Comparative Generalized Logic Modeling Reveals Differential Gene Interactions during Cell Cycle Exit in *Drosophila* Wing Development

Mingzhou (Joe) Song†, Chung-Chien Hong†, Yang Zhang†, Laura Buttitta‡, Bruce A. Edgar‡
†Department of Computer Science, New Mexico State University, Las Cruces, U.S.A.
‡Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, U.S.A.

Abstract: A comparative interaction detection paradigm is proposed to study the complex gene regulatory networks that control cell proliferation during development. Instead of attempting to reconstruct the entire cell cycle regulatory network from temporal transcript data, differential interactions – represented by generalized logic – are detected directly from time course transcript data under two distinct conditions. This comparative approach is scale- and shift-invariant and is capable of detecting nonlinear differential interactions. Simulation studies on *E. coli* circuits demonstrated that the proposed comparative method has substantially increased statistical power over the intuitive reconstruct-then-compare approach. This method was therefore applied to a microarray experiment, profiling gene expression in the fruit fly wing as cells exit the cell cycle, and under a condition which delays this exit, over-expression of the cell cycle regulator E2F. One statistically significant differential interaction was identified between two gene clusters that is strongly influenced by E2F activity, and suggests the involvement of the Hippo signaling pathway in response to E2F, a finding that may provide additional insights on cell cycle control mechanisms. Furthermore, the comparative modeling can be applied to both static and dynamic gene expression data, and is extendible to deal with more than two conditions, useful in many biological studies.

1 Introduction

Comparative experimental designs for gene expression studies have yet to be explored for their full potential in understanding *differential* and *conserved* gene interactions in important biological phenomena such as cell cycle control. An interaction is an association from one or more parent genes to a child gene. Because complex interactions may only stand out when contrasted, we have developed a comparative modeling paradigm to detect novel gene interactions, represented by generalized logic (**glog**), for such experiments. Our strategy, based on heterogeneity and homogeneity chi-square tests, extends meta-analysis which has been traditionally used for comparing data sets of similar studies from different researchers. Our new comparative modeling approach is designed to uncover novel gene interactions, missed by other approaches.

Our goal is to fundamentally increase sensitivity in detecting how gene interactions may be either conserved or shifted in a comparative experiment. As microarray technologies mature, many approaches to gene expression analysis have been developed. Some per-

form single-gene differential expression analysis [TTC01], ignoring either dynamics or gene interactions; while others carry out gene regulatory network reconstruction [Fri04], relying on exhaustive genome-wide perturbation experiments for mathematical accuracy. As pointed out by Bonneau [Bon08], reconstruction is often cost-ineffective and believed to be "beyond our current reach". A step forward is the strategies summarized in [TBB07] that identify conserved and differential interactions by shifted Pearson linear correlation coefficients, which do not integrate temporal associations, nonlinear interactions, or interactions involving more than two genes. Still no rigorous statistical framework exists for comparative gene interaction detection beyond pair-wise linear correlation.

Our innovation is to extend the heterogeneity and homogeneity chi-square tests by associating child gene expression with potential parent gene expression at the same or previous time points. This association takes a non-parametric form that can be highly nonlinear. Our approach generalizes the correlation-based comparisons [BL06], which can be considered a single-parent, linear, zero-delay, and static interaction. We use a glog to represent an interaction. Our approach directly assesses the contrastive strength of a pair of potential interactions, instead of reconstructing-then-comparing the interaction under each condition. An interaction will be selected if it consistently shows either similar or differential patterns. Such a strategy embraces uncertainty in glog, while other approaches assume zero variance. A remarkable property of this strategy is its determination of parents without having to estimate accurately the actual glog. Although this paper explores discrete differential gene interactions, we have also developed a nice analogous approach for continuous differential interactions [OS09]. The discrete approach captures switch-like behaviors of interactions, while the continuous approach is effective for subtle and gradual interactions, complementarily.

