Dynamics of Gene Co-expression Networks in Time-Series Data: A Case Study in *Drosophila melanogaster* Embryogenesis

Li Yieng Lau[1], Antonio Reverter[2], Nicholas J. Hudson[3], Marina Naval-Sanchez[4], Marina R. S. Fortes[1] and Pâmela A. Alexandre[2]*

[1] *School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, QLD, Australia,* [2] *Commonwealth Scientific and Industrial Research Organisation (CSIRO) Agriculture and Food, St Lucia, QLD, Australia,* [3] *School of Agriculture and Food Sciences, The University of Queensland, St Lucia, QLD, Australia,* [4] *Institute for Molecular Bioscience, The University of Queensland, St Lucia, QLD, Australia*

Co-expression networks tightly coordinate the spatiotemporal patterns of gene expression unfolding during development. Due to the dynamic nature of developmental processes simply overlaying gene expression patterns onto static representations of co-expression networks may be misleading. Here, we aim to formally quantitate topological changes of co-expression networks during embryonic development using a publicly available *Drosophila melanogaster* transcriptome data set comprising 14 time points. We deployed a network approach which inferred 10 discrete co-expression networks by smoothly sliding along from early to late development using 5 consecutive time points per window. Such an approach allows changing network structure, including the presence of hubs, modules and other topological parameters to be quantitated. To explore the dynamic aspects of gene expression captured by our approach, we focused on regulator genes with apparent influence over particular aspects of development. Those key regulators were selected using a differential network algorithm to contrast the first 7 (early) with the last 7 (late) developmental time points. This assigns high scores to genes whose connectivity to abundant differentially expressed target genes has changed dramatically between states. We have produced a list of key regulators – some increasing (e.g., *Tusp, slbo, Sidpn, DCAF12,* and *chinmo*) and some decreasing (*Rfx, bap, Hmx, Awh,* and *mld*) connectivity during development – which reflects their role in different stages of embryogenesis. The networks we have constructed can be explored and interpreted within Cytoscape software and provide a new systems biology approach for the Drosophila research community to better visualize and interpret developmental regulation of gene expression.

Keywords: PCIT, regulator genes, RNA-Seq – RNA sequencing, developmental process, transcriptomics

INTRODUCTION

The increasing accessibility of -omics data drives the development of computational methods to integrate different sources of information and connect the underlying molecular mechanisms to complex phenotypes (Greenham and McClung, 2018). Among those methods, co-expression networks gain increasing application for the ability to integrate large transcriptional datasets.

Co-expression network allows the simultaneous identification and clustering of genes with similar expression patterns across multiple/contrasting conditions, providing a global overview of the co-expression relationships between genes that are functionally related or members of the same pathway (Marti-Marimon et al., 2018). While including samples from multiple conditions can straighten relevant correlations, such an approach is prone to Simpson's paradox, where the correlation trend observed in individual networks is reversed when all the samples from multiple conditions are combined in one network. A detailed explanation of the Simpson's paradox can be found in this article (Wang et al., 2018).

Additionally, constructing co-expression networks considering all samples under scrutiny has a limited ability in identifying condition-specific modules because such a correlation signal can be diluted by a possible lack of correlation in other conditions (van Dam et al., 2017). But limiting the sample size to a specific tissue/condition would reduce the sample size, thereby compromising the statistical power in detecting shared co-expression modules (van Dam et al., 2017). Even with differential co-expression analysis, an approach that identifies genes with varying co-expression relationships under different conditions, it remains challenging to infer causality due to the static representation and the fact that correlation does not imply causation (Stuart et al., 2003).

Another challenge with co-expression network is the analysis of time series data. Studying time-series gene expression enables the identification of transient transcriptional changes, temporal patterns of a response and causal relationships between genes (Bar-Joseph et al., 2012). But incorporating time dependent changes to determine causal relationships within a gene network, such as a transcription factor (TF) and its target genes, remain challenging, mostly because different processes gain and lose importance over time in a non-linear fashion (Greenham and McClung, 2018). Investigation of a single network built from all time-point samples leads to an even poorer capture of meaningful correlations than discussed before. This is particularly true for dynamic processes, such as *Drosophila melanogaster* embryogenesis comprising multiple developmental stages. For example, in the early stages of embryogenesis, the zinc finger class of TFs are predominantly expressed and a large number of TFs are maternally contributed (Adryan and Teichmann, 2010). In later stage, however, the homeobox TFs were predominantly expressed and maternally contributed TFs were significantly reduced (Adryan and Teichmann, 2010). Therefore, one can expect different biological processes (and thus different connections among genes) appearing and disappearing over time, while connections for fundamental genes remain for basic cell functions.

In this study, we employed a likelihood-based approach to exploit the dynamics of gene networks over time using a very comprehensive time-series dataset from a recently published study in *Drosophila melanogaster* (Becker et al., 2018). Becker et al. (2018) provided a highly detailed description of embryogenesis assayed across 14 time points and interpreted their data through the conventional approach of contrasting patterns of differential expression between the various time points, with a particular view to connect mRNA and protein abundance data. Nevertheless, this comprehensive dataset of a premier model to study developmental biology provides a great opportunity to better understand the changing topology of the Drosophila networks. The approach described here provides a more dynamic visualization of gene expression over time and captures the relevance of specific genes according to the developmental stage. It can be applied to other time-series - omics data, which we argue is more insightful than overlaying patterns of differential expression onto static representations of co-expression networks.

