supplementary Fig. 11**).

We found similar results when analyzing human and worm developmental expression data (**Supplementary Fig. 10**). We then identified the most conserved TF-target pairs: 77 pairs in the fly (nine TFs, 68 targets, **Supplementary Table 5**) having correlations higher than 0.5 in all species). Among these, the most conserved pairs include the Hox cofactor Exd and the TGF-Beta related factors Mad and Med as TFs. These TFs are predicted to regulate several transcriptional and chromatin factors across all metazoans, like the members of the Brahma complex: *Brm*, *Bap60* and *Dalao*, *Row*, *Glo* and *CG1620*, some splicing factors like Hel25E and B52 and some cell cycle regulators like Mapmodulin and Grp. Also Max, the cofactor of Myc, is in the list of most conserved pairs regulating *RpS30*, involved in translation, and *CG8209* (the ortholog of *UBXN1*, a general negative regulator of protein metabolism.

364 These results, all together, strongly suggest that the fly transcriptional differentiation
365 program is largely conserved within the metazoan lineage.
366

Discussion

Cell determination and differentiation are fundamental to tissue and organ formation and ultimately organism development. To characterize the molecular basis of these processes, we profiled gene expression of imaginal tissues during fly organ differentiation. In contrast to previous work (4,18,21,33,34), we specifically labeled primordial cells with GFP and tracked them along development. This allowed us to monitor changes precisely associated to particular cells while they undergo differentiation (**Fig 1A and Supplementary Fig. 1**). The data produced here, therefore, is a valuable resource to investigate transcriptomic changes in fly development.

Our analyses of this data suggest that the transition from precursor undifferentiated to terminally differentiated cells is the consequence of two, partially orthogonal, transcriptional programs. First, the general down-regulation of early genes and activation of late genes, which is common to all differentiating cells, independently from the differentiation end point. Second, the late specialized activation of genes defining tissue fate (**Fig. 7A**). The temporal differentiation program clearly dominates the spatial program, which is defined by a relatively small number of genes. This suggests that the Waddington landscape is less steep, and the valleys less deep than often assumed, contributing to explain why transdifferentiation from a terminal cell type to another can be forced with relative ease, either directly or through de-differentiation.

We have built a differentiation gene regulatory network (GRN, **Fig. 4**) to help characterize these programs from the functional standpoint. We found that genes generically downregulated during differentiation are preferentially associated with cell cycle, regulation of gene expression and translation. Remarkably, most genes (~ 60%) generically up-regulated during differentiation have not been functionally characterized yet. We identified a cluster, however, in the fly differentiation GRN (cluster 7), mainly composed of late genes involved in oxidative metabolism and ion transport. These include *Vha14-1, Vha36-1, Vha68-2, VhaAC45 and VhaM9.7* (35) that encode for subunits of vacuolar H+ ATPase; and genes related to other metabolic pathways, like *mmy* (36)*, Gs1* (37), *Mfe2* (38) among others (**Fig. 4 and Supplementary Fig. 7**). White et al. (39) previously described a transcriptional metabolic change associated with metamorphosis entrance in *Drosophila*, and hypothesized that this could reflect tissues preparing for cell death, or consequence of the transition from an active larval state to a sessile pupa one. However, recent insights in early embryogenesis and stem cell (SC) differentiation (reviewed in (40)) indicate that in mammalian embryonic SCs (ESCs)

399 oxidative capacity is reduced and glycolysis-dependent anabolic pathways are enriched
400 whereas mitochondrial function and oxidative metabolism positively correlate with SC
401 differentiation. Thus, experimental evidences based on mammalian ESCs reprogramming and
402 differentiation indicate that transcriptional programs regulating stemness influence energy
403 metabolism and metabolic enzymes ((41–44) review in (40)). In agreement with this, we found
404 that mouse orthologues of the early and late gene sets are enriched for genes associated with
405 different metabolic pathways (**Fig. 6B**). Altogether, our results suggest that the fly transcriptional
406 differentiation program may be regulating metabolic changes necessary for tissue differentiation
407 (**Fig. 7A**). Metabolomic assays to characterize the metabolic fluxes occurring in imaginal
408 tissues through differentiation would help to assess this hypothesis.

409 The fly transcriptional differentiation program, thus, appears to be under weak direct
410 epigenetic regulation. Chromatin accessibility appears quite stable and unspecific for genes
411 regulated during fly development, according to previous publication (21) (**Supplementary Fig.**
412 **5C,E)**, and it does not seem therefore to play a key direct regulatory role. However, it could still
413 play an indirect role through the regulation of certain key fate regulators in specific tissues. For
414 example, we found three eye TFs (Scrt, Oli and Hmx) showing open promoters specifically in
415 the eye. These genes are involved in eye and neural fate specification, and regulate multiple
416 genes according to our GRN. For instance, among Scrt direct targets there are essential eye
417 fate regulators like Ey, So and Pnt. Marking by H3K4me3, on the other hand, reflects mostly the
418 breadth, rather than the specificity, of gene expression **(Fig. 3**, **Supplementary Fig. 5**). We
419 specifically found, however, that genes exclusively expressed in L3, tend to maintain the mark in
420 LP. H3K4me3, thus, is not actively erased from the switched-off promoters of these genes.
421 While the consensus in the field is that H3K4me3 is a conserved hallmark of active promoters,
422 recent reports (reviewed in (45)) show that it is dispensable for gene activation; our results
423 further suggest that it may not be sufficient either to drive and/or maintain transcription.

424 In contrast, the fly GRN predicts regulatory TF-target interactions for 74% of the genes
425 regulated during differentiation (DDGs). This suggests that the combinatorial action of TFs on
426 open promoters are likely to play the leading role in the regulation of DDGs during cellular
427 differentiation (reviewed in (46,47)). Further studies on TF occupancy in promoters and
428 enhancers during imaginal tissues differentiation, as well as direct and indirect protein-protein
429 interactions between TFs, are required to fully understand the regulation of the transcriptional
430 output of DDGs.

We found that the fly transcriptional differentiation program is broadly conserved among metazoans (from worm to humans) (**Fig. 6E,D and Supplementary Fig.10**). Specifically in the fly, we found that temporal transcriptional changes common to multiple cell types, as they transition from undifferentiated to fully differentiated types, dominate over cell types specific transcriptional changes. It has been recently shown that also during mammalian organ development there is a temporal transcriptional program common across tissues (24,26). However, in this case, tissue specific transcriptional changes seem to predominate (24,26). This does not necessarily reflect a true biological phenomenon, but could partially be the consequence of the difficulties of measuring in systems of large complexity. First, mammalian organs are composed of a large number of cell types and tissues, which do not necessarily differentiate and mature in a synchronous manner, especially in late development. In contrast, the fly organs, at least those analyzed here, (eye, leg and wing), are simpler, composed by a reduced number of cell types (48–50), which, as a result of metamorphosis, mature in a more synchronized manner. Second, as vertebrates suffered several rounds of whole genome duplications, duplicated genes underwent different paths of neo- and sub-functionalization (51,52), evolving divergent(53–55) or redundant regulatory profiles (56,57) and making difficult to correctly identify TF-target interactions. In *Drosophila,* in contrast, reduced genome complexity facilitates the identification of these interactions. This can actually be seen when comparing the fly TF-target interactions in other species. Thus, in worm, with reduced genome complexity compared to human and mouse, the drop in TF-target correlations when computed over all orthologous TF-target pairs compared to the best pair is less dramatic than in the mammalian species. Thus, the relative simplicity of the fly developmental system allows for the discovery of general trends, which are obscured, and thus more difficult to detect, in more complex (mammalian) systems.

Finally, further analysis at single cell level will contribute to understanding the temporal and spatial transcriptome determinants of cellular differentiation.

Conclusions

In summary, we investigated the regulatory transcriptional programs underlying cell linages differentiation. Our results show that transcriptional changes occurring during cellular differentiation are likely to predominantly reflect the progressive loss of pluripotency and the gain of a mature metabolic state, as cells transition from undifferentiated to terminally differentiated states. These changes are common to most cell types, and dominate those underlying lineage specification and tissue specificity, which affect a relatively small number of genes. A network of TFs regulates, mainly, the cell differentiation transcriptional program and this GRN is conserved across metazoans during development. This novel gene regulatory program will help to better understanding of many differentiation events, both, *in vivo* and *in vitro*, and could contribute to the improvement of differentiation *in vitro* processes.

Methods

Drosophila melanogaster strains

Fly strains used for this study were: *nubbinGAL4*;UAS-GFP (selection of wing primordial cells), *GMRGAL4*;UAS-GFP (selection of eye cells) and *p{GAW}rnGAL4-5*;UAS-GFP (selection of leg cells and antenna cells), *enGAL4;UAS-GFP* (selection of anterior and posterior compartments in wing), *apGAL4;UAS-GFP* (selection of ventral and dorsal compartments in wing), *ubiRFP;p{GawB}C68a;UAS-GFP* (male and female genitalia discs). Flies were grown in standard media at 25ºC.

Cell sorting, RNA isolation, library preparation and sequencing

Imaginal tissues from third instar larvae (110-115h after egg laying), early-pupa (120-130h) and late pharate (225-235h) were dissected in PBS 1x and incubated for 1h in a 10x trypsin solution (Sigma T4174) at room temperature in a rotating wheel. Cells were vigorously pipetted and kept on ice in Schneider's insect medium. To discard dead cells, DAPI was added to the sample at 1 µg/mL final concentration. Cells were sorted in a FACSAria (BD) with the 85 µm nozzle at the Flow Cytometry Unit of the University Pompeu Fabra and the Centre for Genomic Regulation (UPF-CRG, Barcelona, Spain). Cells of interest were collected for subsequent analyses

(**Supplementary Fig. 1B**). RNA from sorted cells was extracted with the ZR-RNA MicroPrep Kit from Zymo Research following the manufacturer's instructions. Sequencing libraries were prepared using TruSeq Stranded mRNA Library Preparation Kit from Illumina and following the manufacturer's instructions. Sequencing was performed in a HiSeq sequencer from Illumina at the Ultrasequencing Unit of the CRG. A minimum of 50 million paired-end 75 bp-long reads were obtained per replicate and two replicates were performed per each tissue.

