SUPPLEMENTARY DATA

Hunting complex differential gene interaction patterns across molecular contexts

Mingzhou Song[1,*], Yang Zhang[1], Alexia J. Katzaroff[2,3], Bruce A. Edgar[4], Laura Buttitta[2,†,*]

1 Department of Computer Science, New Mexico State University, Las Cruces, NM 88003, USA
2 Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA
3 Molecular and Cellular Biology Graduate Program, University of Washington, Seattle, WA 98195, USA
4 German Cancer Research Center (DKFZ)–Center for Molecular Biology Heidelberg (ZMBH) Alliance, Im Neuenheimer Feld 282, D-69120 Heidelberg, Germany

∗ E-mail: Corresponding joemsong@cs.nmsu.edu, buttitta@umich.edu

† Current address: Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA

Contents

1 Supplementary Figures **2**

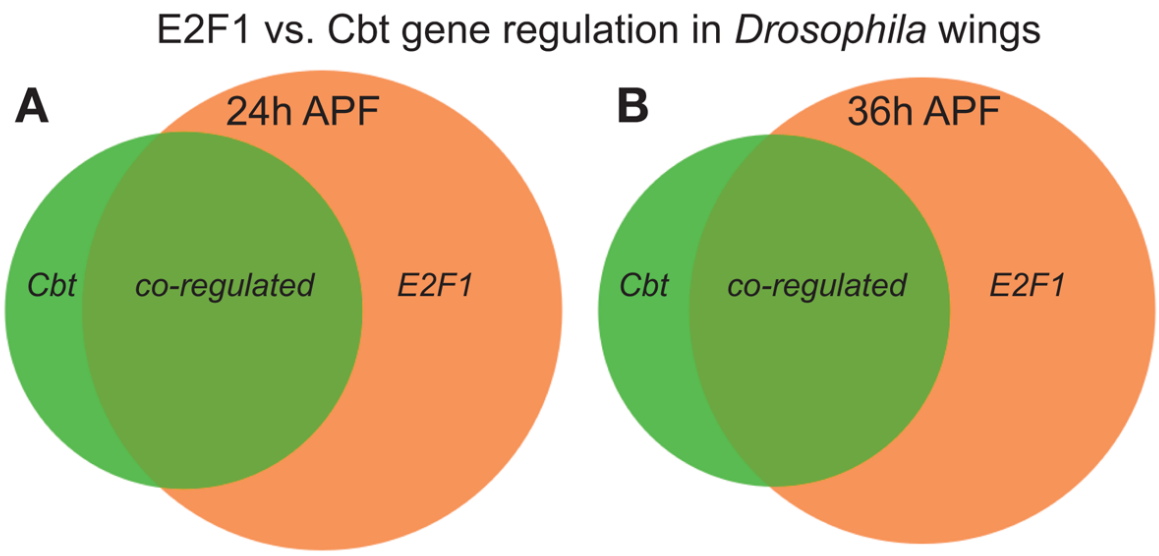
2 Supplementary Results **5**
2.1 Biological support for predicted parent-child interactions 5
2.2 Novel gene interactions promoting proliferation in response to E2F and Cabut 7

3 Supplementary Methods **8**
3.1 The asymptotic null distribution of the heterogeneity statistic 8
3.2 The chi-square test for change in working zone 13
3.3 A discrete noise model and its estimation 13
3.4 Benchmarking on differential gene networks regulating yeast cell cycle 14
3.5 *Drosophila* microarray data set 19

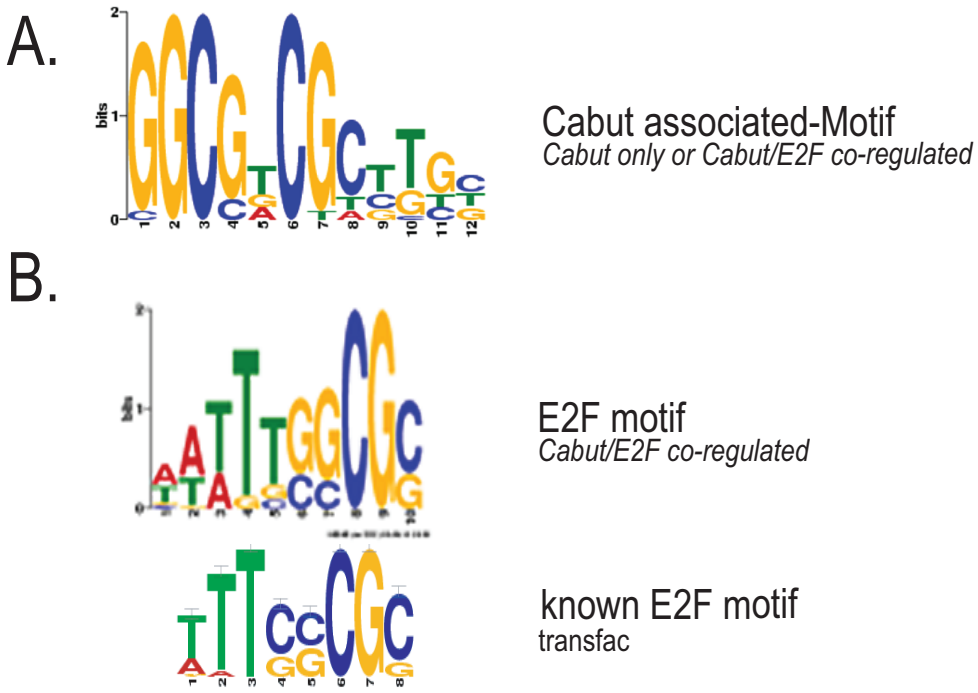
List of Supplementary Figures

S1 E2F1 and Cabut regulate distinct but overlapping sets of genes 2
S2 A novel Mad-like motif is found in Cabut-regulated and Cabut/E2F1 co-regulated genes 3
S3 The histogram of estimated noise levels from the *Drosophila* microarray data 4
S4 Correlation of two heterogeneity chi-square estimates 12
S5 Working zone change of a random variable under two conditions. 13
S6 Two budding yeast cell cycle networks 15
S7 The logical rules for the first budding yeast cell cycle model 15
S8 The logical rules for the second budding yeast cell cycle model 16
S9 Two fission yeast cell cycle networks 16
S10 The logical rules for the first fission yeast cell cycle model 16
S11 The logical rules for the second fission yeast cell cycle model 17