METHODS

The Drosophila embryogenesis dataset used in this study was obtained from the NCBI's Gene Expression Omnibus, GEO Series accession number GSE121160. Further information about sample collection and RNA libraries generation can be found in the original publication (Becker et al., 2018). The mentioned study generated a paired transcriptome/proteome time course dataset with 14 time points during *Drosophila melanogaster* embryogenesis, i.e., 0, 1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 18, and 20 h. The 68 embryo RNA libraries were pooled in equimolar ratio and sequenced on 8 lanes of a HiSeq2500 (1×51 cycles plus 7 cycles for the index read).

Reads were mapped to the BDGP6 fly reference genome and gene expression was estimated as read counts. We considered for the analysis 7,640 genes that presented counts in at least 20 samples and more than 100 counts on average. Data were averaged within each time point and log2 transformed. From the genes that passed quality control, we focused on 3,568 genes clustered according to (Becker et al., 2018), based on pairwise comparison of genes up or down-regulated (relative to the first time point – 0 h) in mRNA and protein data, resulting in groups named to represent the gene status on mRNA/protein: up/up (495 genes), down/up (1,736 genes), down/down (1,022 genes), and up/down (315 genes). Genes were also classified as regulators (791 genes) based on the list provided by Rhee et al. (2014) consisting of essential genes involved in replication and transcription, splicing, DNA repair and cell division.

To identify key regulators of Drosophila embryogenesis among the 791 pre-defined regulators mentioned before, we applied the regulatory impact factor metrics (RIF; Reverter et al., 2010) considering all clusters and regulators as targets (4,133 genes). RIF has been applied in a range of biological circumstances reviewed in Hudson et al. (2012) and Ehsani and Drabløs (2020), including a very recent example relating to how sunitinib drug treatment influences kidney cancer (Al-Lamki et al., 2020). In brief, RIF combines the correlation between a regulator and its potential targets with the degree of differential expression of the targets between the tested conditions. Therefore, RIF requires contrasting conditions, which we defined as early (the first 7 time points) vs. late (the last 7 time points) embryogenesis. The results are comprised by two metrics designed to assign scores to (1) regulator genes consistently differentially co-expressed with target genes

(RIF1), and to (2) those with the most altered ability to predict the abundance of target genes (RIF2). Regulators with scores deviating ± 1.96 SD from the mean (corresponding to a nominal t-test $P < 0.05$) were considered significant and labeled as "key."

The dynamic aspects of gene expression during embryogenesis were explored by creating 10 groups of 5 consecutive time points (i.e., 0–4 h, 1–5 h and so on) which were then used to create 10 networks by applying the Partial Correlation and Information Theory (PCIT) algorithm (Reverter and Chan, 2008) to the 4,133 genes (clustered genes and regulators). PCIT combines partial correlation coefficients with information theory to determine locally significant correlations automatically, avoiding the need for the specification of fixed correlation cut-offs. In short, PCIT is a data driven approach that explored all the correlations between possible triplets of genes prior to the identification of significant correlations that are within extremes of the distribution (Reverter and Chan, 2008). The outputs of PCIT were visualized on Cytoscape Version 3.7.1 (Shannon et al., 2003). One additional network was created considering all the time points to allow us to compare the two approaches.

To further explore the networks, we eliminated connections that were present in six or more networks, as those were considered to be fairly conserved and would not reflect the dynamic aspects of gene expression during embryogenesis. On the other side, we kept only connections that appear in the same direction (positive or negative) in at least 3 consecutive networks, to capture more meaningful correlation and avoid technical noise. Finally, we focused on correlations involving key regulators and explored the changes between networks over time by creating an animated image in Graphics Interchange Format (GIF). Functional enrichment analysis was performed on STRINGv10 online platform (Szklarczyk et al., 2015) using hypergeometric test and correction for multiple tests ($FDR < 0.05$).

RESULTS AND DISCUSSION

By defining 5 consecutive time points as the number of samples used to construct co-expression networks and sliding forward one time point at a time, we generated 10 networks and compared them to recover some of the dynamic aspects of gene expression during Drosophila embryogenesis. The number of consecutive time points and total networks were arbitrary and designed to generate the higher number of networks while keeping a reasonable number of samples to extract meaningful correlations. Our goal was to demonstrate that changes in gene behavior over time can be captured and add value to the interpretation of the underlying biological processes. However, those parameters can be adjusted according to the biological question and number of samples under investigation.

Comparing the characteristics of each network (**Table 1**), although they were created based on the same set of 4,133 genes, they presented different numbers of significant connections (edges) which reflects the appearance and disappearance of significant correlations over time. Most of the other topological parameters remain unchanged or demonstrate very slight change.

The variation in the number of connections per gene in each network was captured in **Figure 1**. The co-expression

networks demonstrated a scale-free Power-law connectivity distribution; a few nodes are highly connected (hubs) and many nodes are lowly connected (Barabási and Oltvai, 2004). Although the distribution of degree (number of significant connections per gene) change according to the network, a higher proportion of genes possess lower degree (between 50 and 150 connections), while some networks present a few genes with possessing as many as 470 significant connections.