RNA-seq experiments

Stranded paired-end RNA-seq data for 27 samples in two bio-replicates were generated. The raw data (FASTQ), mapped data (BAM) and lists of quantified elements are available https://rnamaps.crg.cat/.

RNA-seq data processing and analysis

We processed the data using the in-house pipeline grape-nf (available at https://github.com/guigolab/grape-nf). RNA-seq reads were aligned to the fly genome (dm6) using STAR 2.4.0j(58) allowing up to 4 mismatches per paired alignment. We used the FlyBase genome annotation r6.05(59,60). Only alignments for reads mapping to ten or fewer loci were considered. On average 92% of reads were mapped and 85% of the initial number of reads were uniquely mapped to the fly dm6 genome. Of these, 92% mapped to exonic regions.

Gene and transcripts were quantified in Transcripts Per Kilobase Million (TPMs) using RSEM (61). TPM values were recomputed including only protein coding and long non coding genes (13,920 and 2,470 genes, respectively). Only genes expressed at least 5 TPMs in two samples were considered for subsequent analyses (9334 genes). TF annotation was obtained from FlyFactorSurvey (http://mccb.umassmed.edu/ffs). Plots were made using d3js (available at https://d3js.org/) and ggplot2(62) and R scripts (some available at https://github.com/abreschi/Rscripts).

Gene expression

Variance Decomposition

For each gene, the total variance in expression across samples (total sum of squares, TSSg) can be decomposed into three variances : variance across developmental stages (SSSg), variance across tissues (SSTg), and the residual variance (SSRg) as in the ANOVA type of

analysis: TSSg=SSSg+SSTg+SSSg:SSTg+SSRg(19) The relative contribution of each factor to the total variance in gene expression can then be computed as the relative proportion of each variance with respect to the total. We used a linear model, implemented using the function lm() from R using the in-house wrapper available at https://github.com/abreschi/Rscripts/blob/master/anova.R (19). The TPM matrix with both replicates per sample was used for the analysis.

Profiling the gene expression and DDGs definition

To identify genes whose expression changes across fly differentiation we used different methods to profile gene expression. First, we performed differential gene expression analysis using EdgeR v3.22.5 (63) with stages and tissues as factors. The counts matrix with both replicates per sample was used for the analysis. We required log2 fold change > 1 (at least two-fold change) and FDR < 0.01. Contrasts of every tissue and stage were used to define tissue and stage specific genes (18 contrast in total, herein called EdgeR tissue-stage). Second, to identify tissue-specific genes, EdgeR was used with tissue as factor (tissue gene profiles) and time-specific genes were classified for each tissue independently based on their trajectories (stage gene profiles: up-regulation, down-regulation, peaking or bending). Briefly, we focused on profiles with at least two-fold change and identified monotonic up-regulations and down-regulations; peaking profiles were defined as monotonic increases followed by monotonic decreases, bending profiles as the opposite (script: classification.log2.pl) (64). Third, we used the percentage of contribution from variance decomposition to identify genes for which the sum of tissue and stage contribution explains at least 70% of variation of expression.

Finally, genes classified as differentially expressed in at least two of the three methods used (edgeR tissue-stage, gene profiles, either across tissues or stages, and variance decomposition) were considered Developmental Dynamic Genes (DDGs).

In detail, EdgeR tissue-stage results were used to classify all gene sets: stage genes (differentially expressed in all the tissues in a particular stage/s), tissue genes (differentially expressed in all stages in a particular tissue/s and tissue-stage (differentially expressed in particular tissue and stage). Gene profiles through developmental stages were used to define stage genes and tissue-stage genes, while gene profiles across tissues were used to define tissue genes. Following variance decomposition classification: stage genes have stage-explained variation at least two fold greater than tissue-explained variation, tissue genes have tissue-explained variation at least two fold greater than stage-explained variation and tissue-

stage were the rest of genes above the 70% cut-off. Groups with less than 5 genes were discarded. **Supplementary Table 6** summarizes the number of genes obtained from each analysis.

Restricted vs widespread gene classification

Restricted genes show expression levels equal or higher than 5 TPMs only in the precise tissue/developmental stage where they are considered differentially expressed. The remaining differentially expressed genes are classified as widespread.

GO term enrichment analysis

The GO term enrichment analysis for biological processes hierarchy was performed separately for each set of genes, with respect to all DDGs used as background. The enrichment is tested with the hypergeometric test implemented in the R package GOstats v2.44.0 (65). FlyBase gene IDs are converted to entrez gene IDs via the R package org.Dm.eg.db v3.4.1 (Bioconductor - org.Dm.eg.db), and mapped to gene ontology through the R package GO.db v3.4.1 (Bioconductor - GO.db).

Epigenetic regulation

FAIRE-Seq data processing and classification

For each gene, we define the promoter as the sequence within a window of 250 bp upstream and downstream from the transcription start site (TSS). All TSSes annotated for the genes were considered for classification, but only the genes with all TSSes equally classified were considered for later analyses. FAIRE-Seq data was obtained from NCBI GEO database GSE38727 (21), replicates with higher signal-to-noise ratio were selected for the analysis. Data was processed using the in-house chip-nf pipeline (https://github.com/guigolab/chip-nf). Reads were continuously mapped to the fly genome (dm6) with up to two mismatches using the GEM mapper (66). Only alignments for reads mapping to 10 or fewer loci were reported. Duplicated reads were removed using Picard (http://broadinstitute.github.io/picard/). Peak calling was performed using MACS2 (67), only peaks with fdr < 0.1 were considered. Promoters were classified as specific when open chromatin peaks overlapping the promoter are present only in specific tissue and/or stage, as close when no open chromatin peaks overlap the promoter and as non specific when overlapping peaks are present in several tissues or stages.

H3K4me3 Chromatin Immunoprecipitation

Chromatin from antenna, eye, leg and wing at three stages of development (L3, EP and LP) was fixed with FA1% at RT for 10 min and sonicated with a Diagenode Bioruptor for 15 minutes at high intensity with ON/OFF alternate pulses of 30 second. Sheared chromatin was aliquoted and flash frozen in liquid nitrogen. Chromatin immunoprecipitation assays were performed following iChIP(68) protocol with some modifications. Abcam antibody Abcam_ab8895 was used to immunoprecipitate H3K4me3 attached chromatin. Data was processed using the in-house chip-nf pipeline (https://github.com/guigolab/chip-nf). Reads were continuously mapped to the fly genome (dm6) with up to two mismatches using the GEM mapper (66). Only alignments for reads mapping to 10 or fewer loci were reported. Duplicated reads were removed using Picard (http://broadinstitute.github.io/picard/). Peak calling was performed using MACS2 (67) , only peaks with fdr < 0.1 were considered. The intersection of peaks called in both replicates were used for the analysis. The intersection of such peaks with accessible promoters, described in section "FAIRE-Seq data processing and classification" was used for the epigenetic analysis of promoter regions. The intersection of peaks was performed using BEDtools (69) intersectBed v2.17.0.

Gene Regulatory network

Promoter open chromatin and motif search

As previous studies demonstrate that FAIRE-enriched regions(21,70,71) are bound by multiple regulatory factors, we used FAIRE-Seq data of fly from the same tissues and developmental stages we used for our analyses to identified putative regulators of DDGs. FAIRE-enriched regions overlapping DDGs promoters (window of 250 bp upstream and downstream from the transcription start site (TSS)) and open at least in the respective tissue and/or developmental stage where the gene is differentially expressed were selected for TF binding motif search. In the case of eye late pupa, for which data was not available, peak should be present in eye or CNS at L3 or in late pupa in any other tissue. The overlap between FAIRE-enriched regions and promoters was computed using BEDTools intersectBed v2.17.0 (69). We ran FIMO (72) using all available fly transcription factor (TF) motif matrices from MEME suite (73) against the FAIRE-enriched regions on DDGs promoters. We found conserved motifs for 238 fly TFs in the promoters of 1991 DDGs. Only TFs expressed at least 5 TPMs in two of the experiments were kept for building the network.

Sequence conservation and experimental data filtering

To predict binding sites in DDGs promoters every motif was inspected for conservation using the dm6 27-way multiple alignment (23 *Drosophila* species} sequences, house fly, *Anopheles* mosquito, honey bee and red flour beetle) and the phastCons measurement of evolutionary conservation from the UCSC Genome Browser (74,75). PhastCons scores in this window are averaged from the bigwig file with the bwtool software(76) and this average is taken as a measure of promoter sequence conservation. Alignment coverage should be at least 80% of initial fly input sequence in at least 10 species (at least one species further than *D.pseudoobscura*). Average phastCons over the motif region should be greater than 0.5.

Gene co-expression regulatory network

To build the fly gene co-expression regulatory network (GRN), we computed the correlation of expression between DDGs and potential regulatory TFs across all the samples produced here (including those from the wing compartments, the antenna and the genitalia). To generate GRN for DDGs, the R package WGCNA was used (27). We used expression values of DDGs and 775 fly TFs across the 27 samples generated in this study. Using default parameters of WGCNA package, that is: hierarchical clustering (hclust R function) and Dynamic tree cut R package (77) we identified fifteen clusters of expression (clusters with Pearson's correlation coefficient higher than 0.85 were merged), and interactions were filtered first by coefficient of correlation (connection weight higher than 0.1) and then, by presence of a TF conserved motif in the accessible TSS promoter of the DDG. Software Cytoscape 3.8.0 was used for network visualization (78). Nodes were displayed according to Edge-weighted Spring-Embedded Layout analysis of TF-target correlation of expression (Pearson's correlation coefficient between TF and target expression, averaged between replicates, calculated across the 27 samples) or the odes Betweenness centrality (measure of the amount of influence a node has over the flow of information in a graph calculated as the number of times a node acts as a bridge along the shortest path between two other nodes). Node size was adjusted depending on node closeness centrality (average shortest path length from the node to every other node in the network, it indicates how close a node is to all other nodes). Edges were colored according to TF-target correlation of expression (Pearson's correlation coefficient mentioned above). Edge transparency was adjusted depending on TF-target weight (similarity measure considering levels of expression, averaged between replicates, across the 27 samples used for network generation).