The biological phenomenon we examine with this approach is cell cycle exit, an event critical during the process of organism development, and mis-regulated in cancers. Normally, cells differentiating into their final fates exit the cell cycle and become unresponsive to proliferative cues, but this process is somehow blocked or disrupted in cancer. To uncover how differentiation so potently blocks the cell cycle, we have examined the process of cell cycle exit in the model organism *Drosophila melanogaster*. *Drosophila* has been a key organism for studies of the cell cycle and provides an excellent system for a wide array of genetic manipulations. The *Drosophila* wing is particularly useful for studies of cell cycle control because it is highly homogenous with over 90% of the cells consisting of a single epithelial cell type, which undergo a well-characterized temporally synchronized cell cycle exit [SP87, MCGB96, BKP$^+$07]. Due to this synchrony, it is an excellent system for a time-course study of differential genetic interactions upon cell cycle exit *in vivo*.

The final cell cycle in the wing occurs between 122-144 hours of development. Exactly how this relatively synchronous cell cycle exit is controlled remains unknown, but restraining the activity of the transcription factor complex E2F has been shown to be critical for the proper timing of exit *in vivo* [BKP$^+$07]. The E2F transcription factor complex is a master regulator of cell cycle genes, promoting expression of many genes for G1-S as well as G2-M cell cycle transitions. Consistent with its role in promoting the cell cycle, the E2F complex is a well-established target for negative regulation by tumor suppressor proteins such as Retinoblastoma. It is also positively regulated by oncogenes such as SV40 Large

T and Adenovirus E1A [vdHD08]. We have found the E2F complex to regulate the expression of ∼900 genes, covering a number of cell cycle regulators, chromatin modifiers and other factors comprising the "E2F transcriptional program".

By comparative modeling on gene expression under normal conditions and conditions where E2F activity is high, we successfully identified a significant differential interaction between two clusters of genes influenced by E2F activity during cell cycle exit. We propose that this approach uncovers novel genetic networks that are perturbed upon aberrant E2F activity, providing insight into the global function of this transcription factor *in vivo*.

2 Interactions in generalized logic and their reconstruction

Let child node X have Q quantization levels ranging from 0 to $Q-1$, controlled by K parents $z_1, z_2, \ldots, z_K$ of $Q_1, Q_2, \ldots, Q_K$ quantization levels, respectively. The glog H of node X is a function that maps all possible combinations of parent node values to values of X. We also call glog H an *interaction*. The glog can incorporate temporal dependencies by introducing time t and delays of each parent $\tau_1, \ldots, \tau_K$.

We apply chi-square test to detect an interaction from a contingency table obtained from experimental data. The number of rows in the table is $R = Q_1 Q_2 \cdots Q_K$ and the number of column is Q $n_{r,c}$ is the number of observations in which the parents take the values in the r-th row and X takes the value of c. Let $n_{\cdot,c}$ be the sum of column c. Let $n_{r,\cdot}$ be the sum of row r. Let $\bar{n}_{r,c} = n_{r,\cdot} n_{\cdot,c}/n$ be the expected count when the parents are not associated with X. Then, $\chi^2 = \sum_{r=0}^{R-1} \sum_{c=0}^{Q-1} \frac{(n_{r,c}-\bar{n}_{r,c})^2}{\bar{n}_{r,c}}$ is asymptotically chi-square distributed with $(R-1)(Q-1)$ degrees of freedom (d.f.) when the parents do not influence the child. Further details can be found in [SLLea09].

3 Differential interactions and their detection by heterogeneity tests

An interaction is *conserved* if it does not change from one condition to another; otherwise, it is *differential* if any change occurs in parent identity or strength for any parent. An interaction under two conditions can have both *homogenous* and *heterogenous* components: the former represents an overall agreement of the interaction under the two conditions; the

	0	1	2
0	X	0	0
1	0	X	0
2	0	0	X

	0	1	2
0	0	0	X
1	0	X	0
2	X	0	0

(a) Detectable linear differential interaction: 1 versus -1 for Pearson coefficients.

	0	1	2
0	X	0	X
1	X	0	X
2	0	X	0

	0	1	2
0	0	X	0
1	X	0	X
2	X	0	X

(b) Undetectable nonlinear differential interactions: 0 versus 0 for Pearson coefficients.

Figure 1: Linear correlation differential interaction detection: Detectable and undetectable.

latter represents deviation from the overall agreement.

Existing comparative methods compare interactions numerically, ignoring the variance in the estimated models. For example, pair-wise linear correlation based approaches will be effective on linear differential interaction detection (Fig. 1(a)), but not nonlinear ones

(Fig. 1(b)). Our strategy will instead consider both nonlinearity and uncertainty in two data sets collected under comparative experimental conditions. Such a consideration enables much greater statistical power than other approaches.