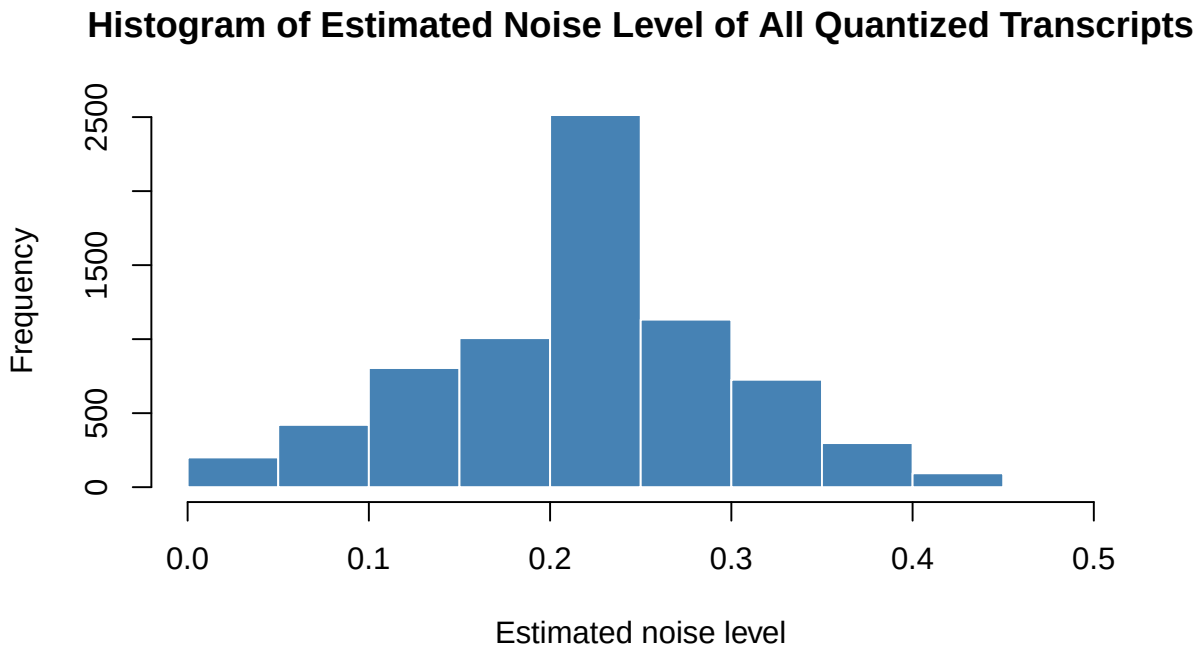
1 Supplementary Figures



Supplementary Figure S1. E2F1 and Cabut regulate distinct but overlapping sets of genes. Using microarrays, we compared gene expression changes in E2F1/DP expressing wings, and Cabut expressing wings, at two time points, 24h after puparium formation (h APF) and 36h APF. Venn diagrams indicate total numbers of transcript changes with a fold change of 1.3 or more ($> \log_2 \pm 0.4$) obtained for E2F1 and Cabut overexpression. Panel A illustrates the set relationship between Cabut and E2F regulated genes at 24h APF; Panel B for the set relationship at 36h APF. The E2F1 and Cabut independently regulate a smaller set of non-overlapping transcripts at each timepoint, but also co-regulate a larger number of transcripts at both time points.



Supplementary Figure S2. A novel Mad-like motif is found in Cabut-regulated and Cabut/E2F1 co-regulated genes. Using MEME we identified motifs enriched in gene clusters displaying differential interactions with working zone changes as well as the top 200 most strongly E2F1 and Cabut co-upregulated genes. A novel Mad-like motif was significantly enriched in the upstream regulatory regions of genes in clusters corresponding to Cabut regulated and E2F1/Cabut co-regulated genes (A). We also readily identified known E2F binding sites in E2F1 regulated genes that are not as strongly regulated by Cabut as well as E2F1/Cabut co-regulated genes (B).



Supplementary Figure S3. The histogram of estimated noise levels of all quantized transcripts from the *Drosophila* microarray data. Each transcript is quantized using its observations across 36 gene expression arrays at time points 0, 24h, 36h APF. The noise level of each transcript under an identical condition is obtained by maximum likelihood estimation on replicates of the transcript under that identical condition. The average estimated noise level is 0.22.

2 Supplementary Results

2.1 Biological support for predicted parent-child interactions

To confirm our findings of differential interactions between over-expression of E2F or Cabut and the control (**Fig. 3** in the main text) obtained from the perturbed and normal cell cycle data, we matched 5 gene interactions to those in BioGRID (Chatr-Aryamontri et al., 2013). We also found 8 known interactions directly involving E2F. The 13 matched interactions are listed in **Supplementary Table S1**. All these interactions are reported in *Drosophila melanogaster*, though we allowed other organisms in our BioGRID search. These interactions suggest a network of genes promoting proliferation in response to E2F and Cabut activity. Interestingly, we note some interactions are also predicted to limit cell cycling in response to high E2F activity (negative regulatory loops), but not in response to Cabut. This is consistent with our previous finding that a robust cell cycle mechanism indeed limits cell cycle exit delay in response to high E2F activity (Buttitta et al., 2010).

Supplementary Table S1. Biological evidence of detected known genetic interactions illustrates how E2F or Cabut promote cell proliferation.

Parent cluster	Child cluster	Parent	Child	References
C137(6)	C131(9)	CG3008	Ebi	(Guruharsha et al., 2011)
C39(34)	Dah	CG8247	Dah	(Giot et al., 2003)
C21(50)	CG6084	Ntf-2	CG6084	(Guruharsha et al., 2011)
C51(45)	tos	CG9938 (NDC80)	tos	(Giot et al., 2003)
C10(61)	ncd	sub	ncd	(Giunta et al., 2002)
E2F	C40(20)	E2F	DREF	(Hochheimer et al., 2002)
E2F	C28(48)	E2F	CycA	(Hayashi and Yamaguchi, 1999) (Staehling-Hampton et al., 1999)
E2F	C39(34)	E2F	brm	(Staehling-Hampton et al., 1999)
E2F	C39(34)	E2F	dap	(Frolov et al., 2003) (Weng et al., 2003)
E2F	C131(9)	E2F	Ebi	(Boulton et al., 2000)
E2F	C288(9)	E2F (E2F2)	CG13900	(Stanyon et al., 2004)
E2F	Rbf2	E2F (E2F2)	Rbf2	(Korenjak et al., 2004) (Taylor-Harding et al., 2004)
E2F	CG13806	E2F (E2F2)	CG13806	(Stanyon et al., 2004)

The differential interactions can be detected between clusters, a cluster and a gene, or two genes. In the case of a cluster being involved, all members in the cluster are considered in search of a match to known interactions in BioGRID.