Conservation of differentiation regulatory program in metazoans

Fly gene identifiers were mapped to mouse, human, and worm orthologs using Ensembl79 (http://mar2015.archive.ensembl.org) (79). Genes mapping to one or more orthologs in each species were analyzed in a fly-oriented manner.

Transcriptional profiling of mouse and human were obtained from ArrayExpress (E-MTAB-6798, and E-MTAB-6814) (24). Worm data was obtained from NCBI BioProject database (PRJNA477006) (32). The profile of gene expression of orthologs of *Drosophila* early and late genes along tissue development was analyzed using self-organizing maps (R package kohonen v3.0.8) (80,81). Orthologs that mapped into both early and late fly genes were excluded from all analyses. We compared the profiles of gene expression of fly early and late genes with the mouse orthologs based on the gene profile classification provided by the authors (24).

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and analyzed during the current study from fly are available in the Array Express (https://www.ebi.ac.uk/arrayexpress/) repository, under accession number E-MTAB-10879 (for RNA-Seq samples) (http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-10879); and under accession number E-MTAB-11307 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-11307) (for ChIP-Seq data).

All the fly generated data is also available through the RNAmaps dashboard (https://rnamaps.crg.cat/).

The datasets analyzed during the current study from fly modENCODE are available in modENCODE portal repository (www.modencode.org) (28,30).

The datasets analyzed during the current study from fly midgut cell types is available in NCBI GEO repository under accession number GSE61361 (31).

The datasets analyzed during the current study from fly FAIRE-Seq data is available in NCBI GEO repository under accession number GSE38727 (21).

694 The datasets analyzed during the current study from mouse and human development are
695 available in the from ArrayExpress (https://www.ebi.ac.uk/arrayexpress/) repository, under
696 accession number E-MTAB-6798 and E-MTAB-6814 (24).
697 The dataset analyzed during the current study from worm development is available in the NCBI
698 BioProject database repository under accession number PRJNA477006 (32).
699

700 **Competing interests**

701 The authors declare that they have no competing interests.

702 **Funding**

703 This work was supported by the European Community under the FP7 program (ERC-2011-
704 AdG-294653-RNA-MAPS to R.G.), by the Spanish Ministry of Economy, Industry and
705 Competitiveness (MEIC) (BIO2011-26205 to R.G) to the EMBL partnership and the Centro de
706 Excelencia Severo Ochoa and by the CERCA Programme (Generalitat de Catalunya).

707 **Author's contributions**

708 R.G. conceived the project. M.R-R. and R.G designed the study. C.C.K, M.R-R. and A.B.
709 performed the computational analyses. S.P-L., M.R-R. and A.A. performed the RNA-Seq
710 experiments. S.P-L. and M.R-R. performed the ChIP-Seq experiments. M.R-R., C.C.K and R.G.
711 wrote the manuscript with the contribution of all authors. All authors read and approved the final
712 manuscript.

713 **Acknowledgements**

714 We thank Emilio Palumbo for his help with pipeline development and data processing, Bruna R.
715 Correa, Ramil Nurtdinov and Beatrice Borsari for helpful discussions about the data and the
716 manuscript and Romina Garrido for administrative assistance. We also thank the CRG
717 Genomics Unit and the CRG/UPF Flow Cytometry Unit (Barcelona, Spain).

718

References

1. Waddington CH. The Strategy of the Genes: a discussion of some aspects of theoretical biology [Internet]. The Strategy of the Genes. 1957 [cited 2020 Jul 10]. Available from: https://books.google.es/books?id=odl1AwAAQBAJ&source=gbs_similarbooks
2. Rajagopal J, Stanger BZ. Plasticity in the Adult: How Should the Waddington Diagram Be Applied to Regenerating Tissues? Developmental Cell. 2016.
3. Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, et al. Spatio-temporal transcriptome of the human brain. Nature. 2011;478(7370):483–9.
4. Li JJ, Huang H, Bickel PJ, Brenner SE. Comparison of D. melanogaster and C. elegans developmental stages, tissues, and cells by modENCODE RNA-seq data. Genome Res [Internet]. 2014 Jul [cited 2019 Aug 6];24(7):1086–101. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24985912
5. van de Leemput J, Boles NC, Kiehl TR, Corneo B, Lederman P, Menon V, et al. CORTECON: A temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells. Neuron [Internet]. 2014 Jul 2 [cited 2020 Jul 10];83(1):51–68. Available from: https://pubmed.ncbi.nlm.nih.gov/24991954/
6. Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, et al. Genetic effects on gene expression across human tissues. Nature [Internet]. 2017 Oct 11 [cited 2020 Jul 10];550(7675):204–13. Available from: www.gtexportal.org
7. Li Y, Wang R, Qiao N, Peng G, Zhang K, Tang K, et al. Transcriptome analysis reveals determinant stages controlling human embryonic stem cell commitment to neuronal cells. J Biol Chem [Internet]. 2017 Dec 1 [cited 2020 Jul 10];292(48):19590–604. Available from: https://pubmed.ncbi.nlm.nih.gov/28972157/
8. Li B, Qing T, Zhu J, Wen Z, Yu Y, Fukumura R, et al. A Comprehensive Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq. Sci Rep [Internet]. 2017 Dec 1 [cited 2020 Jul 10];7(1):1–10. Available from: www.nature.com/scientificreports
9. Semrau S, Goldmann JE, Soumillon M, Mikkelsen TS, Jaenisch R, Van Oudenaarden A. Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. Nat Commun [Internet]. 2017 Dec 1 [cited 2020 Jul 10];8(1):1–16. Available from: www.nature.com/naturecommunications
10. García-Bellido A. Pattern reconstruction by dissociated imaginal disk cells of Drosophila melanogaster. Dev Biol. 1966;
11. Klebes A, Biehs B, Cifuentes F, Kornberg TB. Expression profiling of Drosophila imaginal

discs. Genome Biol [Internet]. 2002 [cited 2020 Jul 10];3(8):research0038.1. Available from: /pmc/articles/PMC126232/?report=abstract

12. Potier D, Davie K, Hulselmans G, NavalSanchez 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;

13. Jasper H, Benes V, Atzberger A, Sauer S, Ansorge W, Bohmann D. A genomic switch at the transition from cell proliferation to terminal differentiation in the Drosophila eye. Dev Cell [Internet]. 2002 Oct 1 [cited 2020 Jul 10];3(4):511–21. Available from: http://www.cell.com/article/S1534580702002976/fulltext

14. Michaut L, Flister S, Neeb M, White KP, Certa U, Gehring WJ. Analysis of the eye developmental pathway in Drosophila using DNA microarrays. Proc Natl Acad Sci U S A [Internet]. 2003 Apr 1 [cited 2020 Jul 10];100(7):4024–9. Available from: www.ncbi.nlm.nih.govgeo

15. Pérez-Lluch S, Blanco E, Tilgner H, Curado J, Ruiz-Romero M, Corominas M, et al. Absence of canonical marks of active chromatin in developmentally regulated genes. Nat Genet. 2015;47(10).

16. O'keefe DD, Thomas SR, Bolin K, Griggs E, Edgar BA, Buttitta LA. Combinatorial control of temporal gene expression in the Drosophila wing by enhancers and core promoters [Internet]. 2012 [cited 2020 Jul 10]. Available from: http://www.biomedcentral.com/1471-2164/13/498

17. Arbeitman MN, Krasnow MA, Furlong EEM, Imam F, Johnson E, Null BH, et al. Gene expression during the life cycle of Drosophila melanogaster. Science (80-). 2002;

18. Ren N, Zhu C, Lee H, Adler PN. Gene expression during Drosophila wing morphogenesis and differentiation. Genetics. 2005;171(2):625–38.

19. Breschi A, Djebali S, Gillis J, Pervouchine DD, Dobin A, Davis CA, et al. Gene-specific patterns of expression variation across organs and species. Genome Biol [Internet]. 2016 [cited 2019 Aug 5];17(1):151. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27391956

20. Robinson MD, Mccarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinforma Appl NOTE [Internet]. 2010 [cited 2020 Jul 10];26(1):139–40. Available from: http://bioconductor.org

21. McKay DJ, Lieb JD. A Common Set of DNA Regulatory Elements Shapes Drosophila Appendages. Dev Cell. 2013;

22. Negre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, et al. A cis-regulatory map

of the Drosophila genome. Nature [Internet]. 2011;471(7339):527–31. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=21430782

23. Hödl M, Basler K. Transcription in the absence of histone H3.2 and H3K4 methylation. Curr Biol [Internet]. 2012 Dec 4 [cited 2021 Sep 9];22(23):2253–7. Available from: https://pubmed.ncbi.nlm.nih.gov/23142044/

24. Cardoso-Moreira M, Halbert J, Valloton D, Velten B, Chen C, Shao Y, et al. Gene expression across mammalian organ development. Nature. 2019;

25. Janssens J, Aibar S, Taskiran II, Ismail JN, Gomez AE, Aughey G, et al. Decoding gene regulation in the fly brain. Nat 2022 6017894 [Internet]. 2022 Jan 5 [cited 2022 Feb 3];601(7894):630–6. Available from: https://www.nature.com/articles/s41586-021-04262-z

26. He P, Williams BA, Trout D, Marinov GK, Amrhein H, Berghella L, et al. The changing mouse embryo transcriptome at whole tissue and single-cell resolution. Nature [Internet]. 2020 Jul 30 [cited 2020 Sep 8];583(7818):760–7. Available from: https://doi.org/10.1038/s41586-020-2536-x

27. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics. 2008 Dec 29;9:559.

28. Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, et al. The developmental transcriptome of Drosophila melanogaster. Nature [Internet]. 2012;471(7339):473–9. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=21179090

29. Yafei Zhang YX. Fat Body Development and its Function in Energy Storage and Nutrient Sensing in Drosophila melanogaster. J Tissue Sci Eng. 2014;06(01).