3.1 Detect the differential interaction of a child with known parents

We develop a procedure based on chi-square statistics to determine whether a fixed topology interaction shows any significant shift under two conditions. The null hypothesis assumes no interaction between the parents and the child. The test statistics measure the homogenous and heterogenous components in interactions, illustrated in Fig. 2.

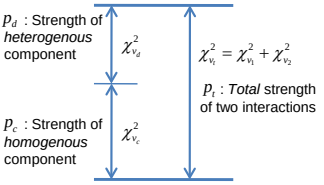


Figure 2: Components in an interaction.

Let the two data sets (temporal or static), collected under two different conditions, be $\mathcal{T}_1$ and $\mathcal{T}_2$. Let $\pi \rightarrow X$ represent an interaction from parents π to child X. We first obtain the contingency tables C_1 and C_2 from $\mathcal{T}_1$ and $\mathcal{T}_2$, respectively, associated with $\pi \rightarrow X$. $\chi^2_{v_1}$, d.f. v_1, and p-value p_1 are computed from C_1, so do $\chi^2_{v_2}$, v_2, and p_2.

The total chi-square from two interactions is $\chi^2_{v_t} = \chi^2_{v_1} + \chi^2_{v_2}$, with d.f. $v_t = v_1 + v_2$ and p-value p_t. This statistic measures by p_t the total strength of any interaction under either conditions, regardless of differential or conserved.

The homogenous component is the conserved portion of an interaction under two conditions. A contingency table C_{pool} is filled, using parent and child values, from both $\mathcal{T}_1$ and $\mathcal{T}_2$. From C_{pool}, one can compute $\chi^2_{v_c}$ with d.f. v_c and p-value p_c, which is the strength of interaction homogeneity under different conditions.

Algorithm 1 Decide-Interaction-Type(X, π, $\mathcal{T}_1$, $\mathcal{T}_2$, α)

1: Form contingency table C_1 for $\pi \rightarrow X|\mathcal{T}_1$, C_2 for $\pi \rightarrow X|\mathcal{T}_2$, and C_{pool} for $\pi \rightarrow X|\mathcal{T}_1, \mathcal{T}_2$
2: Calculate heterogenous component $\chi^2_{v_d}$ and strength p_d
3: Calculate homogenous component $\chi^2_{v_c}$ and strength p_c
4: Calculate total chi-square $\chi^2_{v_t}$ and total strength p_t
5: **if** heterogenous component is significant ($p_d \leq \alpha$) **then**
6: **if** total chi-square is significant ($p_t \leq \alpha$) **then**
7: comparative interaction type $\leftarrow$ absolute differential
8: **else**
9: comparative interaction type $\leftarrow$ relative differential
10: **end if**
11: **else if** homogenous component is significant ($p_c \leq \alpha$) **then**
12: comparative interaction type $\leftarrow$ conserved
13: **else**
14: comparative interaction type $\leftarrow$ null
15: **end if**
16: Return the comparative interaction type and $\{C_1, C_2, C_{\text{pool}}, \chi^2_{v_d}, \chi^2_{v_c}, \chi^2_{v_t}, p_d, p_c, p_t\}^{\pi}$

The heterogenous component is the differential portion of an interaction under two con-

ditions, defined by $\chi^2_{v_d} = \chi^2_{v_1} + \chi^2_{v_2} - \chi^2_{v_c}$ with d.f. $v_d = v_1 + v_2 - v_c$ and p-value p_d, which is the strength of interaction heterogeneity under different conditions.

For a parent set π, Algorithm 1 determines the interaction type: conserved, absolute or relative differential, and null. Our principle is that a pair of interactions is considered differential if it has a significant heterogenous component regardless of the significance of its homogenous component. We further classify a differential pair to be *relative* differential if the total chi-square is insignificant and otherwise *absolute* differential.

A simulation study to demonstrate the power advantage is shown in Fig. 3. The power gain can be as high as about 40% when the noise is at an intermediate level.