The 13 verified gene interactions comprise a network of components predicted to act cooperatively to promote S-phase and proliferation, thereby delaying cell cycle exit in response to ectopic E2F or Cabut activity. The biological functions of these 13 gene interactions are summarized as follows:

CG3008 – Ebi: This interaction was reported from the protein complex network derived by Guruharsha et al. (2011) using yeast two-hybrid screening. CG3008 is the fly ortholog of the human RIO kinases which were recently shown to promote aberrant growth and proliferation in a *Drosophila* model of glioblastoma (Read et al., 2013). Ebi is a component of a transcriptional co-repressor complex that mediates histone deacetylation to repress gene expression via HDAC3 (Qi et al., 2008) and was recently shown to interact with the E2F repressor Rbf to inhibit cell cycle gene expression (Lim et al., 2013). This interaction suggests

RIO kinases may play a role in modifying the Ebi co-repressor to modulate its repressive activity on G1-S phase cell cycle gene expression

CG8247 – Dah: This interaction was previously reported from the protein interaction map of *Drosophila* based on fly proteome data (Giot et al., 2003). CG8247 is an uncharacterized gene. Dah is a component of actin contractile ring assembly, essential for completion of the cell cycle at cytokinesis (Albertson et al., 2008)

Ntf-2 – CG6084: This interaction was reported from the protein complex network derived by Guruharsha et al. (2011), Ntf-2 is a part of a multiprotein complex that facilitates import of the Relish transcription factors into the nucleus (Bhattacharya and Steward, 2002). CG6084 is an uncharacterized gene with homology to human Aldehyde reductases

CG9938 (NDC80) – tos: This interaction was previously reported from the fly proteome data (Giot et al., 2003). CG9938, also known as NDC80, is an outer kinetochore protein, critical during the cell cycle in mitosis for proper spindle assembly. tos (tosca) is an endonuclease involved in DNA repair pathways, which are induced in response to cell cycle entry

sub – ncd: sub (subito) and ncd are both kinesin-family microtubule motor proteins required for chromosome segregation during mitosis. Both interact in mitotic spindle formation (Giunta et al., 2002)

E2F – DREF: DREF is a promoter binding component that interacts with the TRF2 complex to direct core promoter recognition of hundreds of cell cycle and cell growth genes including the proliferating cell nuclear antigen (PCNA) gene (Hochheimer et al., 2002). The upstream regions of most direct E2F transcriptional targets also contain DREF binding sites (O'Keefe et al., 2012). This interaction suggests a positive-regulatory loop to promote cell cycle progression in response to E2F activity

E2F – CycA: CycA is the S and G2 cyclin promoting cell cycle progression (Lehner and O'Farrell, 1989). CycA with Cdc2 also cooperatively inhibits transcriptional activity of E2F to limit S-phase (Hayashi and Yamaguchi, 1999). This interaction suggests a potential negative-feedback loop to partially limit abnormal cell cycle progression in response to aberrant E2F activity

E2F – brm: brm modifies chromatin structure facilitating DNA binding of transcription factors (Staehling-Hampton et al., 1999), and both physically and genetically interacts with the cell cycle regulator and E2F target gene CycE (Brumby et al., 2002)

E2F – dap: Dacapo (dap) is the *Drosophila* homolog of p21, which inhibits S-phase entry by repression of CycE/cdk activity (Frolov et al., 2003; De Nooij et al., 1996; Lane et al., 1996). This interaction suggests a negative-feedback loop to limit abnormal cell cycle progression in response to E2F activity

E2F – Ebi: As mentioned above, Ebi was recently shown to interact with the E2F repressor Rbf to inhibit cell cycle gene expression (Lim et al., 2013). Ebi is highly expressed only in E2F+ but not in Cabut+. This interaction suggests a negative-feedback loop to limit abnormal cell cycle progression in response to E2F activity, which is not observed in response to Cabut activity

E2F – CG13900: This interaction was identified through high throughput yeast two-hybrid screening using the E2F paralog E2F2 as bait and CG13900 as prey (Stanyon et al., 2004). CG13900 is the *Drosophila* ortholog of the splicing factor 3b, subunit 3 (SF3B3), conserved from yeast to humans. In yeast, SF3B3 has been suggested to be involved in splicing transcripts encoding a protein (or proteins) required for G2/M transition (Habara et al., 2001)

E2F – Rbf2: E2F2 and Rbf2 form a complex to suppress cell cycle gene expression (Taylor-Harding et al., 2004). Rbf2 is highly expressed only in E2F+, but not in Cabut+. This interaction suggests a negative-feedback loop to limit abnormal cell cycle progression in response to E2F activity, which is not observed in response to Cabut activity.

E2F – CG13806: This interaction was identified through high throughput yeast two-hybrid screening using the E2F paralog E2F2 as bait and CG13806 as prey (Stanyon et al., 2004). CG13806 is an uncharacterized gene encoding a protein with a chitin binding domain. CG13806 is repressed in both E2F+ and Cabut+ conditions, suggesting it is a novel common target for repression upon cell cycle exit delay.

2.2 Novel gene interactions promoting proliferation in response to E2F and Cabut

In addition to the interactions confirmed from the literature above, we also predicted parent-child interactions with genes that do not have any known interactions within the BioGRID database (Chatr-Aryamontri et al., 2013) or literature search. Function via GO terms and expression patterns of the target child genes are summarized in **Supplementary Table S2**. These suggest novel interactions for future investigation. Importantly, certain classes of genes were conspicuous within this group, further suggesting a coherent network modulated by E2F and Cabut to promote proliferation. These include several genes involved in DNA replication and repair during S-phase, as well as centrosome assembly, duplication and function during G2 and mitosis.

Supplementary Table S2. Function and gene expression patterns of target genes of the shared differential interactions under E2F+ and Cabut+.

Gene	GO terms	References	Relative Expression		
			E2F+	Cabut+	Control
RanGap	meiotic chromosome segregation	(Kusano et al., 2003)	++	-	- -
RpA-70	DNA-dependent DNA replication	(Mitsis et al., 1993)	++	-	- -
tefu	cell cycle check point	(Sekelsky et al., 2000)	+	-	- -
CG13745	mitotic G2 DNA damage checkpoint	(Kondo and Perrimon, 2011)	0	-	- -
DnaJ-60	protein folding	(Iliopoulos et al., 1997)	0	+	++
skl	apoptotic process; cell death; developmental programmed cell death	(Bergmann et al., 2003)	+	0	- -
l(3)s2214	asymmetric cell division; centriole replication; centrosome duplication	(Basto et al., 2006)	+	-	- -
CG11175	centrosome organization	(Dobbelaere et al., 2008)	-	-	+ -
IM10	defense response; humoral immune response; Toll signaling pathway	(Uttenweiler-Joseph et al., 1998)	-	0	+
CG13162	centrosome duplication; mitotic spindle organization	(Goshima et al., 2007)	0	-	- -
CG9919	N/A		0	0	++
CG15234	N/A		0	0	- -
CG13806	N/A		-	-	++

The functions are described by GO terms. In the relative expression column, the signs of "-", "0", "+" indicate relatively under-, intermediate-, and over-expressed among the three conditions of E2F+, Cabut+, and control, with "- -" or "++" suggesting consistently under- or over-expressed in the replicates. "+ -" suggests both under- and over-expressed replicates are observed.

Here we provide further detail on some novel interactions with clear links to the cell cycle.