30. Roy S, Ernst J, Kharchenko P V., Kheradpour P, Negre N, Eaton ML, et al. Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science (80-) [Internet]. 2010 Dec 24 [cited 2020 Sep 16];330(6012):1787–97. Available from: www.sciencemag.org/cgi/content/science.1196914/DC1

31. Dutta D, Dobson AJ, Houtz PL, Gläßer C, Revah J, Korzelius J, et al. Regional Cell-Specific Transcriptome Mapping Reveals Regulatory Complexity in the Adult Drosophila Midgut. Cell Rep [Internet]. 2015 Jul 14 [cited 2020 Sep 16];12(2):346–58. Available from: https://pubmed.ncbi.nlm.nih.gov/26146076/

32. Warner AD, Gevirtzman L, Hillier LDW, Ewing B, Waterston RH. The C. elegans embryonic transcriptome with tissue, time, and alternative splicing resolution. Genome

Res. 2019;29(6):1036–45.

33. White K, Grether ME, Abrams JM, Young L, Farrell K, Steller H. Genetic control of programmed cell death in Drosophila. Science (80-) [Internet]. 1994;264(5159):677–83. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8171319

34. Quiquand M, Rimesso G, Qiao N, Suo S, Zhao C, Slattery M, et al. New regulators of Drosophila eye development identified from temporal transcriptome changes. Genetics [Internet]. 2021 Apr 15 [cited 2022 Feb 4];217(4). Available from: /pmc/articles/PMC8049564/

35. Allan AK, Du J, Davies SA, Dow JAT. Genome-wide survey of V-ATPase genes in Drosophila reveals a conserved renal phenotype for lethal alleles [Internet]. Vol. 22, Physiological Genomics. Physiol Genomics; 2005 [cited 2020 Sep 17]. p. 128–38. Available from: https://pubmed.ncbi.nlm.nih.gov/15855386/

36. Humphreys GB, Jud MC, Monroe KM, Kimball SS, Higley M, Shipley D, et al. Mummy, A UDP-N-acetylglucosamine pyrophosphorylase, modulates DPP signaling in the embryonic epidermis of Drosophila. Dev Biol [Internet]. 2013 Sep 15 [cited 2020 Sep 17];381(2):434–45. Available from: /pmc/articles/PMC3775589/?report=abstract

37. Caggese C, Barsanti P, Viggiano L, Bozzetti MP, Caizzi R. Genetic, molecular and developmental analysis of the glutamine synthetase isozymes of Drosophila melanogaster. Genetica [Internet]. 1994 Jun [cited 2020 Sep 17];94(2–3):275–81. Available from: https://pubmed.ncbi.nlm.nih.gov/7896146/

38. Haataja TJK, Koski MK, Hiltunen JK, Glumoff T. Peroxisomal multifunctional enzyme type 2 from the fruitfly: Dehydrogenase and hydratase act as separate entities, as revealed by structure and kinetics. Biochem J [Internet]. 2011 May 1 [cited 2020 Sep 17];435(3):771–81. Available from: https://pubmed.ncbi.nlm.nih.gov/21320074/

39. White KP, Rifkin SA, Hurban P, Hogness DS. Microarray analysis of Drosophila development during metamorphosis. Science (80-) [Internet]. 1999 Dec 10 [cited 2020 Jul 10];286(5447):2179–84. Available from: http://science.sciencemag.org/

40. Folmes CDL, Dzeja PP, Nelson TJ, Terzic A. Metabolic plasticity in stem cell homeostasis and differentiation [Internet]. Vol. 11, Cell Stem Cell. NIH Public Access; 2012 [cited 2020 Sep 18]. p. 596–606. Available from: /pmc/articles/PMC3593051/?report=abstract

41. Folmes CDL, Nelson TJ, Martinez-Fernandez A, Arrell DK, Lindor JZ, Dzeja PP, et al.

Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming. Cell Metab [Internet]. 2011 Aug 3 [cited 2022 Feb 3];14(2):264–71. Available from: https://pubmed.ncbi.nlm.nih.gov/21803296/

42. Panopoulos AD, Yanes O, Ruiz S, Kida YS, Diep D, Tautenhahn R, et al. The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. Cell Res [Internet]. 2012 Jan [cited 2022 Feb 3];22(1):168–77. Available from: https://pubmed.ncbi.nlm.nih.gov/22064701/

43. Tormos K V., Anso E, Hamanaka RB, Eisenbart J, Joseph J, Kalyanaraman B, et al. Mitochondrial complex III ROS regulate adipocyte differentiation. Cell Metab [Internet]. 2011 Oct 5 [cited 2022 Feb 3];14(4):537–44. Available from: https://pubmed.ncbi.nlm.nih.gov/21982713/

44. Chung S, Dzeja PP, Faustino RS, Perez-Terzic C, Behfar A, Terzic A. Mitochondrial oxidative metabolism is required for the cardiac differentiation of stem cells. Nat Clin Pract Cardiovasc Med [Internet]. 2007 Feb [cited 2022 Feb 3];4 Suppl 1(Suppl 1). Available from: https://pubmed.ncbi.nlm.nih.gov/17230217/

45. Morgan MAJ, Shilatifard A. Reevaluating the roles of histone-modifying enzymes and their associated chromatin modifications in transcriptional regulation. Nat Genet [Internet]. 2020 Dec 1 [cited 2021 Sep 9];52(12):1271–81. Available from: https://pubmed.ncbi.nlm.nih.gov/33257899/

46. Spitz F, Furlong EEM. Core promoter Transcription factors: from enhancer binding to developmental control. Nat Publ Gr [Internet]. 2012 [cited 2021 Jul 8];13:613. Available from: www.nature.com/reviews/genetics

47. Inukai S, Kock KH, Bulyk ML. Transcription factor–DNA binding: beyond binding site motifs. Curr Opin Genet Dev [Internet]. 2017 Apr 1 [cited 2022 Feb 3];43:110. Available from: /pmc/articles/PMC5447501/

48. Cagan R. Chapter 5 Principles of Drosophila Eye Differentiation. Vol. 89, Current Topics in Developmental Biology. NIH Public Access; 2009. p. 115–35.

49. Currie DA, Milnert MJ, Evans CW. The growth and differentiation in vitro of leg and wing imaginal disc cells from Drosophila melanogaster. Vol. 102, Development. 1988.

50. Kiger JA, Natzle JE, Kimbrell DA, Paddy MR, Kleinhesselink K, Green MM. Tissue Remodeling During Maturation Of The Drosophila Wing. Dev Biol [Internet]. 2007 Jan 1 [cited 2022 Feb 3];301(1):178. Available from: /pmc/articles/PMC1828914/

51. DeLuna A, Vetsigian K, Shoresh N, Hegreness M, Colón-González M, Chao S, et al. Exposing the fitness contribution of duplicated genes. Nat Genet. 2008 May

13;40(5):676–81.

52. Otto SP, Yong P. 16 The evolution of gene duplicates. Vol. 46, Advances in Genetics. Academic Press; 2002. p. 451–83.

53. Gu X, Zhang Z, Huang W. Rapid evolution of expression and regulatory divergences after yeast gene duplication. Proc Natl Acad Sci U S A. 2005 Jan 18;102(3):707–12.

54. Acharya D, Ghosh TC. Global analysis of human duplicated genes reveals the relative importance of whole-genome duplicates originated in the early vertebrate evolution. BMC Genomics. 2016 Jan 22;17(1).

55. Li WH, Yang J, Gu X. Expression divergence between duplicate genes [Internet]. Vol. 21, Trends in Genetics. 2005 [cited 2020 Apr 24]. p. 602–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16140417

56. Dean EJ, Davis JC, Davis RW, Petrov DA. Pervasive and Persistent Redundancy among Duplicated Genes in Yeast. McVean G, editor. PLoS Genet [Internet]. 2008 Jul 4 [cited 2020 Apr 24];4(7):e1000113. Available from: http://dx.plos.org/10.1371/journal.pgen.1000113

57. Vavouri T, Semple JI, Lehner B. Widespread conservation of genetic redundancy during a billion years of eukaryotic evolution. Vol. 24, Trends in Genetics. 2008. p. 485–8.

58. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics [Internet]. 2013 Jan [cited 2020 Sep 16];29(1):15–21. Available from: https://pubmed.ncbi.nlm.nih.gov/23104886/

59. Attrill H, Falls K, Goodman JL, Millburn GH, Antonazzo G, Rey AJ, et al. FlyBase: establishing a Gene Group resource for Drosophila melanogaster. Nucleic Acids Res [Internet]. 2016 [cited 2022 Feb 3];44(D1):D786–92. Available from: https://pubmed.ncbi.nlm.nih.gov/26467478/

60. Larkin A, Marygold SJ, Antonazzo G, Attrill H, dos Santos G, Garapati P V, et al. FlyBase: updates to the Drosophila melanogaster knowledge base. Nucleic Acids Res [Internet]. 2021 Jan 8 [cited 2021 Oct 1];49(D1):D899–907. Available from: https://academic.oup.com/nar/article/49/D1/D899/5997437

61. Li B, Dewey CN. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics [Internet]. 2011 Aug 4 [cited 2020 Sep 16];12(1):323. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-323

62. Wickham H. ggplot2. ggplot2. Springer New York; 2009.

63. Chen Y, Mccarthy D, Ritchie M, Robinson M, Smyth GK. edgeR□: differential expression

analysis of digital gene expression data User ' s Guide. 2016;(April).