Comparing the connections between genes across the networks, only 74 were present in all the networks as illustrated in **Supplementary Figure S1A**. Those conserved connections involved a set of 131 genes which presented enrichment for KEGG pathways ribosome ($FDR = 1.63e-21$), spliceosome ($FDR = 3.92e-08$), and DNA replication ($FDR = 2.45e-05$, **Supplementary Figure S1B**), all terms related to basic cell functions for maintenance of life. Although in this study our aim was to explore the dynamic aspects of genes expression and therefore focus on connections changing over time, exploring conserved connections across several consecutive networks is also a valid strategy to understand the biological mechanisms behind the process under investigation. In this study, however, most of the connections were present in only one of the networks (1,520,837 out of 2,400,823 total significant connections; **Supplementary Figure S1A**) and the highest level of overlap occurred at around 4 consecutive networks (**Supplementary Figure S1C**), reflecting the changes in correlations as samples are gradually replaced.

The RIF analysis performed by contrasting the first 7 developmental time points against the last 7 resulted in 59 key regulator genes ($P < 0.05$; **Supplementary Figure S2**). The need to define contrasting conditions and the high overlap between consecutive networks led us to perform RIF considering all samples as opposed to comparing consecutive networks. Although all 59 key regulators were only significant according to RIF2, if we look at the top 10 regulators according to RIF1 (**Supplementary Table S1**) we will find *HmgD*, *Ubx*, and *Sp1*, all genes previously reported for their roles in Drosophila embryogenesis. In details, *HmgD* is a high mobility group protein that is highly concentrated at the beginning of Drosophila embryogenic development but the exact role remain unclear (Renner et al., 2000). On the other hand, *Ubx* plays an important developmental role throughout Drosophila embryogenesis by affecting abdominal identities during early stage of embryogenesis and thoracic segments identities in later stage of embryogenesis (Lamka et al., 1992). Epigenetic enhancer silencing was found to regulate *Ubx* expression at different embryonic stage by responding to *Ubx* levels and genetic variation (Crickmore et al., 2009). Similar to *Ubx*, *Sp1* also plays an important developmental role during Drosophila embryogenesis where the effect is confined to ventral appendage specification that affects the leg development at larval stage (Estella and Mann, 2010). These results reflect the aptitude of RIF analysis to put forward genes with regulatory potential.

RIF and PCIT are related in the sense they are both assessing patterns of connectivity via co-expression. By sorting the 59 key regulators according to their connectivity on network 1

TABLE 1 | Topological parameters of co-expression networks.

	Net1	Net2	Net3	Net4	Net5	Net6	Net7	Net8	Net9	Net10
Nodes	4,133	4,133	4,133	4,133	4,133	4,133	4,133	4,133	4,133	4,133
Edges	414,539	354,486	450,098	380,369	332,086	352,702	337,808	421,647	370,491	331,339
CC	0.50	0.49	0.47	0.45	0.45	0.44	0.45	0.45	0.44	0.44
NCC	1	1	1	1	1	1	1	1	1	1
Diameter	7	6	6	6	6	6	6	6	6	6
Centrz	4	5	4	4	4	4	4	4	4	4
avSpl	0.05	0.04	0.06	0.06	0.04	0.05	0.05	0.06	0.05	0.04
avNeighb	3.52	3.51	3.32	3.21	3.32	3.29	3.29	3.10	3.27	3.30
Density	201	172	218	184	161	171	163	204	179	160
Heterog	0.05	0.04	0.05	0.04	0.04	0.04	0.04	0.05	0.04	0.04

CC, clustering coefficient[a]; NCC, connected components[b]; Centrz, centralization[c]; avSpl, average shortest path length[d]; avNeighb, average number of neighbors[e]; Heterog, heterogeneity[f]. [a]Clustering coefficient – a ratio N/M where N is the number of edges between the neighbors of a node and M is the maximum number of edges that could possibly exist between the neighbors of a node. [b]Connected components – an indication of network connectivity where a lower number of connected components suggests a stronger connectivity. [c]Centralization – the degree to which a few members hold the greatest number of connections in the network. High centralization score represents a highly centralized network where only a few members hold the most central positions. [d]Average shortest path length – expected distance between two connected nodes. [e]Average number of neighbors – average connectivity of a node in the network. [f]Heterogeneity – tendency of a network to contain hub nodes.

(representing the earliest embryogenesis stage) it is possible to notice two blocks of key regulators: the ones increasing the number of connections over time and the ones decreasing the number of connections over time (**Figure 2**). Considering all the genes used for network construction, 12 out of the 100 least connected genes are key regulators that increase connectivity over time, representing a significant enrichment (P = 19.02e-09). On the other side, key regulators that decrease connectivity over time are more spread among all the genes in network1.

Some of the key regulators were previously reported for their roles in Drosophila embryogenesis. For example, *Tusp* (tubby domain superfamily proteins) expression was detected in sensory neurons and brain cells during the later stages of Drosophila embryogenesis and this is consistent with our findings, showing an increased connectivity of *Tusp* during late embryogenesis (Ronshaugen et al., 2002). Other key regulators with increased connectivity over time include *slbo* (slow border cells) and *DCAF12* (DDB1 and CUL4 associated factor 12). The *slbo* locus is vital for regulated cell migration during Drosophila development and a null mutation can lead to lethality in late embryonic or early larval life (Montell et al., 1992). This is consistent with the increased connectivity of this gene over time, indicating an increased activity toward late developmental stages. Concordantly, *DCAF12,* a regulator of apoptosis in Drosophila, is also reported in the literature as having a significant increase of connectivity during late embryogenesis (Hwangbo et al., 2016). Thus, it is reasonable to hypothesize the increased connectivity of *slbo* and *DCAF12* are important for the survival of the embryo.