Interactions associated with priority gene cluster C15(7): Cluster C15(7) contains seven genes and are implicated in five differential interactions from the control consistent between E2F+ and Cabut+. Two genes in C15(7) are XRCC1 and kuk known to have cell cycle related function:

XRCC1 is involved in the repair of DNA damage and interacts with several E2F transcriptional targets including PCNA (Ravi et al., 2009)

kuk is involved in cell aging (Brandt et al., 2008), nucleus organization and nuclear breakdown at mitosis (Hampoelz et al., 2011)

The targets (children) of C15(7) include the following genes that may be implicated in cell cycle:

RanGap is involved in mitotic spindle assembly and segregation. The ran GTPase regulates mitotic spindle assembly (Kalab et al., 1999)

RpA-70 is involved in DNA-dependent DNA replication (Mitsis et al., 1993)

tefu is the *Drosophila* homolog of human ATM, which is a kinase activated in response to DNA damage for repair and cell cycle regulation (Sekelsky et al., 2000)

CG13745 is the fly homolog of FANCI, a components of the Fanconia Anemia Complementation Group I, involved in DNA replication and the response to DNA damage (Kondo and Perrimon, 2011)

C221(16) – DnaJ-60: DnaJ-60 is an uncharacterized gene encoding a protein with a DNA-J domain, which is a characteristic domain in chaperone proteins that assists in protein folding (Iliopoulos et al., 1997)

C32(2) – skl: skl binds to the Inhibitor of Apoptosis (IAP) family of proteins to block their anti-apoptotic effects (Bergmann et al., 2003)

C39(34) – l(3)s2214: Sas-4 (l(3)s2214) is essential for centrosome organization, specifically centriole replication during the G2-phase of the cell cycle (Basto et al., 2006)

C380(4) – CG11175: Rcd6 (CG11175) is essential for centrosome organization, specifically recruitment of peri-centriolar material during centrosome replication (Dobbelaere et al., 2008)

C188(5) – IM10: IM10 is involved in fly defense and immune response via the Toll signaling pathway (Uttenweiler-Joseph et al., 1998)

C39(34) – CG13162: ana3 (CG13162) is essential for centrosome assembly and duplication (Goshima et al., 2007)

3 Supplementary Methods

3.1 The asymptotic null distribution of the heterogeneity statistic

Here, we prove the asymptotic null distribution of the heterogeneity test statistic χ^2_d to be chi-squared as stated in the main text. The was loosely introduced in a couple of textbooks (Steel et al., 1997; Zar, 2010; Sheskin, 2011). The proof is based on the work of Lancaster (1949) and Irwin (1949). We first summarize their results on partitioning a contingency table into a number of i.i.d. standard normal variables, and then give our proof for χ^2_d to be asymptotically chi-square distributed.

Chi-square partition of a contingency table

We summarize the theoretical result on partitioning the chi-square of an $r \times s$ contingency table to the sum of independent chi-squares of 1 degree of freedom. Lancaster (1949) and Irwin (1949) laid the theoretical foundation for chi-square partition of contingency table by applying two orthogonal transforms.

Lemma 1. *A contingency table can be asymptotically partitioned into a matrix of independent standard normal random variables, and the sum of their squares is the same as the chi-square of the original contingency table, when the row and column variables have no associations.*

This lemma is proved in (Lancaster, 1949). We summarize the proof below with changed notation for consistency within this paper.

Proof. Let $[p_{ij}]$ be a matrix of the cell population probabilities of an $r \times s$ contingency table following a multinomial distribution. Let row probabilities $p_{i\cdot}$ be the sum of row i of $[p_{ij}]$, and column probabilities $p_{\cdot j}$ the sum of column j of $[p_{ij}]$. Our derivation is based on the non-association of the row and column variables, that is,

$$p_{ij} = p_{i\cdot} p_{\cdot j}$$

Let $[n_{ij}]$ be an observed contingency table sampled from the above multinomial distribution with n independent trials. Let matrix $A = [a_{ij}]$ be the normalized version of $[n_{ij}]$ defined as

$$a_{ij} = \frac{n_{ij} - np_{ij}}{\sqrt{np_{ij}}} \tag{S1}$$

where $n = \sum_{ij} n_{ij}$.

Irwin (1949) defined the $r \times r$ row-Helmert matrix V by

$$V_{i,j} = \begin{cases} \sqrt{p_{i\cdot}} & i = 1 \\ 0 & j > i, i \neq 1 \\ -\sqrt{\frac{p_{1\cdot}+\cdots+p_{i-1\cdot}}{p_{1\cdot}+\cdots+p_{i\cdot}}} & j = i, i \neq 1 \\ \sqrt{\frac{p_{i\cdot}p_{j\cdot}}{(p_{1\cdot}+\cdots+p_{i-1\cdot})(p_{1\cdot}+\cdots+p_{i\cdot})}} & j < i \end{cases} \tag{S2}$$

and the $s \times s$ column-Helmert matrix W by replacing $p_{i\cdot}$ by $p_{\cdot j}$ in V.

Lancaster (1949) and Irwin (1949) applied two orthogonal transforms to A using

$$E = VAW^\top \tag{S3}$$

and showed that as $n \to \infty$, elements in $r \times s$ matrix $E = [e_{ij}]$ are independent standard normal variables, except that $e_{11} = 0$. We now define

$$\chi^2 = \sum_{i=2}^{r} \sum_{j=2}^{s} e_{ij}^2$$

and also

$$\chi^2_{\text{all}} = \sum_{i=1}^{r} \sum_{j=1}^{s} e_{ij}^2, \quad \chi^2_{\text{row}} = \sum_{i=1}^{r} e_{i1}^2, \quad \chi^2_{\text{col}} = \sum_{j=1}^{s} e_{1j}^2$$

With $p_{ij} = p_{i\cdot} p_{\cdot j}$ and via Stirling's approximation, we have asymptotically

$$\chi^2_{\text{all}} = \chi^2_{\text{row}} + \chi^2_{\text{col}} + \chi^2 \tag{S4}$$

which is proved on page 123 of (Lancaster, 1949) with an illustrating example on page 126 of the same paper.

If $p_{i\cdot}$ and $p_{\cdot j}$ are not estimated from the sample, we asymptotically have χ^2_{all} is chi-squared with $rs - 1$ degrees of freedom, χ^2 is chi-squared with $(r-1)(s-1)$ degrees of freedom, χ^2_{row} is chi-squared with $r - 1$ degrees of freedom, and χ^2_{col} is chi-squared with $s - 1$ degrees of freedom. The -1 in all above degrees of freedom is due to $e_{11} = 0$.

Otherwise if $p_{i\cdot}$ and $p_{\cdot j}$ are sample row and column probabilities, we have $\chi^2_{\text{row}} = \chi^2_{\text{col}} = 0$ and $\chi^2_{\text{all}} = \chi^2$ is chi-squared with $(r-1)(s-1)$ degrees of freedom.