64. Borsari B, Abad A, C. Klein C, Nurtinov R, Esteban A, Palumbo E, et al. Dynamics of gene expression and chromatin marking during cell state transition. [cited 2021 Sep 9]; Available from: https://www.biorxiv.org/content/10.1101/2020.11.20.391524v1.full.pdf

65. Falcon S, Gentleman R. Using GOstats to test gene lists for GO term association. Bioinformatics [Internet]. 2007 [cited 2022 Feb 4];23(2):257–8. Available from: https://pubmed.ncbi.nlm.nih.gov/17098774/

66. Marco-Sola S, Sammeth M, Guigó R, Ribeca P. The GEM mapper: fast, accurate and versatile alignment by filtration. Nat Methods [Internet]. 2012 Dec [cited 2022 Feb 4];9(12):1185–8. Available from: https://pubmed.ncbi.nlm.nih.gov/23103880/

67. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol [Internet]. 2008 Sep 17 [cited 2021 Dec 9];9(9). Available from: https://pubmed.ncbi.nlm.nih.gov/18798982/

68. Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, Zaretsky I, Jaitin DA, David E, et al. Immunogenetics. Chromatin state dynamics during blood formation. Science [Internet]. 2014 [cited 2020 Sep 16];345(6199):943–9. Available from: /pmc/articles/PMC4412442/?report=abstract

69. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics [Internet]. 2010 Mar 15 [cited 2021 Dec 9];26(6):841–2. Available from: https://academic.oup.com/bioinformatics/article/26/6/841/244688

70. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res [Internet]. 2007 Jun 1 [cited 2021 May 17];17(6):877–85. Available from: http://www.genome.org/cgi/doi/10.1101/gr.5533506.;www.genome.orgwww.genome.org

71. Song M, Zhang Y, Katzaroff AJ, Edgar BA, Buttitta L. Hunting complex differential gene interaction patterns across molecular contexts. Nucleic Acids Res [Internet]. 2014 Jan 29 [cited 2014 Feb 21]; Available from: http://www.ncbi.nlm.nih.gov/pubmed/24482443

72. Grant CE, Bailey TL, Noble WS. FIMO: Scanning for occurrences of a given motif. Bioinformatics. 2011;

73. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res [Internet]. 2015 Jul 1 [cited 2021 Oct 1];43(W1):W39–49. Available from: https://academic.oup.com/nar/article/43/W1/W39/2467905

74. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al.

Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res [Internet]. 2005 Aug 1 [cited 2020 Sep 16];15(8):1034–50. Available from: www.genome.org

75. C T, GP B, J C, H C, M D, C E, et al. The UCSC Genome Browser database: 2017 update. Nucleic Acids Res [Internet]. 2017 [cited 2021 Oct 1];45(D1). Available from: https://pubmed.ncbi.nlm.nih.gov/27899642/

76. Pohl A, Beato M. B: A tool for bigWig files. Bioinformatics [Internet]. 2014 Jun 1 [cited 2020 Sep 16];30(11):1618–9. Available from: http://cromatina.crg.eu/

77. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics [Internet]. 2008 Mar [cited 2022 Feb 3];24(5):719–20. Available from: https://pubmed.ncbi.nlm.nih.gov/18024473/

78. 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 [Internet]. 2003 Nov [cited 2020 Sep 16];13(11):2498–504. Available from: /pmc/articles/PMC403769/?report=abstract

79. Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, et al. Ensembl 2016. Nucleic Acids Res [Internet]. 2016 [cited 2020 Sep 16];44(D1):D710–6. Available from: /pmc/articles/PMC4702834/?report=abstract

80. Wehrens R, Buydens LMC. Self- and super-organizing maps in R: The kohonen package. J Stat Softw [Internet]. 2007 Oct 8 [cited 2020 Sep 16];21(5):1–19. Available from: https://www.jstatsoft.org/index.php/jss/article/view/v021i05/v21i05.pdf

81. Wehrens R, Kruisselbrink J. Flexible self-organizing maps in kohonen 3.0. J Stat Softw [Internet]. 2018 Oct 1 [cited 2020 Sep 16];87(7):1–18. Available from: https://www.jstatsoft.org/index.php/jss/article/view/v087i07/v87i07.pdf

Figure legends

Fig. 1 Transcriptional profiling of *Drosophila melanogaster* imaginal discs. **(A)** Overview of tissues, wing compartments and developmental stages profiled along this work. **(B)** Workflow of the RNA-Seq data generation. Briefly, imaginal tissues are manually dissected, disaggregated with trypsin treatment and sorted to collect the cells of interest (i.e., the precursor cells that differentiate into the adult tissues) from each imaginal disc. After that, RNA is extracted and processed for library preparation. **(C)** Expression profiles of genes with known tissue or stage specific expression patterns. **(D)** Principal component analysis (PCA) based on the expression of the 1,000 most variable genes across tissues and developmental stages. Gene expression is computed as log10-normalized Transcripts Per Kilobase Million (TPMs) with pseudocount of 0.01. Only genes with at least 5 TPMs in at least two samples were considered. PC1 separates the early and the late stages. PC2 separates neural and non-neural tissues. PC3 separates the late stage in the eye from the rest of the samples.

Fig 2. Gene expression dynamics along fly development. **(A)** Proportion of the variance in gene expression explained by tissue (x-axis) and by developmental stage (y-axis). Each dot corresponds to one of the 9,334 genes that are expressed at least 5 TPMs in at least two samples). Genes changing expression across tissues and/or developmental stages (Developmentally Dynamic Genes, DDGs) along fly development are highlighted in grey (differentially expressed across stages, SGs), pink (differentially expressed across tissues, TGs) and purple (differentially expressed across tissues and stages, TSGs). (**B)** Proportion of gene expression variance explained by tissue, stage and the interaction between the two. **(C)** Expression of Developmentally Dynamic Genes (DDGs) along fly development. Gene expression values are normalized to z-score values. **(D)** Dynamics of genes differentially expressed across tissues (TGs) and across tissues and stages (TSGs) represented as a Sankey diagram. At each developmental stage, we represent the genes that are differentially expressed in each tissue and those that are not (yellow). The arrows represent the number of genes that transition from one developmental stage to the next one and from not differentially to differentially expressed in a given tissue (or vice versa). Many tissue specific genes are already differentially expressed at L3 (TGs), but many which are not differentially expressed at this developmental stage become tissue specific at LP (TSGs). There is, in particular, a large expansion of eye specific genes. Overall, the transcriptome diverges as tissues become specified.

1029

Fig. 3. Regulation of DDGs. **(A)** Epigenetic regulation of DDGs. The innermost circle labels genes according to DDGs classification; the second circle displays the breadth of gene expression (profile class), and the third circle displays H3K4me3 marking. **(B** and **C)** Percentage of TFs in DDGs categories.

Fig. 4. *Drosophila* gene regulatory network (GRN). **(A)** GRN. Edges are colored according to TF-target Pearson's correlation coefficient (red > 0.3, blue < -0.3). While most correlations are positive, negative interactions are predicted between early and late gene clusters as well as between clusters corresponding to different tissues. Node size reflects node closeness centrality (that is, how close a node is to all other nodes). Nodes are colored according to DDG classification. **(B)** Network clusters. Proportion of DDG categories and TFs in each cluster. The class column indicates the most abundant DDGs category within each cluster. **(C)** GRN. Nodes are colored according to the clusters to which the genes belong and the GO categories associated. **(D)** Connectivity between clusters. Arrows indicate directionality from TFs to targets. The width of the arrows is proportional to the number of TF-target pairs. For ease in interpretation, we have included inner colored bars, representing the target's clusters.

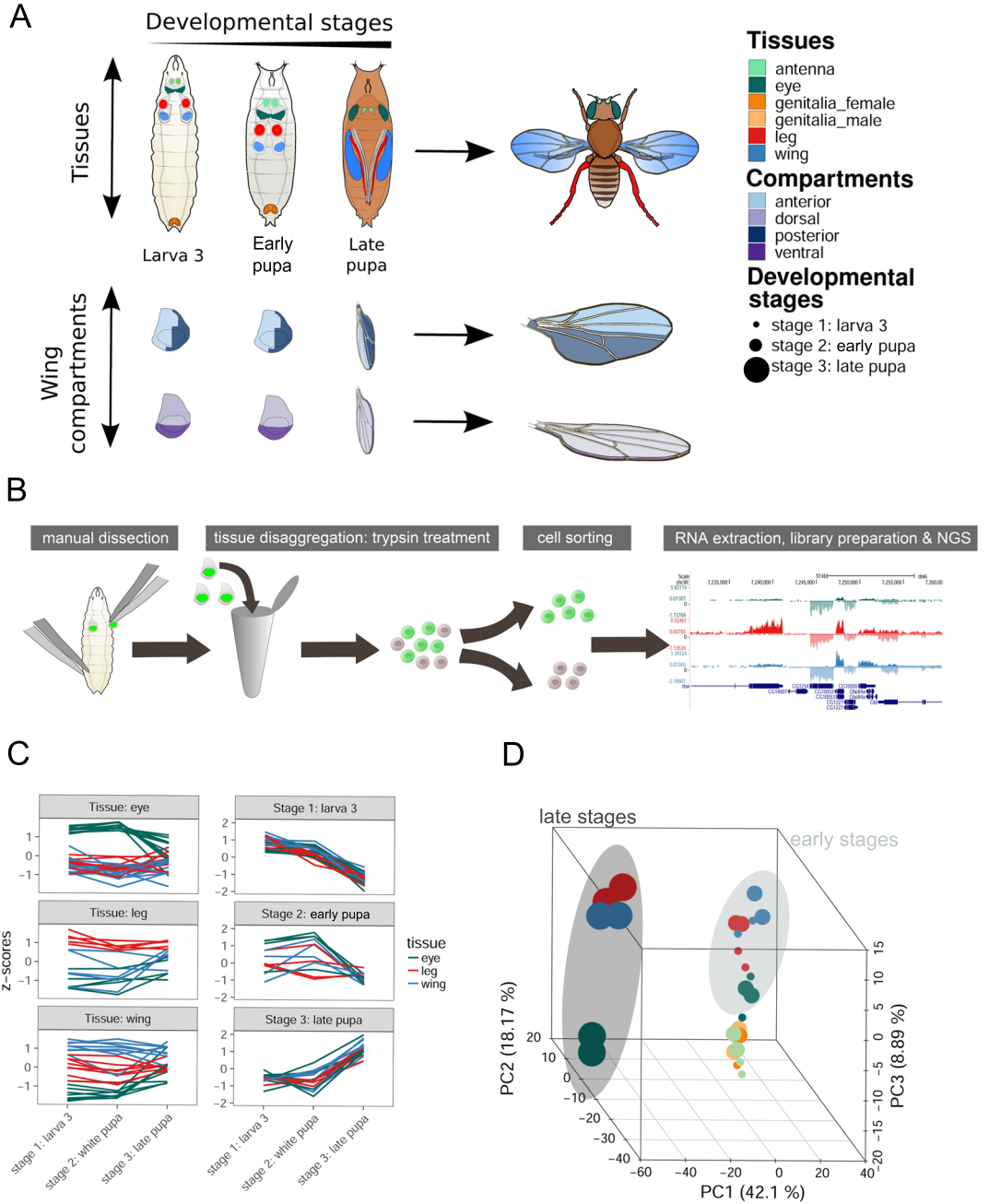
Figure 5. Gene expression dynamics of SGs in tissues and whole animals during fly development. (A) Principal component analysis (PCA) of SGs based on modENCODE tissue data (CNS and fat body from L3 and LP and carcass from L3 and adult) and the imaginal tissue data produced here at L3, EP and LP. **(B)** Expression of SGs in modENCODE whole animals from embryo to adult stages, and in the imaginal tissues monitored here at L3, EP and LP.