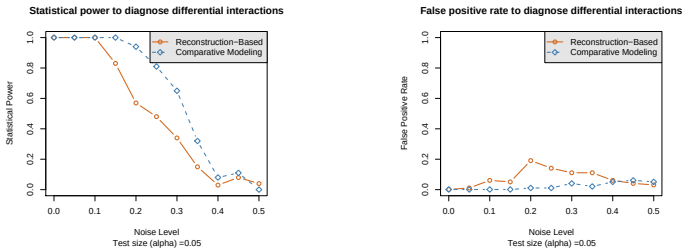


Figure 3: Advantage of comparative modeling versus reconstruct-then-compare in statistical power and false positive rate of differential interactions when parent are fixed. The example was based on a 2-parent binary interaction.

3.2 *Ab initio* comparative interaction modeling in networks of unknown topology

In *ab initio* comparative modeling, we find differential or conserved interactions of each child when parents identities are unknown. This is thus both a modeling problem to determine the most likely parents for each child, as well as a detection problem to check differential interactions. The rationale of such an approach lies in that it is unlikely for non-parents to show consistently differential or conserved interactions with a child.

Table 1: Selection of parent sets.

Interaction $\pi_1 \to X$	**Interaction** $\pi_2 \to X$	**Condition**	**True**	**False**
Conserved	Conserved	$p_c^{\pi_1} \leq p_c^{\pi_2}$	π_1	π_2
Abs. differential	Abs. differential	$p_t^{\pi_1} \leq p_t^{\pi_2}$	π_1	π_2
Rel. differential	Rel. differential	$p_d^{\pi_1} \leq p_d^{\pi_2}$	π_1	π_2
Conserved	Abs. differential	$p_c^{\pi_1} \leq p_t^{\pi_2}$	π_1	π_2
Abs. differential	Conserved	$p_c^{\pi_2} \leq p_t^{\pi_1}$	π_2	π_1
Conserved or abs. diff.	Rel. diff.	-	π_1	π_1
Rel. diff.	Conserved or abs. diff.	-	π_2	π_2
null	null	$p_t^{\pi_1} \leq p_t^{\pi_2}$	π_1	π_2
non-null	null	-	π_1	π_1
null	non-null	-	π_2	π_2

We compare two parent sets, π_1 and π_2, for child X, using Table 1. The interaction types under each parent set are determined first. We assume all p-values have been adjusted for multiple comparisons. The selection of parent set is based on four interaction types. The same type is compared by the p-values associated with that type. When the types differ, p-values are not compared but a prioritized list of conserved or absolute differential, relative

differential, and null is used. The principle is that a conserved or absolute differential parent set is selected over a relative differential one and non-null is over null. If both are null, the parent set with smaller p_t is selected.

4 Simulation study on comparing 78 pairs of *E. coli* circuits

The simulation study in Fig. 3 indicates that for known parents, the statistical power for comparative modeling is higher to reconstruct-then-compare under the same false positive rate. We now evaluate the performance of *ab initio* comparative modeling (Section 3.2), in reference to the reconstruct-then-compare approach. We used 13 *E. coli* networks [GEHL02], also called circuits, to form 78 pairs of circuits for comparison. Each circuit has four binary nodes: the inducible repressors (LacI and TetR), λ CI, and GFP. All circuits are first Markovian with a maximum of one parent for each child.

We evaluate the performance node-wise. An interaction for a node is true negative (TN) if it is conserved and announced so, false positive (FP) if conserved announced differential, false negative (FN) if differential announced conserved, and true positive (TP) if differential announced so. The performance of comparing two networks is accumulated over all the nodes in them. The network TNs is the total number of TN children, FPs the total number of FP children, FNs the total number of FN children, and TPs the total number of TP children.

The noise model is defined such that one node at a particular expression level is more likely to jump to its adjacent levels:

$$P(j|i,\theta) = \begin{cases} \left(1 - \frac{|j-i|}{\sum_{d=0}^{K-1}|d-i|}\right)\frac{\theta}{K-1}, & j \neq i \\ 1-\theta, & j = i \end{cases} \quad (1)$$

where θ denotes the noise level (from 0 to 1)[1], j denotes the noisy version of true value i, and K is the number of quantization levels.

Figure 4 shows the performance advantage of comparative modeling versus the reconstruct-then-compare approach.[2] When noise level is relatively high (0.1), comparative modeling significantly outperformed: its TPs is almost twice of reconstruct-then-compare. This implies comparative modeling can detect differential interactions more accurately without increasing FPs.