□

The interpretation of the four chi-squares are as follows. χ^2_{all} is due to all causes of variation from np_{ij}, χ^2_{row} due to row variation from $np_{i\cdot}$, χ^2_{col} due to column variation from $np_{\cdot j}$, and χ^2 due to association or interactions beyond $n_{i\cdot}n_{\cdot j}/n$, where $n_{i\cdot}$ and $n_{\cdot j}$ are sample row and column sums of $[n_{ij}]$.

Heterogeneity of multiple contingency tables

Let K $r \times s$ contingency tables have sample sizes $n_1, \ldots, n_K$ each. Let $n = n_1 + \cdots + n_K$ be the sum of the K sample sizes. We define the pooled contingency table to be $[n_{ij}] = [\sum_{k=1}^{K} n_{ij,k}]$, where $n_{ij,k}$ is the observed count in cell (i, j) of contingency table k.

We first give a lemma on quadratic forms of i.i.d standard normal variables. A proof of this lemma can be found in (Mathai and Provost, 1992).

Lemma 2. *When $x = [x_1, \ldots, x_D]^\top$ is a vector of D i.i.d. standard normal random variables and $D \times D$ matrix A is idempotent, the quadratic form $x^\top Ax$ is chi-square distributed with* $\operatorname{rank} A$ *degrees of freedom.*

With the two lemmas, we are ready to prove the following theorem regarding the asymptotic null distribution of the heterogeneity statistic χ^2_d across K contingency tables.

Theorem 3 (Heterogeneity chi-square of K contingency tables)**.** *Under the null hypothesis of K homogenous non-interacting $r \times s$ contingency tables, the heterogeneity statistic $\chi^2_d = \sum_{k=1}^{K} \chi^2_k - \chi^2_{pool}$ is asymptotically chi-square distributed with $(K-1)(r-1)(s-1)$ degrees of freedom.*

Proof. By Lemma 1 and let $A_k = [a_{ij,k}]$ be the normalized matrix of contingency table k, we can partition A_k to residual matrix $E_k = [e_{ij,k}]$ of independent standard normal random variables.

Under the null hypothesis of homogeneous contingency tables, we assume the K conditions share the same cell probabilities $[p_{ij}]$, then the normalized pooled element

$$a_{ij}^{\text{pool}} = \frac{n_{ij} - np_{ij}}{\sqrt{np_{ij}}} \tag{S5}$$

$$= \sum_{k=1}^{K} \frac{n_{ij,k} - n_k p_{ij}}{\sqrt{np_{ij}}} \tag{S6}$$

$$= \sum_{k=1}^{K} \sqrt{\frac{n_k}{n}} \frac{n_{ij,k} - n_k p_{ij}}{\sqrt{n_k p_{ij}}} \tag{S7}$$

$$= \sum_{k=1}^{K} \sqrt{\frac{n_k}{n}} a_{ij,k} \tag{S8}$$

Thus we can represent the normalized pooled matrix A_{pool} by the normalized matrices under the K conditions:

$$A_{\text{pool}} = \sqrt{\frac{n_1}{n}} A_1 + \cdots + \sqrt{\frac{n_K}{n}} A_K \tag{S9}$$

By Lemma 1, the row and column Helmert matrices V defined in Eq. (S2) and W are all identical given that the contingency tables are all sampled from the same multinomial distribution with parameter $[p_{ij}]$. Thus the pooled contingency table is also from the same population. Therefore the partitioned residual for A_{pool} in cell (i, j) is

$$e_{ij}^{\text{pool}} = [v_{i1}, \ldots, v_{ir}] A_{\text{pool}} [w_{j1}, \ldots, w_{jr}]^\top \tag{S10}$$

$$= [v_{i1}, \ldots, v_{ir}] \left(\sqrt{\frac{n_1}{n}} A_1 + \cdots + \sqrt{\frac{n_K}{n}} A_K \right) [w_{j1}, \ldots, w_{jr}]^\top \tag{S11}$$

$$= \sqrt{\frac{n_1}{n}} [v_{i1}, \ldots, v_{ir}] A_1 [w_{j1}, \ldots, w_{jr}]^\top + \ldots + \sqrt{\frac{n_K}{n}} [v_{i1}, \ldots, v_{ir}] A_K [w_{j1}, \ldots, w_{jr}]^\top \tag{S12}$$

$$= \sqrt{\frac{n_1}{n}} e_{ij,1} + \ldots + \sqrt{\frac{n_1}{n}} e_{ij,K} \tag{S13}$$

When $i > 1$ or $j > 1$, the difference between the sum of individual condition residual squares and the square of pooled residual for cell (i, j) is

$$\chi^2_{ij} = e^2_{ij,1} + \ldots + e^2_{ij,K} - e^2_{ij,\text{pool}} \tag{S14}$$

$$= e^2_{ij,1} + \ldots + e^2_{ij,K} - \left(\sqrt{\frac{n_1}{n}} e_{ij,1} + \ldots + \sqrt{\frac{n_1}{n}} e_{ij,K}\right)^2 \tag{S15}$$

$$= [e_{ij,1}, \ldots, e_{ij,K}] \left(I - \left[\sqrt{\frac{n_1}{n}}, \ldots, \sqrt{\frac{n_K}{n}}\right]^\top \left[\sqrt{\frac{n_1}{n}}, \ldots, \sqrt{\frac{n_K}{n}}\right]\right) [e_{ij,1}, \ldots, e_{ij,K}]^\top \tag{S16}$$

Since $e_{ij,k}$ $(k = 1, \ldots, K)$ are all i.i.d. standard normal asymptotically (Lemma 1) and

$$I - \left[\sqrt{\frac{n_1}{n}}, \ldots, \sqrt{\frac{n_K}{n}}\right]^\top \left[\sqrt{\frac{n_1}{n}}, \ldots, \sqrt{\frac{n_K}{n}}\right]$$

is idempotent with rank $K - 1$, by Lemma 2 the quadratic form χ^2_{ij} gives rise to a chi-square random variable with $K - 1$ degrees of freedom.

When $i = j = 1$, $e_{11,1} = \cdots = e_{11,K} = e_{11,\text{pool}} = 0$ (Lemma 1). Therefore $\chi^2_{11} = 0$.

As $e_{ij,k}$ are independent across i, j, k, it implies all χ^2_{ij} are independent across i, j.