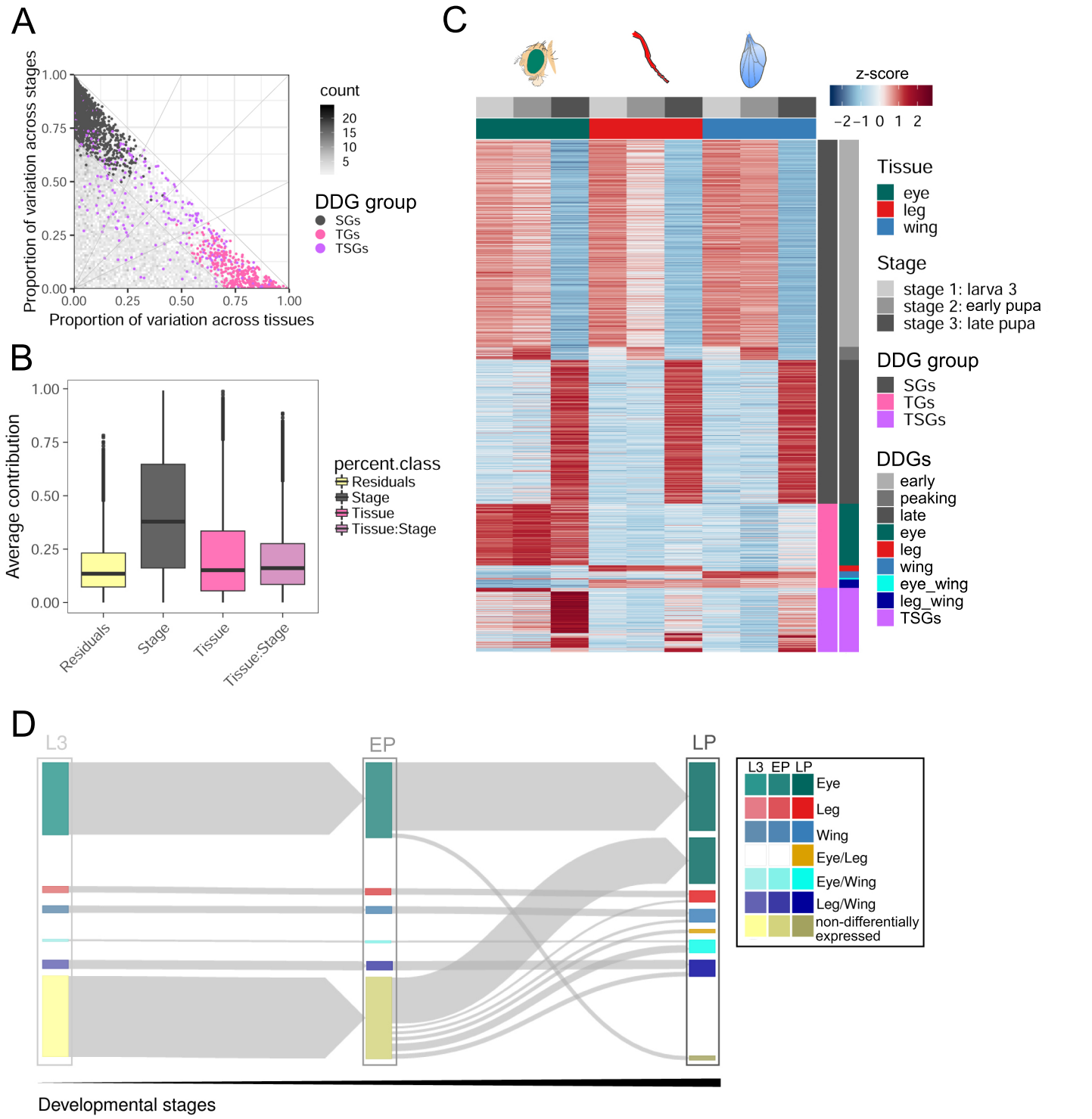
Fig. 6. Conservation of the *Drosophila* GRN in mouse. (A) Venn diagrams (left panels) showing the number of fly early and late genes, the orthologs of which are also classified as early and late during mouse development according to Cardoso-Moreira et al. (*24*) The bar plots (right panels) show the early and late orthologs classified by expression profile in mouse tissues according to Cardoso-Moreira et al.(*24*) **(B)** GO term enrichment analysis of early and late orthologs in mouse. **(C)** Mouse orthologs of fly early and late genes are clustered using self-organizing maps (SOM) based on the RNA-seq derived expression from a number of organs during mouse development (brain, heart, kidney and liver from embryo 10.5/11.5 days through 3 days post-natal stage). **(D)** SOM clustering of mouse orthologs of fly early and late genes. Each cell corresponds to a gene cluster. We are considering 12x12 cells, each containing between 2 and 23 genes. The grey intensity of cells denotes the proportion of early (light grey) versus late

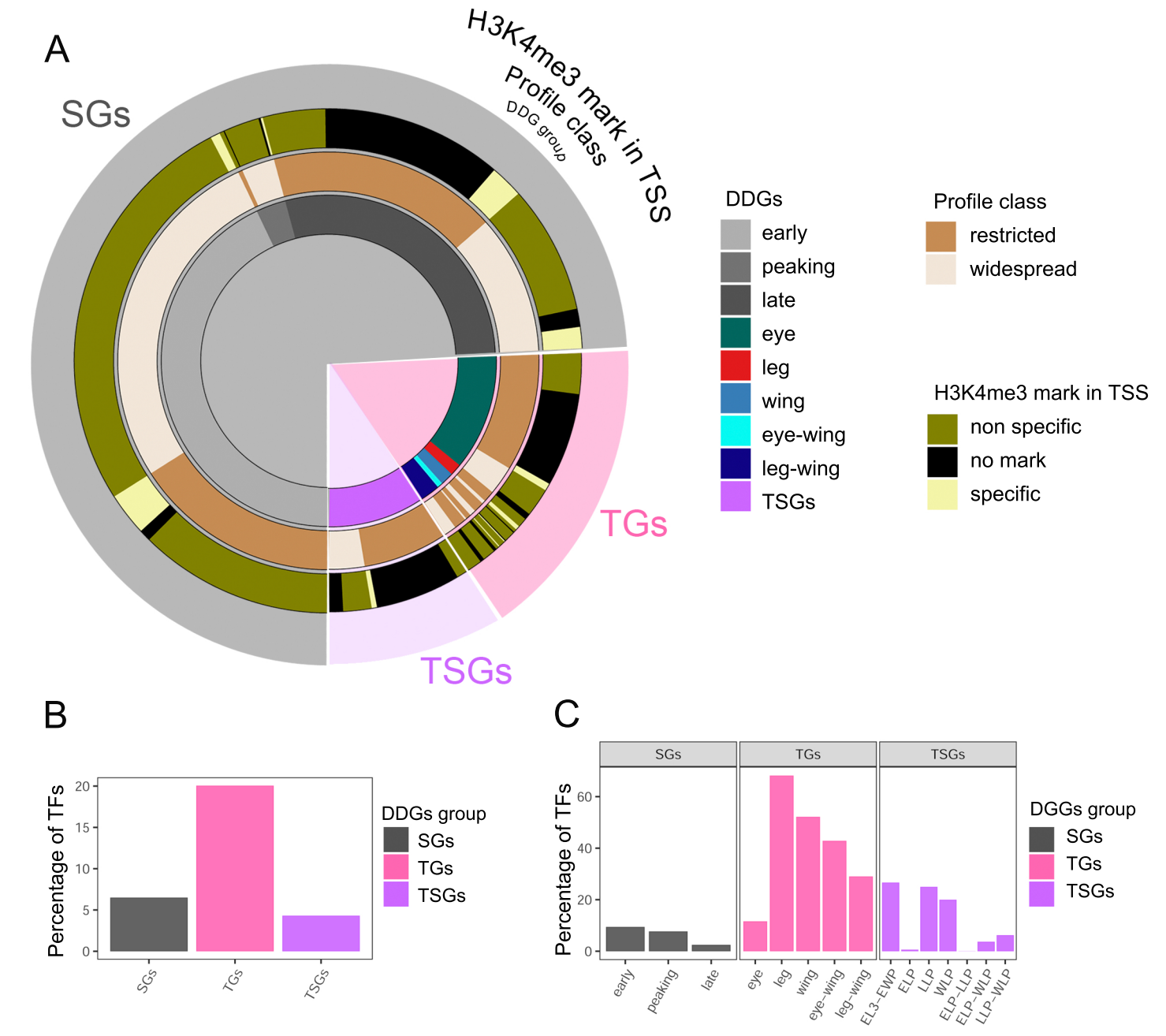
(dark grey) genes in the cell/cluster. Early and late orthologs cluster separately regarding gene expression along mouse organ development. Early genes cluster preferentially on the left part of the SOM representation, while late genes cluster on the right. **(E)** Changes in expression along mouse development in brain, heart, kidney and liver of early and late orthologs. **(F)** Scatter plots of the correlation in fly and in mouse of TF-target pair. The correlations have been computed independently in each fly-mouse ortholog tissue. When multiple mouse orthologous targets are found for the same orthologous TF, the TF-target pair with the closest correlation is employed. Pearson's correlation coefficient (cc) between fly and mouse correlations is shown on the top of the plot, p-values of the correlations are 1.09^{e-317} for brain and 0 for heart, kidney and liver. (Figure partially created with BioRender.com).