Some of the key regulators with decreased connectivity over time were also previously studied in the context of Drosophila embryogenesis. One example is *bcd* (bicoid), a gene that has been well studied for its role in the anterior-posterior specification during Drosophila embryogenesis (Frohnhöfer and Nüsslein-Volhard, 1986). According to our results, *bcd* demonstrated the highest connectivity during early embryogenesis which is consistent with a previous study where mRNA of *bcd* were highly

expressed in early embryo for anterior specification (Berleth et al., 1988). Likewise, *Hmx* (homeobox) gene has been shown to be expressed in developing Drosophila brain during early embryonic stages and it was suggested to be paramount for the development of the Drosophila central nervous system (Wang et al., 2000). The Drosophila *Rfx* (regulatory factor X) gene was previously identified as being a peripheral neuron marker and can also be found in the brain, although its presence is not restricted to embryogenesis but throughout development (Vandaele et al., 2001). The distinct change in connectivity of *Rfx* between early and late embryogenesis in our analysis implies its possible role during embryogenesis that is yet to be elucidated. We also found the *mld* (molting defective) gene to decrease connectivity over time. The *mld* gene is required for the production of ecdysone, a hormone that controls molting during Drosophila larval development (Ono et al., 2006). Nevertheless, there is lack of evidence of *mld* role during embryogenesis.

By focusing only on connections involving key regulators and present in less than 6 networks, we can observe the changes in the co-expression network over time regarding key regulators gaining and losing connections, as well as the change in the direction of the connections. When a gene is highly connected (present significant correlation with many other genes in the network), it is considered a central regulator of gene expression, since a slight change in their expression will simultaneously influence several other genes. This tightly connected cluster of genes is expected to work coordinately to a specific biological function or pathway relevant to the trait under investigation (Stuart et al., 2003). When in a different condition or, in this case, a different set of time points, the same gene does not demonstrate high connectivity, the lack of significant correlations indicates that this gene, and therefore the biological process it represents, is not so relevant to the phenotype anymore. In our study, by sliding one time point at a time and creating consecutive networks, we were able to visualize the gradual increase and decrease of connectivity of key regulators over time, which is related to the