Let $\chi^2_{k,\text{all}}$, $\chi^2_{k,\text{row}}$, $\chi^2_{k,\text{col}}$, and χ^2_k be as derived by Lemma 1 for contingency table k. Let $\chi^2_{\text{pool,all}}$, $\chi^2_{\text{pool,row}}$, $\chi^2_{\text{pool,col}}$, and χ^2_{pool} be as derived by Lemma 1 for the pooled contingency table. It follows that we can derive four heterogeneity chi-square random variables as follows. Summing up χ^2_{ij} over all i and j, we obtain the heterogeneity chi-square due to all causes of variations across conditions

$$\chi^2_{d,\text{all}} = \sum_{k=1}^{K} \chi^2_{k,\text{all}} - \chi^2_{\text{pool,all}} = \sum_{i=1}^{r} \sum_{j=1}^{s} e^2_{ij,1} + \ldots + e^2_{ij,K} - e^2_{ij,\text{pool}} \sim \chi^2_{(K-1)(rs-1)} \tag{S17}$$

Summing up χ^2_{ij} over j, we obtain the heterogeneity chi-square due to differences between rows across conditions

$$\chi^2_{d,\text{row}} = \sum_{k=1}^{K} \chi^2_{k,\text{row}} - \chi^2_{\text{pool,row}} = \sum_{i=1}^{r} e^2_{i1,1} + \ldots + e^2_{i1,K} - e^2_{i1,\text{pool}} \sim \chi^2_{(K-1)(r-1)} \tag{S18}$$

Summing up χ^2_{ij} over i, we obtain the heterogeneity chi-square due to differences between columns across conditions

$$\chi^2_{d,\text{col}} = \sum_{k=1}^{K} \chi^2_{k,\text{col}} - \chi^2_{\text{pool,col}} = \sum_{j=1}^{s} e^2_{1j,1} + \ldots + e^2_{1j,K} - e^2_{1j,\text{pool}} \sim \chi^2_{(K-1)(s-1)} \tag{S19}$$

Summing up χ^2_{ij} over $i, j > 1$, we obtain the heterogeneity chi-square due to interactions or associations across conditions

$$\chi^2_d = \sum_{k=1}^{K} \chi^2_k - \chi^2_{\text{pool}} = \sum_{i=2}^{r} \sum_{j=2}^{s} e^2_{ij,1} + \ldots + e^2_{ij,K} - e^2_{ij,\text{pool}} \sim \chi^2_{(K-1)(r-1)(s-1)} \tag{S20}$$

□

Corollary 4. *The heterogeneity chi-square due to all causes of variations is the sum of the heterogeneity chi-squares due to row, column, and association variations.*

Proof. As it follows from Lemma 1 that

$$\chi^2_{k,\text{all}} = \chi^2_{k,\text{row}} + \chi^2_{k,\text{col}} + \chi^2 \tag{S21}$$

$$\chi^2_{\text{pool,all}} = \chi^2_{\text{pool,row}} + \chi^2_{\text{pool,col}} + \chi^2_{\text{pool}} \tag{S22}$$

and with Eq. (S17), (S18), (S19), and (S20), we obtain

$$\chi^2_{d,\text{all}} = \chi^2_{d,\text{row}} + \chi^2_{d,\text{col}} + \chi^2_d \tag{S23}$$

□

Approximation errors

As implied in Eq. (S20), there are two ways to compute the heterogeneity chi-square χ^2_d due to interactions or associations. The first is to use the pooled marginals to estimate $p_{i\cdot}$ and $p_{\cdot j}$ and compute $\chi^2_{d,\text{all}}$, $\chi^2_{d,\text{row}}$, and $\chi^2_{d,\text{col}}$. Then using Eq. (S23), we get

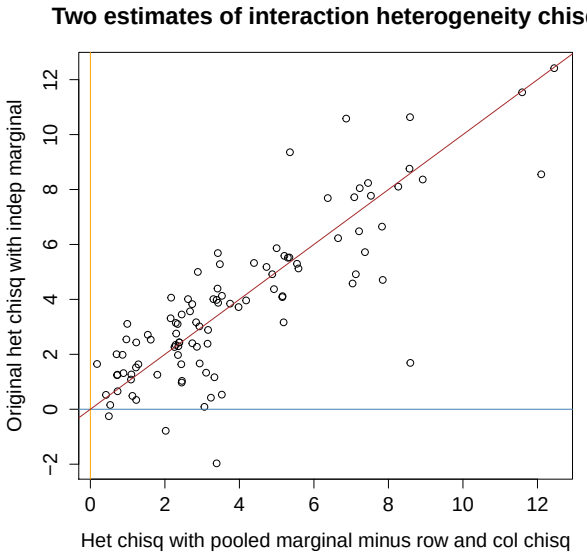
$$\chi^2_d = \chi^2_{d,\text{all}} - \chi^2_{d,\text{row}} - \chi^2_{d,\text{col}} \tag{S24}$$

By the above steps, we can avoid computing the two orthogonal transforms using the row- and column-Helmert matrices.

The second way is to compute directly χ^2_k and χ^2_{pool} using their respective marginals to estimate $p_{i\cdot}$ and $p_{\cdot j}$ for each contingency table. This is what has been used in the literature (Zar, 2010; Sheskin, 2011). However, this method would not guarantee non-negativity of χ^2_d and therefore one must take the absolute value of Eq. (S20) by

$$\chi^2_d = \left| \sum_{k=1}^{K} \chi^2_k - \chi^2_{\text{pool}} \right| \tag{S25}$$

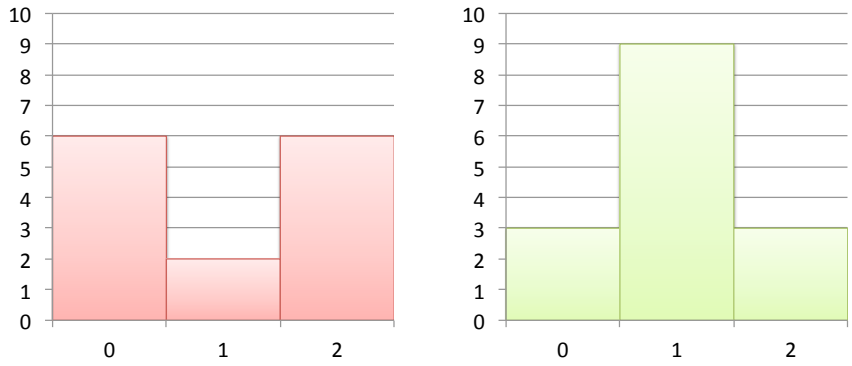
Both strategies converge to the same result as sample size $n \to \infty$. When sample size is small, the two methods are highly correlated as demonstrated in **Supplementary Fig. S4**, where some negative χ^2_d appeared by the second approach.



Supplementary Figure S4. Correlation of two heterogeneity chi-square χ^2_d estimates at small sample sizes. The horizontal axis is χ^2_d computed using Eq. (S24). The vertical axis is χ^2_d computed using Eq. (S20). The data is from two 3×3 tables of 16 versus 8 samples drawn randomly from two identical equal probability multinomial distributions. The two estimates are close but the one calculated using Eq. (S20) can be negative, though both are asymptotically accurate.