5 Differential gene interactions in cell cycle exit in *Drosophila* wings

We next applied comparative modeling to the study of cell cycle control during development *in vivo*. For this study, we obtained transcriptomic profiles of *Drosophila* wings

[1] For comparative modeling, the worst noise is 0.5.

[2] The non-monotonic ROC of the reconstruct-then-compare approach is expected as a pair of differential interactions involving a null and a non-null can become conserved of two null interactions when α increases.

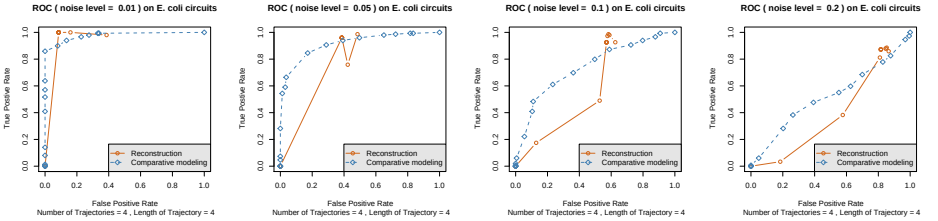


Figure 4: ROC advantage of comparative modeling versus reconstruct-then-compare, demonstrated by a simulation study on 78 pairs of *E. coli* circuits.

during cell cycle exit under normal conditions, and conditions of excessive E2F activity.

Microarray experiments – Ten pupal wings from either control animals (E2F-) or animals expressing the E2F/DP transcription factor complex under the control of the Gal4/UAS system (E2F+) were dissected at 0h, 24h, and 36h after pupa formation (APF). RNA was isolated using Trizol, and cDNA synthesis was performed with one subsequent round of T7-dependent linear RNA amplification using the commercially available Message AmpTM kit from Ambion. Amplified RNA labeled and hybridized to Nimblegen *Drosophila* expression arrays of 15,473 probes according to the manufacturer's specifications. Hybridizations were repeated 4 times with independently obtained samples. Microarray scanning and normalization was performed as recommended by the manufacturer. Importantly, cell cycle exit occurs at 24h APF under normal conditions (E2F-), while under E2F+ conditions cells go through an extra cycle and instead exit at 36h APF [BKP$^+$07].

Preprocessing – Two-way ANOVA on time (24h/36h), condition (E2F+/-), and their interaction was applied to filtered out genes insignificantly differentially expressed, resulting in 5,867 selected out of 15,473. We performed hierarchical clustering to form 127 groups of linearly correlated transcripts at 24h and 36h. A total of 127 representatives that best represent transcripts in each cluster were selected. Genes in the same cluster are considered mathematically equivalent and only the representatives were used in the subsequent modeling. A joint quantization was applied to convert continuous gene expression levels at 0h, 24h, and 36h to discrete levels of low, intermediate, and high.

Comparative modeling – Comparative glog interaction modeling was applied to data at 24h and 36h to contrast the interactions under E2F+ versus E2F-. The α-level used was 0.05. The maximum number of parents is 1.

Differential interactions – The only significant differential interaction detected is from cluster C125(22) to C34(59). The number in the parentheses is the total number of genes in that cluster. The original gene expression levels in the two clusters are shown in Fig. 5.

Table 2 shows the observed differential interaction between C125(22) to C34(59). The C125(22)→C34(59) interaction contains a significant heterogenous component ($p_d = 0.031$) and is also overall significant ($p_t = 0.039$), indicating a consistent shift in the way the two clusters interact under E2F+ or - conditions.

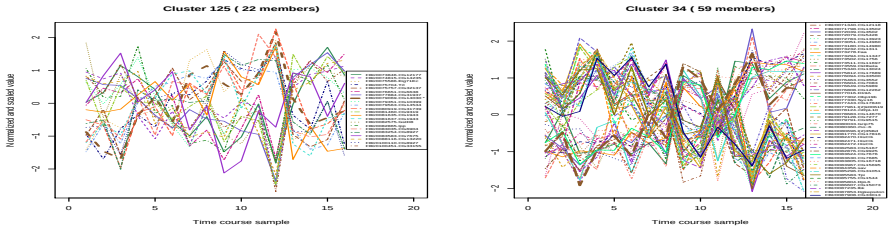


Figure 5: Expression levels (scaled and shifted) of transcripts in clusters C125(22) and C34(59). Time course sample index 1 to 8 represent 8 replicates under E2F- (1 to 4 at 24h; 5 to 8 at 36h). Sample index 9 to 16 represent 8 replicates under E2F+ (9 to 12 at 24h; 13 to 16 at 36h).