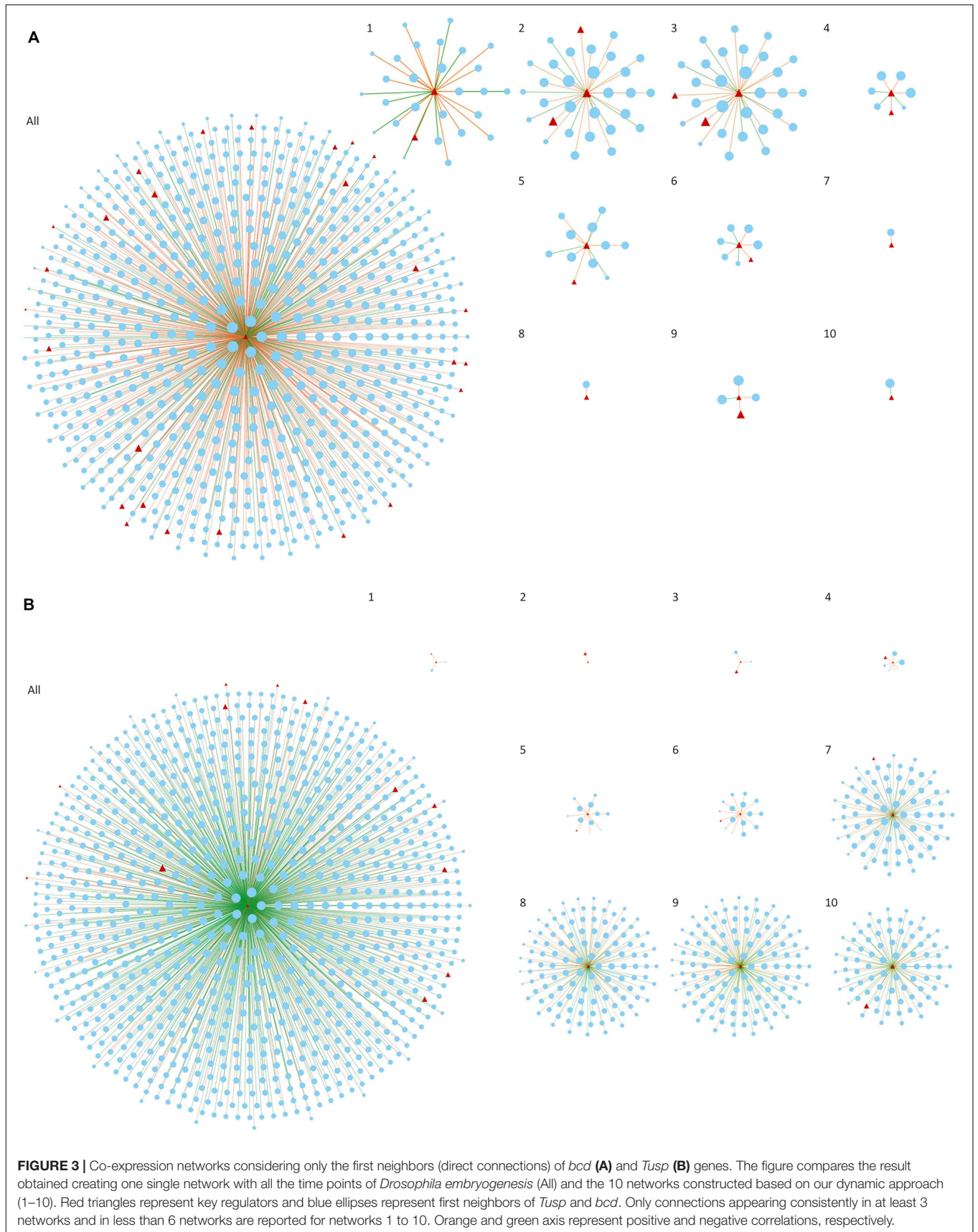| Degree range	Number of genes									
	Network1	Network2	Network3	Network4	Network5	Network6	Network7	Network8	Network9	Network10
0	4	3	2	0	0	0	0	0	0	0
10	21	22	30	1	8	10	10	9	8	3
20	54	33	50	39	21	26	15	22	47	55
30	45	35	63	100	58	84	24	60	95	86
40	53	47	46	128	98	112	62	121	97	106
50	46	61	65	139	156	109	165	170	108	117
60	60	67	72	177	174	143	247	206	122	141
70	65	95	67	217	215	145	292	199	130	151
80	75	122	83	267	269	201	227	149	160	197
90	83	169	109	185	253	203	168	145	192	171
100	99	225	128	153	247	212	155	112	194	244
110	115	294	173	129	215	190	181	120	167	199
120	148	325	143	114	176	191	204	120	147	173
130	155	274	148	125	152	160	172	129	133	182
140	176	209	125	102	130	118	149	104	158	167
150	194	186	112	120	138	130	150	107	149	185
160	201	151	106	110	118	122	155	95	133	140
170	191	132	109	114	138	125	141	114	143	147
180	167	104	122	102	98	113	145	118	144	147
190	169	116	123	111	103	141	104	104	118	168
200	159	104	120	109	107	135	106	98	136	146
210	141	114	116	107	106	128	129	94	133	152
220	131	113	132	117	108	147	115	83	131	124
230	122	148	120	99	109	139	101	100	134	160
240	154	147	116	76	100	153	99	91	120	124
250	161	137	152	66	101	141	113	78	112	115
260	145	143	127	63	103	112	109	75	109	137
270	145	91	130	86	111	113	87	106	98	133
280	131	114	93	87	116	92	102	80	91	95
290	116	105	90	95	125	109	90	80	92	61
300	103	103	128	70	103	81	80	81	99	46
310	97	66	91	80	97	92	67	74	91	42
320	118	51	103	76	53	76	56	80	101	15
330	89	13	121	67	24	44	48	68	74	3
340	75	12	102	83	3	26	33	83	56	1
350	68	2	84	101	0	9	17	77	55	0
360	41	0	69	97	0	1	9	76	41	0
370	14	0	82	76	0	0	4	73	11	0
380	0	0	58	74	0	0	2	81	3	0
390	2	0	53	45	0	0	0	74	1	0
400	0	0	41	20	0	0	0	80	0	0
410	0	0	30	6	0	0	0	68	0	0
420	0	0	33	0	0	0	0	46	0	0
430	0	0	15	0	0	0	0	39	0	0
440	0	0	26	0	0	0	0	31	0	0
450	0	0	14	0	0	0	0	9	0	0
460	0	0	8	0	0	0	0	4	0	0
470	0	0	3	0	0	0	0	0	0	0

FIGURE 1 | Number of genes within different degree ranges in each of the 10 networks. Degree refers to the number of significant connections per gene. All networks were comprised of 4,133 genes/nodes.

increase or decrease of their role in each particular moment. To illustrate that, **Figure 3** shows two of our key regulators, *bcd* (**Figure 3A**) and *Tusp* (**Figure 3B**), and their first neighbors (direct connections). For a dynamic representation of the changes in the networks over time, please refer to **Supplementary File S2** (*bcd*) and **Supplementary Files S3** (*Tusp*).

As discussed earlier, the roles of *bcd* and *Tusp* in the beginning and in the end of embryogenesis, respectively, are well known and revealed here by the changes in number of connections over time. In contrast, combining connections from all the 10 networks in a single network hinders the observation of such changes which consequently complicates the extraction of