3.2 The chi-square test for change in working zone

It is often necessary, such as in differential gene interaction analysis, to require a parent set or the child to show change in dynamic range or working zone. Such changes can be important for time course data when the dynamic range of a random variable has shifted between two conditions. However, testing working zone change is beyond testing for difference between means such as the t-test and its variants. This is due to the possibility that a random variable with two different distributions can still have the same mean, for example, in the two histograms in **Supplementary Fig. S5**. To test the working zone change of a discrete random variable Z of Q levels under K



Supplementary Figure S5. Working zone change of a random variable under two conditions. In the left histogram, the variable has a mean of 1 with a valley-like distribution; in the right histogram, the variable has the same mean of 1 but a ridge-like distribution. Thus the variable has different distributions, yet with same mean under two conditions. A t-test on above two data sets will not detect mean difference and result in a p-value of 1. But in reality the two variable have very different dynamic ranges or changed working zones.

conditions, we perform a chi-square test (Press et al., 2007) to compare the K histograms of Z by the statistic

$$\chi_z^2 = \sum_{k=1}^{K} \sum_{z} \frac{(n_k(z) - \bar{n}_k(z))^2}{\bar{n}_k(z)} \tag{S26}$$

where $n_k(z)$ is the observed frequency of $Z = z$ in condition k, and $\bar{n}_k(z)$ is the expected frequency of $Z = z$ in condition k under the null hypothesis of same distributions given by

$$\bar{n}_k(z) = \frac{n_1(z) + \ldots + n_K(z)}{n_1 + \ldots + n_K} n_k \tag{S27}$$

where n_k is the total number of observations in condition k. It has been shown that under the null hypothesis of identical distributions under the K conditions, χ_z^2 is chi-square distributed with $v_z = K(Q-1)$ d.f.

The above test can be applied to examine both parent and child working zone change. For parent set $\Pi = \{Y_1, \ldots, Y_M\}$, we create a compound discrete random variable whose levels are permutation of the individual parent variable levels. We denote the parent working zone test statistic by $\chi_z^{2,\pi}$ with p-value p_z^{π}. For child X, we call the child working zone test statistic $\chi_z^{2,x}$ with p-value p_z^{x}.

3.3 A discrete noise model and its estimation

We define a probabilistic noise model that changes a true ordinal discrete value to a noised version. The model applies to discrete ordinal variables. The probability of observing a perturbed value decreases as it deviates further

way from the true value. This model is designed to capture both variation in the original continuous variable and quantization errors.

Let $x \in [0, Q-1]$ be the true value and $x' \in [0, Q-1]$ be a noisy observation. Let θ be a noise level between 0 and 1. Specifically, we define the noise model by the conditional probability

$$P(x'|x,\theta) = \begin{cases} \theta, & x' \neq x \\ 1-\theta, & x' = x \end{cases} \quad x, x' \in \{0,1\}, \quad \text{or} \begin{cases} \left(1 - \frac{|x'-x|}{\sum_{u=0}^{2} |u-x|}\right)\theta, & x' \neq x \\ 1-\theta, & x' = x \end{cases} \quad x, x' \in \{0,1,2\} \tag{S28}$$

where $0 \leq \theta \leq 1$ denotes the noise level.

For the purpose of comparative analysis of binary variable, the worst-case noise level θ is 0.5, not 1, as $\theta = 1$ leads to preservation of all original interactions and the only nominal change is that the truth table values become complementary to the original. In the yeast models all but one variable (Cdc2_Cdc13 in the fission yeast networks) are binary. The worst-case noise for a three-level variable is about 0.7.

Given independent n noisy observations $x_1, \ldots, x_n$ of a discrete random variable, we estimate both the true value and the noise level using maximum likelihood estimation. Let $\hat{x}$ be the estimated value of x. Given $\hat{x} = q$, we can prove by maximum likelihood principle that

$$\hat{\theta}(q) = 1 - \frac{|\{i \mid x_i = q, 1 \leq i \leq n\}|}{n} \tag{S29}$$

To obtain the best estimate of the truth value $\hat{x}$, we enumerate all possible values $q \in [0, Q-1]$ to obtain the corresponding $\hat{\theta}(q)$, then select best $\hat{x}$ by

$$\hat{x} = \max_{0 \leq q < Q} L(q, \hat{\theta}(q) \mid x_1, \ldots, x_n) = \max_{0 \leq q < Q} \prod_{i=1}^{n} P(x_i | q, \hat{\theta}(q)) \tag{S30}$$

and the optimal estimate of θ is thus $\hat{\theta}(\hat{x})$.

We estimate the noise level for each gene using replicates under the same condition to make sure the true value is consistent. Then we estimate noise levels for all genes. The histogram of noise levels estimated for the fruit fly data is shown in **Supplementary Fig. S3**.

3.4 Benchmarking on differential gene networks regulating yeast cell cycle

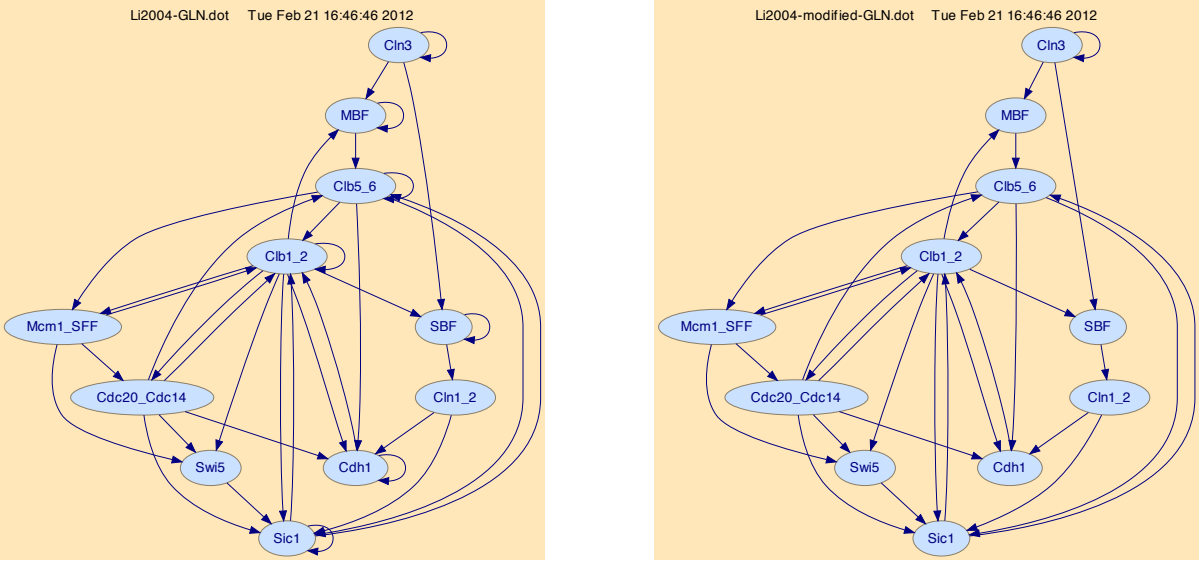
To benchmark the performance of CPχ^2 method for detecting conserved and differential gene interactions, we used two pairs of generalized logical network (GLN) models for yeast cell cycle published in the literature, against the reconstruct-then-compare method.