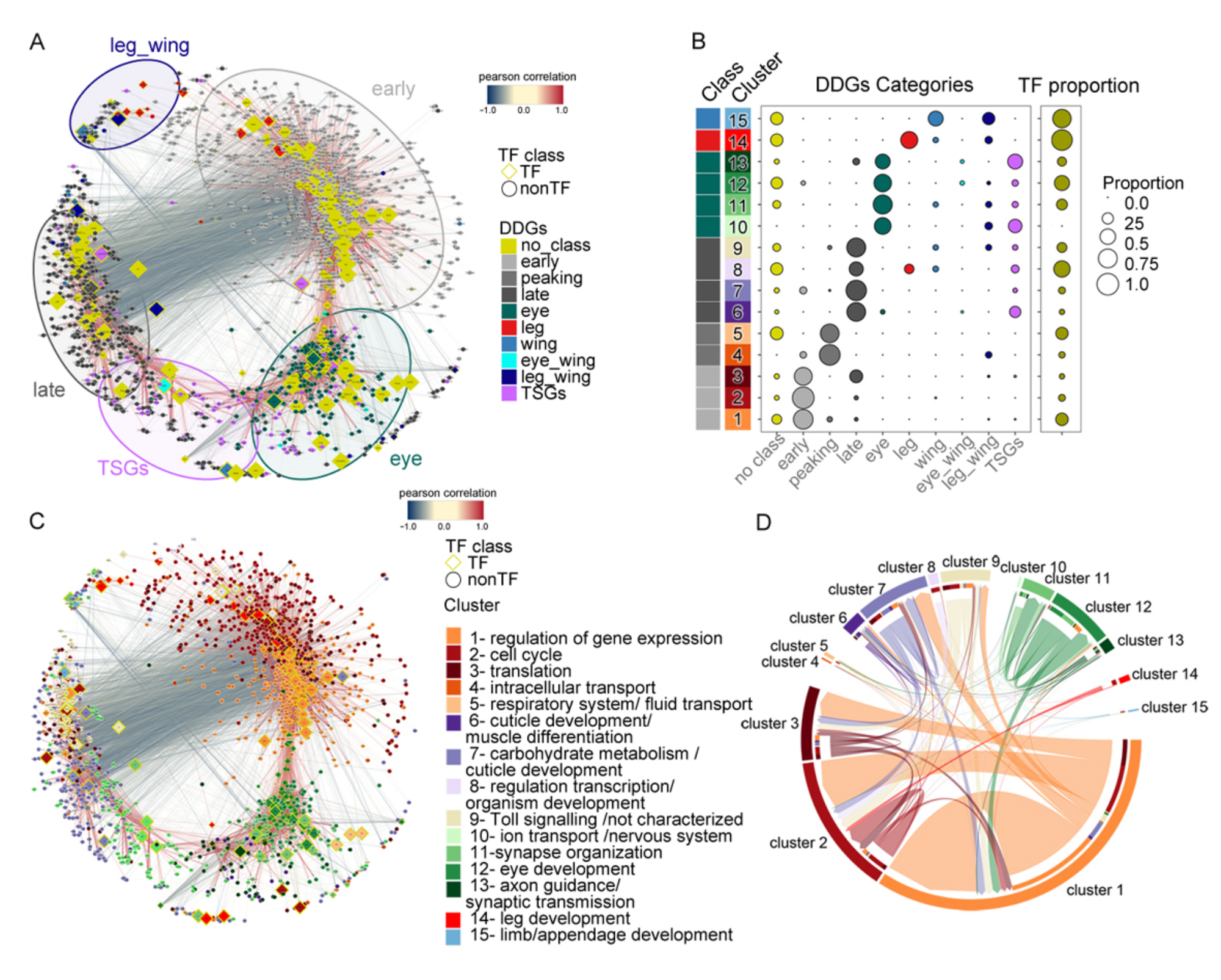
Fig. 7. Model of gene regulation in *Drosophila* differentiation.

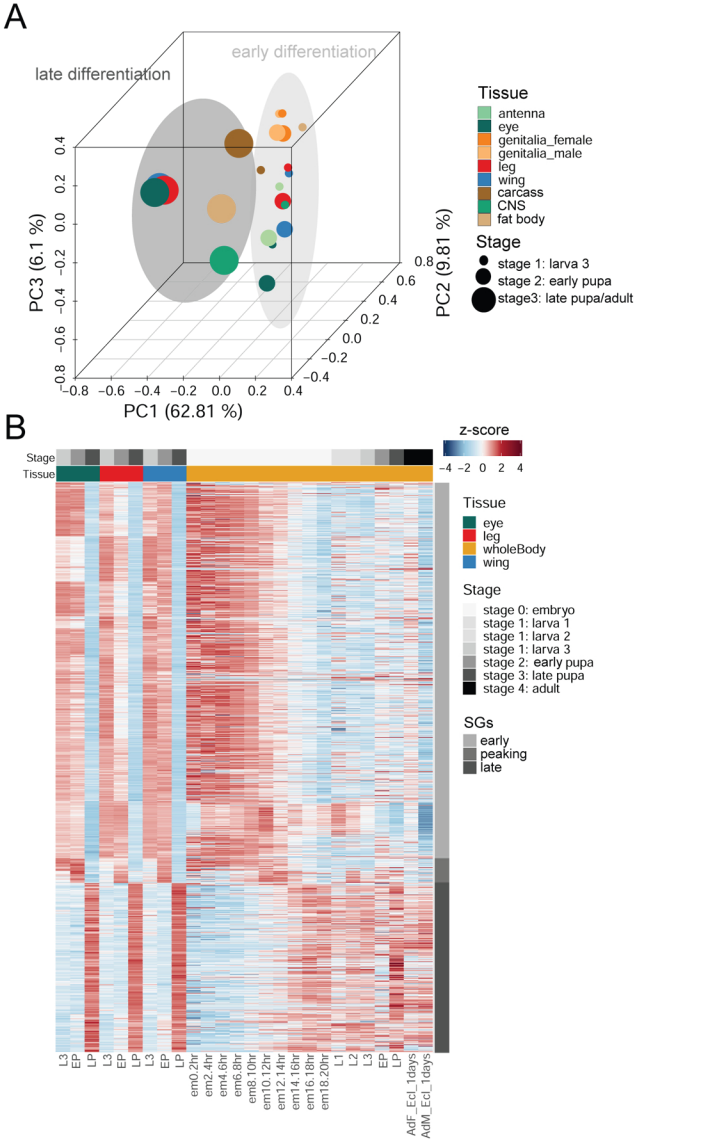
(A) General view of tissue differentiation (e.g. eye differentiation, green). Differentiation requires a transcriptional switch from early genes (light grey) towards late genes (dark grey) that could be mediated by stress, hormonal signals or other uncharacterized systemic and external signals (??). Gene expression, translation and cell cycle genes decrease while cuticle, organism differentiation, oxidative metabolic genes and functionally uncharacterized genes (??) are up-regulated upon differentiation. Eye genes, mainly involved in cell fate regulation, are expressed from precursor cells to fully differentiated ones along the process. Eye specific TFs regulate eye genes as well as TSGs. Early eye genes (green) are associated with axon guidance function and late eye genes are related to synaptic transmission and nervous system development functions. Negative regulation occurs between early and late genes and among tissue-specific TFs.