Key regulators	Connectivity									
	Network1	Network2	Network3	Network4	Network5	Network6	Network7	Network8	Network9	Network10
CG6118	0.04	0.03	0.06	0.07	0.10	0.08	0.23	0.48	0.38	0.28
CG7099	0.04	0.33	0.06	0.13	0.22	0.60	0.65	0.72	0.65	0.60
Mlp84B	0.04	0.10	0.07	0.08	0.09	0.04	0.35	0.48	0.42	0.82
CG12075	0.05	0.10	0.08	0.06	0.19	0.34	0.60	0.89	0.86	0.56
slbo	0.05	0.03	0.11	0.16	0.13	0.06	0.22	0.50	0.43	0.69
Tusp	0.05	0.03	0.05	0.06	0.15	0.27	0.56	0.78	0.83	0.81
Sidpn	0.06	0.07	0.03	0.07	0.14	0.21	0.45	0.86	0.81	0.49
DCAF12	0.06	0.17	0.07	0.14	0.78	0.80	0.53	0.47	0.59	0.61
CG3328	0.06	0.07	0.04	0.09	0.11	0.07	0.24	0.66	0.58	0.62
CG43347	0.07	0.28	0.04	0.06	0.10	0.22	0.48	0.95	0.71	0.59
CG15365	0.07	0.14	0.24	0.13	0.13	0.21	0.42	0.63	0.83	0.65
chinmo	0.08	0.16	0.16	0.15	0.28	0.46	0.60	0.78	0.79	0.69
Oli	0.16	0.64	0.49	0.76	0.83	0.73	0.37	0.21	0.18	0.35
bi	0.17	0.40	0.54	0.89	0.96	0.75	0.45	0.21	0.11	0.39
CG18265	0.18	0.37	0.42	0.87	0.90	0.70	0.30	0.14	0.08	0.35
HLH54F	0.24	0.79	0.63	0.82	0.91	0.68	0.35	0.14	0.18	0.25
svp	0.31	0.41	0.58	0.84	0.95	0.70	0.30	0.07	0.09	0.13
Myb	0.32	0.06	0.15	0.47	0.39	0.30	0.61	0.97	0.84	0.83
croc	0.37	0.62	0.39	0.33	0.55	0.57	0.19	0.11	0.38	0.26
h	0.39	0.50	0.24	0.54	0.33	0.19	0.22	0.12	0.31	0.12
cact	0.40	0.46	0.60	0.66	0.68	0.85	0.33	0.15	0.15	0.24
REPTOR-BP	0.40	0.54	0.65	0.87	0.63	0.27	0.16	0.15	0.13	0.32
Hey	0.40	0.70	0.44	0.53	0.72	0.67	0.37	0.21	0.10	0.52
POLDIP2	0.43	0.51	0.29	0.08	0.28	0.42	0.19	0.20	0.24	0.23
luna	0.45	0.54	0.54	0.88	0.97	0.80	0.32	0.11	0.20	0.61
so	0.48	0.78	0.77	0.43	0.38	0.32	0.15	0.10	0.11	0.12
EcR	0.54	0.57	0.39	0.22	0.90	0.65	0.22	0.13	0.36	0.24
exex	0.55	0.88	0.73	0.22	0.33	0.51	0.36	0.45	0.08	0.38
CG6689	0.56	0.66	0.84	0.20	0.15	0.23	0.16	0.13	0.25	0.13
grn	0.56	0.84	0.88	0.43	0.92	0.71	0.37	0.09	0.12	0.36
esg	0.57	0.44	0.53	0.25	0.21	0.21	0.14	0.27	0.30	0.29
net	0.59	0.80	0.43	0.42	0.34	0.26	0.13	0.26	0.09	0.45
bcd	0.60	0.74	0.42	0.12	0.19	0.17	0.09	0.12	0.25	0.29
CG11414	0.60	0.42	0.15	0.36	0.47	0.63	0.31	0.60	0.19	0.19
CG8765	0.60	0.68	0.56	0.22	0.21	0.16	0.17	0.10	0.17	0.24
C15	0.62	0.77	0.72	0.34	0.22	0.30	0.19	0.22	0.40	0.25
toy	0.62	0.82	0.78	0.33	0.17	0.27	0.18	0.06	0.08	0.29
CG32982	0.62	0.82	0.62	0.29	0.49	0.52	0.20	0.12	0.17	0.10
drm	0.63	0.84	0.69	0.53	0.32	0.29	0.13	0.12	0.42	0.23
usp	0.65	0.54	0.19	0.60	0.80	0.47	0.20	0.23	0.31	0.32
rdx	0.65	0.45	0.08	0.32	0.25	0.64	0.72	0.19	0.07	0.33
CG15011	0.69	0.57	0.20	0.78	0.94	0.67	0.37	0.19	0.15	0.27
CG2678	0.69	0.71	0.73	0.36	0.42	0.44	0.28	0.10	0.10	0.45
Atf6	0.69	0.66	0.53	0.40	0.86	0.63	0.26	0.09	0.10	0.06
brwl	0.71	0.49	0.47	0.12	0.42	0.89	0.24	0.12	0.36	0.71
CG34376	0.72	0.46	0.57	0.93	0.92	0.56	0.35	0.45	0.47	0.30
Mnt	0.73	0.90	0.56	0.33	0.85	0.73	0.39	0.09	0.11	0.17
CG4404	0.73	0.85	0.53	0.23	0.28	0.13	0.04	0.06	0.08	0.10
Rfx	0.73	0.87	0.49	0.47	0.73	0.62	0.33	0.17	0.15	0.26
CG10209	0.75	0.56	0.70	0.53	0.28	0.14	0.33	0.47	0.23	0.26
CG10462	0.77	0.71	0.32	0.72	0.67	0.61	0.41	0.16	0.14	0.68
fs(1)Ya	0.78	0.84	0.53	0.42	0.46	0.32	0.18	0.12	0.08	0.11
CG10979	0.79	0.64	0.70	0.18	0.45	0.60	0.34	0.12	0.33	0.62
bap	0.80	0.71	0.57	0.41	0.24	0.23	0.19	0.12	0.53	0.24
Hmx	0.83	0.74	0.74	0.13	0.08	0.15	0.04	0.07	0.12	0.20
Awh	0.84	0.60	0.56	0.59	0.76	0.57	0.34	0.26	0.19	0.37
mld	0.84	0.91	0.59	0.36	0.82	0.66	0.30	0.11	0.08	0.15
CG8089	0.90	0.90	0.44	0.55	0.87	0.48	0.33	0.14	0.24	0.35
maf-S	0.94	0.79	0.71	0.21	0.40	0.68	0.20	0.17	0.25	0.15