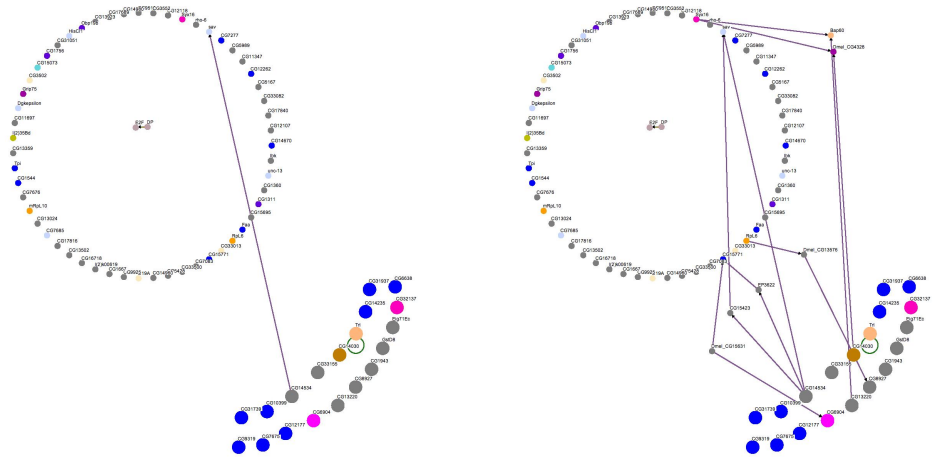
Table 2: Interaction of E2F and C125(22) with C34(59). The numbers in the table represent the occurrences of the associations in the observed expression data.

		C34(59)		
E2F	C125(22)	Low	Intermediate	High
-	Low	4	1	0
-	Intermediate	0	1	1
-	High	1	0	0
+	Low	0	0	2
+	Intermediate	0	2	0
+	High	4	0	0

6 Discussion

From this preliminary analysis we identified two clusters of genes, C125(22) and C34(59), that display a differential interaction under high E2F activity. Surprisingly, few of the genes in these clusters have known roles in cell cycle control, and none have known interactions with E2F. We have examined whether any genes within these clusters have any known genetic or physical interactions with each other, either directly or through secondary partners, using the FlyGRID database and the Osprey network visualization program. Figs. 6(a) and 6(b) show the results of the analysis. While we found no direct known interactions between E2F and the two clusters, we do find a single known direct interaction between the clusters, via CG14534 binding to Salvador (Sav) in a yeast two-hybrid protein binding assay (Fig. 6(a)). Sav is a scaffolding protein, known to be a key component of the Hippo pathway, a pathway involved in cell growth and proliferation [Edg06]. This interaction will therefore have the highest priority for further validation.

Additionally, these two clusters have multiple interactions through secondary partners. For example, CG14534 interacts with 3 targets in C34: CG15771, Syx16 and Sav, via protein-protein binding through secondary partners. CG14030 interacts with Syx16 (C34) through a secondary partner CG4328 and CG13220 (C125) also interacts with Syx16 (C34) via the chromatin modifier Bap60. Thus additional connections can be drawn through intermediate partners with Sav and Syx16 being the most highly connected targets in C34. However, two more interactions between these clusters are independent of the highly connected Syx16 and Sav nodes. They are: the CG6904 (C125) interaction with CG15771 (C34) via CG15631 and the CG8927 (C125) interaction with RpL6 (C34) through CG13576. These



(a) A known genetic and protein-protein interaction from CG14534 in C125(22) to Sav in C34(59).

(b) Genetic and protein-protein interactions including secondary interactors.

Figure 6: Known direct and secondary interactions among genes in C125(22) and C34(59) were provided by FlyGRID and Osprey. Genes in C34 are displayed as small nodes in a circular array where colors indicate different gene ontology annotations. Genes in C125 are displayed as large nodes aligned at right. E2F and DP are displayed as small nodes in the center of the C34 circular array. Known genetic or physical interactions are represented as edges between nodes with purple edges indicating a physical interaction via yeast two-hybrid assays.

results suggest several potential networks for further investigation (Fig. 6(b)).