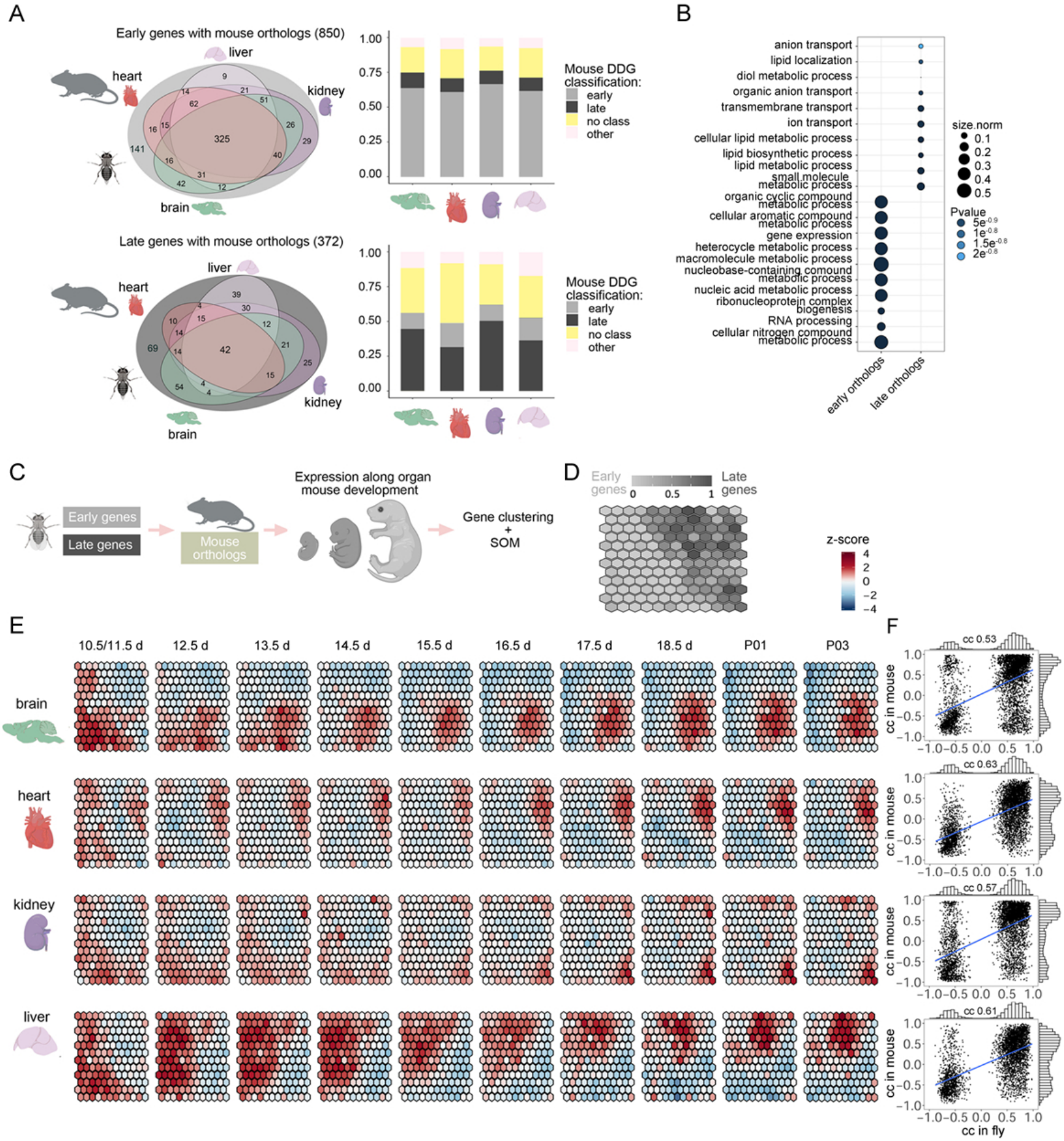

A
Developmental stages
Tissues
Larva 3
Early pupa
Late pupa
Wing compartments
Tissues
antenna
eye
genitalia_female
genitalia_male
leg
wing
Compartments
anterior
dorsal
posterior
ventral
Developmental stages
stage 1: larva 3
stage 2: early pupa
stage 3: late pupa
B
manual dissection
tissue disaggregation: trypsin treatment
cell sorting
RNA extraction, library preparation & NGS
C
Tissue: eye
Tissue: leg
Tissue: wing
Stage 1: larva 3
Stage 2: early pupa
Stage 3: late pupa
z-scores
tissue
eye
leg
wing
stage 1: larva 3
stage 2: white pupa
stage 3: late pupa
D
late stages
early stages
PC1 (42.1 %)
PC2 (18.17 %)
PC3 (8.89 %)

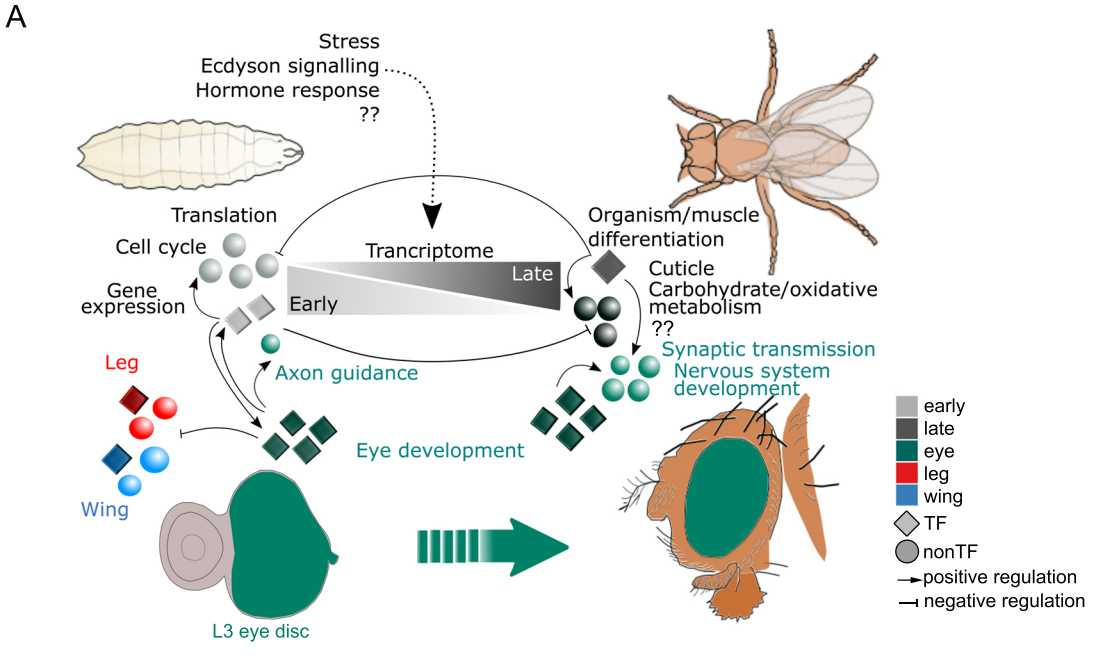

A
Proportion of variation across stages
Proportion of variation across tissues
0.00
0.25
0.50
0.75
1.00
count
20
15
10
5
DDG group
SGs
TGs
TSGs
B
Average contribution
percent.class
Residuals
Stage
Tissue
Tissue:Stage
C
z-score
−2 −1 0 1 2
Tissue
eye
leg
wing
Stage
stage 1: larva 3
stage 2: early pupa
stage 3: late pupa
DDG group
SGs
TGs
TSGs
DDGs
early
peaking
late
eye
leg
wing
eye_wing
leg_wing
TSGs
D
L3
EP
LP
L3 EP LP
Eye
Leg
Wing
Eye/Leg
Eye/Wing
Leg/Wing
non-differentially expressed
Developmental stages


A
SGs
H3K4me3 mark in TSS
Profile class
DDG group
TGs
TSGs
DDGs
early
peaking
late
eye
leg
wing
eye-wing
leg-wing
TSGs
Profile class
restricted
widespread
H3K4me3 mark in TSS
non specific
no mark
specific
B
Percentage of TFs
20
15
10
5
0
SGs
TGs
TSGs
DDGs group
SGs
TGs
TSGs
C
SGs
TGs
TSGs
Percentage of TFs
60
40
20
0
early
peaking
late
eye
leg
wing
eye−wing
leg−wing
EL3−EWP
ELP
LLP
WLP
ELP−LLP
ELP−WLP
LLP−WLP
DGGs group
SGs
TGs
TSGs


A
leg_wing
early
late
TSGs
eye
pearson correlation
-1.0 0.0 1.0
TF class
TF
nonTF
DDGs
no_class
early
peaking
late
eye
leg
wing
eye_wing
leg_wing
TSGs
B
Class
Cluster
DDGs Categories
TF proportion
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
no class
early
peaking
late
eye
leg
wing
eye_wing
leg_wing
TSGs
Proportion
0.0
25
0.5
0.75
1.0
C
pearson correlation
-1.0 0.0 1.0
TF class
TF
nonTF
Cluster
1- regulation of gene expression
2- cell cycle
3- translation
4- intracellular transport
5- respiratory system/ fluid transport
6- cuticle development/ muscle differentiation
7- carbohydrate metabolism / cuticle development
8- regulation transcription/ organism development
9- Toll signalling /not characterized
10- ion transport /nervous system
11-synapse organization
12- eye development
13- axon guidance/ synaptic transmission
14- leg development
15- limb/appendage development
D
cluster 8
cluster 9
cluster 10
cluster 11
cluster 7
cluster 12
cluster 6
cluster 13
cluster 5
cluster 4
cluster 14
cluster 15
cluster 3
cluster 2
cluster 1


A
late differentiation
early differentiation
PC3 (6.1 %)
PC1 (62.81 %)
PC2 (9.81 %)
Tissue
antenna
eye
genitalia_female
genitalia_male
leg
wing
carcass
CNS
fat body
Stage
stage 1: larva 3
stage 2: early pupa
stage3: late pupa/adult
B
z-score
Stage
Tissue
Tissue
eye
leg
wholeBody
wing
Stage
stage 0: embryo
stage 1: larva 1
stage 1: larva 2
stage 1: larva 3
stage 2: early pupa
stage 3: late pupa
stage 4: adult
SGs
early
peaking
late
L3
EP
LP
em0.2hr
em2.4hr
em4.6hr
em6.8hr
em8.10hr
em10.12hr
em12.14hr
em14.16hr
em16.18hr
em18.20hr
L1
L2
AdF_Ecl_1days
AdM_Ecl_1days


A
Early genes with mouse orthologs (850)
liver
heart
kidney
brain
9
14
21
62
51
16
15
26
325
29
141
16
40
31
42
12
Mouse DDG classification:
early
late
no class
other
1.00
0.75
0.50
0.25
0.00
Late genes with mouse orthologs (372)
liver
heart
kidney
brain
39
4
30
10
15
12
14
69
14
42
21
25
15
54
4
4
Mouse DDG classification:
early
late
no class
other
B
anion transport
lipid localization
diol metabolic process
organic anion transport
transmembrane transport
ion transport
cellular lipid metabolic process
lipid biosynthetic process
lipid metabolic process
small molecule metabolic process
organic cyclic compound metabolic process
cellular aromatic compound metabolic process
gene expression
heterocycle metabolic process
macromolecule metabolic process
nucleobase-containing comound metabolic process
nucleic acid metabolic process
ribonucleoprotein complex biogenesis
RNA processing
cellular nitrogen compound metabolic process
early orthologs
late orthologs
size.norm
0.1
0.2
0.3
0.4
0.5
Pvalue
5e-0.9
1e-0.8
1.5e-0.8
2e-0.8
C
Early genes
Late genes
Mouse orthologs
Expression along organ mouse development
Gene clustering + SOM
D
Early genes
Late genes
0
0.5
1
z-score
4
2
0
−2
−4
E
10.5/11.5 d
12.5 d
13.5 d
14.5 d
15.5 d
16.5 d
17.5 d
18.5 d
P01
P03
brain
heart
kidney
liver
F
cc 0.53
cc 0.63
cc 0.57
cc 0.61
cc in mouse
cc in fly
1.0
0.5
0.0
−0.5
−1.0


A
Stress
Ecdyson signalling
Hormone response
??
Translation
Cell cycle
Gene expression
Trancriptome
Late
Early
Organism/muscle differentiation
Cuticle
Carbohydrate/oxidative metabolism
??
Synaptic transmission
Nervous system development
Axon guidance
Leg
Wing
Eye development
L3 eye disc
early
late
eye
leg
wing
TF
nonTF
positive regulation
negative regulation