FIGURE 2 | Connectivity of the 59 key regulators in each network. Connectivity was calculated as the number of significant connections of a gene divided by the number of significant connections of the most connected gene in the network.

FIGURE 3 | Co-expression networks considering only the first neighbors (direct connections) of *bcd* **(A)** and *Tusp* **(B)** genes. The figure compares the result obtained creating one single network with all the time points of *Drosophila embryogenesis* (All) and the 10 networks constructed based on our dynamic approach (1–10). Red triangles represent key regulators and blue ellipses represent first neighbors of *Tusp* and *bcd*. Only connections appearing consistently in at least 3 networks and in less than 6 networks are reported for networks 1 to 10. Orange and green axis represent positive and negative correlations, respectively.

information regarding regulatory role of the gene at specific time points. Although considering all time points leads to numerous significant connections that are statistically more robust and can be important to understand the overall function of a gene, the dynamic aspects expected to be represented in a time series data is actually lost. To compensate for the small number of time points used in each network, our approach focused on connections appearing consistently in at least three consecutive networks. It is important to note that tissues samples collected for any RNA-Seq experiment are prone to bias due to cellular heterogeneity (Kukurba and Montgomery, 2015). In our particular example, each time point represents the combined expression of multiple cell types with specific functions in the embryo. Avoiding such bias would involve approaches such as single-cell RNA sequencing, which would significantly increase the costs of the experiment (Nguyen et al., 2018).

With a likelihood-based approach, we were able to capture the dynamicity of gene networks across different time points in Drosophila embryogenesis, focusing on key regulatory genes. Our approach provides a novel and complementary strategy to understanding the topology of gene networks by sliding smoothly from early to late development. One can focus on specific dynamic aspects such as genes with increasing or decreasing connectivity over time, or even explore conserved mechanisms along the biological process under investigation. Although it is out of the scope of this work to discuss specific biological aspects of Drosophila embryogenesis, we were able to capture some known biological signals regarding early and late developmental stages. Our results recapitulate the known molecular biology of Drosophila embryogenesis and revealed

new insights for further studies. Being able to extract such comprehensive information justify the value of this approach. We anticipate the dynamic investigation proposed here being applied to other time-series-omics data, as a way to further explore regulatory aspects of gene expression changes over time. We argue this approach is preferable to overlaying patterns of differential expression onto static representations of co-expression network.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE121160.

AUTHOR CONTRIBUTIONS

AR conceived and designed the study. AR, LL, NH, and PA performed formal analysis, investigation and visualization of the presented data. LL and PA wrote the manuscript with contributions from all authors. MF and MN-S reviewed the manuscript and added substantial information and insights.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00517/full#supplementary-material

REFERENCES

Adryan, B., and Teichmann, S. A. (2010). The developmental expression dynamics of Drosophila melanogaster transcription factors. Genome Biol. 11, R40–R40. doi: 10.1186/gb-2010-11-4-r40

Al-Lamki, R. S., Hudson, N. J., Bradley, J. R., Warren, A. Y., Eisen, T., Welsh, S. J., et al. (2020). The efficacy of sunitinib treatment of renal cancer cells is associated with the protein PHAX in vitro. Biology 9, doi: 10.3390/biology9040074

Barabási, A.-L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. Nat. Rev. Genet. 5, 101–113. doi: 10.1038/nrg1272

Bar-Joseph, Z., Gitter, A., and Simon, I. (2012). Studying and modelling dynamic biological processes using time-series gene expression data. Nat. Rev. Genet. 13, 552–564. doi: 10.1038/nrg3244

Becker, K., Bluhm, A., Casas-Vila, N., Dinges, N., Dejung, M., Sayols, S., et al. (2018). Quantifying post-transcriptional regulation in the development of Drosophila melanogaster. Nat. Commun. 9:4970. doi: 10.1038/s41467-018-07455-9

Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., et al. (1988). The role of localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. EMBO J. 7, 1749–1756. doi: 10.1002/j.1460-2075.1988.tb03004.x

Crickmore, M. A., Ranade, V., and Mann, R. S. (2009). Regulation of Ubx expression by epigenetic enhancer silencing in response to Ubx levels and genetic variation. PLoS Genet. 5:e1000633. doi: 10.1371/journal.pgen.1000633

Ehsani, R., and Drabløs, F. (2020). Enhanced identification of significant regulators of gene expression. BMC Bioinformatics 21:134. doi: 10.1186/s12859-020-3468-z