Interestingly, a genetic interaction between the Hippo signaling pathway and E2F was recently described [NF08], where repression of Hippo signaling resulted in increased E2F expression and activity. In contrast, our work suggests that activation of E2F also alters the level of Hippo signaling via changes in Sav expression. Together these genetic interactions could result in a feedback loop, stably coordinating changes in E2F activity during development *in vivo* with compensatory alterations in Hippo signaling. We plan to further test this hypothesis by direct genetic experiments examining Hippo signaling *in vivo*.

Importantly, our modeling approach allows new interactions present only under certain conditions to be uncovered. Therefore we do not expect that many of the important interactions will be identified by the genome-wide analyses present in the database, which are done exclusively under normal conditions. To address this in future work, we can systematically test the requirement for certain genes in cluster C125 on the induction of genes in C34 under high E2F activity at 36h. This could be carried out using gene specific RNAis to knock-down the levels of highly connected genes in C125 to test the effects on transcripts in C34 by quantitative RT-PCR.

We have demonstrated that novel genetic interactions can be proposed from modeling gene expression associations at the same time point. However our total sample size of 16 for the comparative analysis is small. In future work, by doubling the sample size, the statistical

power is expected to improve substantially. By increasing the number of time points, we will expand our efforts to detect differential temporal interactions. We anticipate comparative modeling will enable more fundamental understanding of gene expression programs either within a species under different conditions or across species under same conditions.

References

[BKP+07] L Buttitta, A Katzaroff, C Perez, A de la Cruz, and BA Edgar. A double-assurance mechanism controls cell cycle exit upon terminal differentiation in *Drosophila*. *Dev. Cell*, 12(4):631–643, 2007.

[BL06] J Berg and M Lässig. Cross-species analysis of biological networks by Bayesian alignment. *Proc Natl Acad Sci U S A*, 103(29):10967–10972, 2006.

[Bon08] R Bonneau. Learning biological networks: from modules to dynamics. *Nat Chem Biol*, 4(11):658–64, 2008.

[Edg06] BA Edgar. From cell structure to transcription: hippo forges a new path. *Cell*, 124(2):267–273, 2006.

[Fri04] N Friedman. Inferring cellular networks using probabilistic graphical models. *Science*, 303:799–805, 2004.

[GEHL02] CC Guet, MB Elowitz, W Hsing, and S Leibler. Combinatorial synthesis of genetic networks. *Science*, 296:1466–1470, 2002.

[MCGB96] M Milan, S Campuzano, and A Garcia-Bellido. Cell cycling and patterned cell proliferation in the *Drosophila* wing during metamorphosis. *Proc Natl Acad Sci U S A*, 93(21):11687–11692, 1996.

[NF08] BN Nicolay and MV Frolov. Context-dependent requirement for dE2F during oncogenic proliferation. *PLoS Genet.*, 4(10):e1000205, 2008.

[OS09] Z Ouyang and M Song. Comparative identification of differential interactions from trajectories of dynamic biological networks. In *Proceedings of German Conference on Bioinformatics*, Halle, Germany, September 2009.

[SLLea09] M Song, CK Lewis, ER Lance, and et al. Reconstructing generalized logical networks of transcriptional regulation in mouse brain from temporal gene expression data. *EURASIP J Bioinform Syst Biol*, 2009. Article ID 545176, 13 pages.

[SP87] M Schubiger and J Palka. Changing spatial patterns of DNA replication in the developing wing of *Drosophila*. *Dev Biol.*, 123(1):145–153, 1987.

[TBB07] I Tirosh, Y Bilu, and N Barkai. Comparative biology: beyond sequence analysis. *Curr Opin Biotechnol*, 18(4):371–377, 2007.

[TTC01] VG Tusher, R Tibshirani, and G Chu. Significance analysis of microarrays applied to the ionizing radiation response. *PNAS*, 98:5116–5121, 2001.

[vdHD08] S van den Heuvel and N Dyson. Conserved functions of the pRb and E2F families. *Nat Rev Mol Cell Biol.*, 9(9):713–724, 2008.