Estella, C., and Mann, R. S. (2010). Non-redundant selector and growth-promoting functions of two sister genes, buttonhead and Sp1, in Drosophila leg development. PLoS Genet. 6:e1001001. doi: 10.1371/journal.pgen.1001001

Frohnhöfer, H. G., and Nüsslein-Volhard, C. (1986). Organization of anterior pattern in the Drosophila embryo by the maternal gene bicoid. Nature 324, 120–125. doi: 10.1038/324120a0

Greenham, K., and McClung, C. R. (2018). Time to build on good design: resolving the temporal dynamics of gene regulatory networks. Proc. Natl. Acad. Sci. 115:6325. doi: 10.1073/pnas.1807707115

Hudson, N. J., Dalrymple, B. P., and Reverter, A. (2012). Beyond differential expression: the quest for causal mutations and effector molecules. BMC Genomics 13:356. doi: 10.1186/1471-2164-13-356

Hwangbo, D.-S., Biteau, B., Rath, S., Kim, J., and Jasper, H. (2016). Control of apoptosis by Drosophila DCAF12. Dev. Biol. 413, 50–59. doi: 10.1016/j.ydbio.2016.03.003

Kukurba, K. R., and Montgomery, S. B. (2015). RNA Sequencing and Analysis. Cold Spring Harb. Protoc. 2015, 951–969. doi: 10.1101/pdb.top084970

Lamka, M. L., Boulet, A. M., and Sakonju, S. (1992). Ectopic expression of UBX and ABD-B proteins during Drosophila embryogenesis: competition, not a functional hierarchy, explains phenotypic suppression. Development 116, 841–854.

Marti-Marimon, M., Vialaneix, N., Voillet, V., Yerle-Bouissou, M., Lahbib-Mansais, Y., and Liaubet, L. (2018). A new approach of gene co-expression network inference reveals significant biological processes involved in porcine

muscle development in late gestation. *Sci. Rep.* 8:10150. doi: 10.1038/s41598-018-28173-8

Montell, D. J., Rorth, P., and Spradling, A. C. (1992). slow border cells, a locus required for a developmentally regulated cell migration during oogenesis, encodes Drosophila C/EBP. *Cell* 71, 51–62. doi: 10.1016/0092-8674(92)90265-e

Nguyen, Q. H., Pervolarakis, N., Nee, K., and Kessenbrock, K. (2018). Experimental considerations for single-cell RNA sequencing approaches. *Front. Cell Dev. Biol.* 6:108. doi: 10.3389/fcell.2018.00108

Ono, H., Rewitz, K. F., Shinoda, T., Itoyama, K., Petryk, A., Rybczynski, R., et al. (2006). Spook and Spookier code for stage-specific components of the ecdysone biosynthetic pathway in Diptera. *Dev. Biol.* 298, 555–570. doi: 10.1016/j.ydbio.2006.07.023

Renner, U., Ghidelli, S., Schafer, M. A., and Wisniewski, J. R. (2000). Alterations in titer and distribution of high mobility group proteins during embryonic development of *Drosophila melanogaster*. *Biochim. Biophys. Acta* 1475, 99–108. doi: 10.1016/s0304-4165(00)00054-4

Reverter, A., and Chan, E. K. (2008). Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* 24, 2491–2497. doi: 10.1093/bioinformatics/btn482

Reverter, A., Hudson, N. J., Nagaraj, S. H., Perez-Enciso, M., and Dalrymple, B. P. (2010). Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. *Bioinformatics* 26, 896–904. doi: 10.1093/bioinformatics/btq051

Rhee, D. Y., Cho, D.-Y., Zhai, B., Slattery, M., Ma, L., Mintseris, J., et al. (2014). Transcription factor networks in Drosophila melanogaster. *Cell Rep.* 8, 2031–2043. doi: 10.1016/j.celrep.2014.08.038

Ronshaugen, M., McGinnis, N., Inglis, D., Chou, D., Zhao, J., and McGinnis, W. (2002). Structure and expression patterns of Drosophila TULP and TUSP, members of the tubby-like gene family. *Mech. Dev.* 117, 209–215. doi: 10.1016/S0925-4773(02)00211-3

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Stuart, J. M., Segal, E., Koller, D., and Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302, 249–255. doi: 10.1126/science.1087447

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452. doi: 10.1093/nar/gku1003

van Dam, S., Võsa, U., van der Graaf, A., Franke, L., and de Magalhães, J. P. (2017). Gene co-expression analysis for functional classification and gene–disease predictions. *Briefings Bioinform.* 19, 575–592. doi: 10.1093/bib/bbw139

Vandaele, C., Coulon-Bublex, M., Couble, P., and Durand, B. (2001). Drosophila regulatory factor X is an embryonic type I sensory neuron marker also expressed in spermatids and in the brain of Drosophila. *Mech. Dev.* 103, 159–162. doi: 10.1016/s0925-4773(01)00340-9

Wang, B., Wu, P., Kwan, B., Tu, X. M., and Feng, C. (2018). Simpson's paradox: examples. *Shanghai Arch. Psychiatry* 30, 139–143. doi: 10.11919/j.issn.1002-0829.218026

Wang, W., Lo, P., Frasch, M., and Lufkin, T. (2000). Hmx: an evolutionary conserved homeobox gene family expressed in the developing nervous system in mice and Drosophila. *Mech. Dev.* 99, 123–137. doi: 10.1016/s0925-4773(00)00488-3