Groundtruth differential yeast cell cycle network models

The first pair of GLN models (Li et al., 2004; Faure and Thieffry, 2009) is for the budding yeast cell cycle. Their network topologies are shown in **Supplementary Fig. S6**. The generalized logical rules in each model are given in **Supplementary Fig. S7 and S8**.

The second pair of GLN models (Davidich and Bornholdt, 2008; Faure and Thieffry, 2009) is for the fission yeast cell cycle. Their network topologies are shown in **Supplementary Fig. S9**. The generalized logical rules in each model are given in **Supplementary Fig. S10 and S11**. This first model (Davidich and Bornholdt, 2008) has been drastically simplified from the original published model, by eliminating fictitious genes. Although this model is simplified, it offers a contrast with the second fission yeast model and still provides a good example for performance evaluation of our algorithms.

The four GLN models are obtained from the GINSim (Gonzalez et al., 2006) website (`http://gin.univ-mrs.fr`), with modifications as described in the figure captions.



Supplementary Figure S6. Two budding yeast cell cycle networks. The left one is based on (Li et al., 2004) and the right one based on (Faure and Thieffry, 2009). Both are different from the originally published models in the self-cycle at Cln3, which was added in our models to indicate that Cln3 will maintain its initial value during the dynamical progression of the system.

The 1st budding yeast cell cycle model (Li et al., 2004)

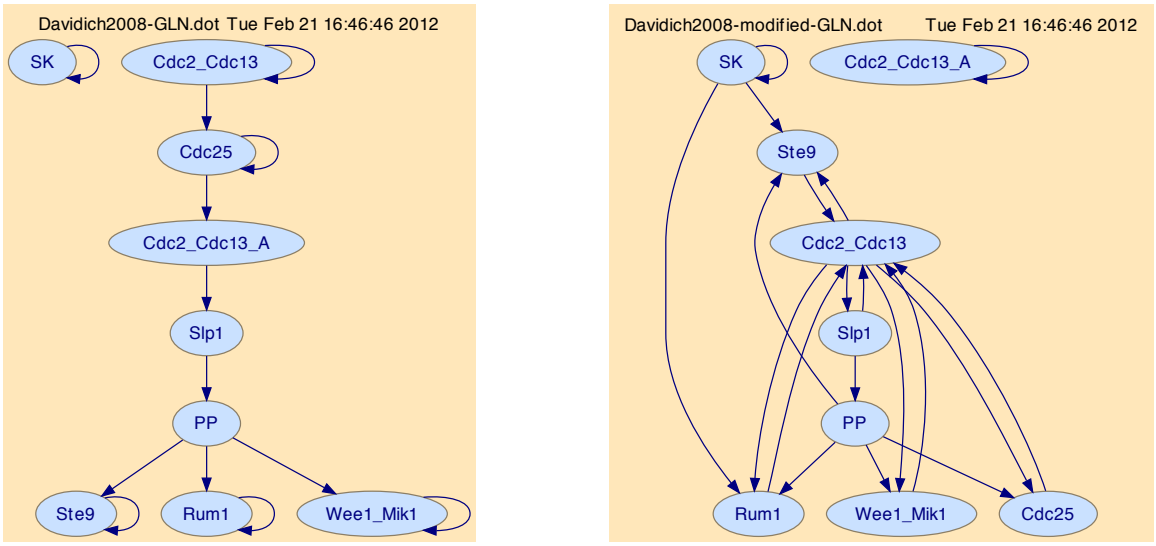
```
Cln3    <- Cln3
MBF     <- ((Cln3 | MBF) & !Clb1_2) | (Cln3 & MBF & Clb1_2))
SBF     <- ((Cln3 | SBF) & !Clb1_2) | (Cln3 & SBF & Clb1_2))
Cln1_2 <- SBF
Cdh1    <- ((Cdc20_Cdc14 | Cdh1) & !(Cln1_2 | Clb1_2 | Clb5_6))
          | (Cdc20_Cdc14 & Cdh1 & !((Cln1_2 & Clb1_2) | (Cln1_2 & Clb5_6)
                                    | (Clb1_2 & Clb5_6)))
Swi5    <- ( (Mcm1_SFF | Cdc20_Cdc14) & !Clb1_2 )
          | (Mcm1_SFF & Cdc20_Cdc14 & Clb1_2)
Cdc20_Cdc14 <- Clb1_2 | Mcm1_SFF
Clb5_6 <- ( (MBF | Clb5_6) & !(Sic1 | Cdc20_Cdc14) )
          | ( MBF & Clb5_6 & !(Sic1 & Cdc20_Cdc14) )
Sic1    <- ((Swi5 | Cdc20_Cdc14 | Sic1) & !(Cln1_2 | Clb1_2 | Clb5_6) )
          | ( ((Swi5 & Cdc20_Cdc14) | ((Swi5 | Cdc20_Cdc14) & Sic1))
            & !((Cln1_2 & Clb1_2) | (Cln1_2 & Clb5_6) | (Clb1_2 & Clb5_6)))
          | (Swi5 & Cdc20_Cdc14 & Sic1 & !(Cln1_2 & Clb1_2 & Clb5_6) )
Clb1_2 <- ((Clb1_2 | Clb5_6 | Mcm1_SFF) & !(Cdc20_Cdc14 | Cdh1 | Sic1))
         | (((Clb1_2 & (Clb5_6 | Mcm1_SFF)) | (Clb5_6 & Mcm1_SFF))
          & !((Cdc20_Cdc14 & Cdh1) | (Cdc20_Cdc14 & Sic1) | (Cdh1 & Sic1)))
        | (Clb1_2 & Clb5_6 & Mcm1_SFF & !(Sic1 & Cdh1 & Cdc20_Cdc14))
Mcm1_SFF <- Clb1_2 | Clb5_6
```

Supplementary Figure S7. The logical rules for the first budding yeast cell cycle model. They are based on (Li et al., 2004). The operator "&" stands for logical and, "|" for logical or, and "!" for logical negation. The operator "<-" stands for assignment.

The 2nd budding yeast cell cycle model (Faure and Thieffry, 2009)

```
Cln3         <- Cln3
MBF          <- Cln3 & !Clb1_2
SBF          <- Cln3 & !Clb1_2
Cln1_2       <- SBF
Cdh1         <- !(Cln1_2 | Clb5_6 | Clb1_2) | Cdc20_Cdc14
Swi5         <- ( (Mcm1_SFF | Cdc20_Cdc14) & !Clb1_2 )
                | (Mcm1_SFF & Cdc20_Cdc14 & Clb1_2)
Cdc20_Cdc14 <- Clb1_2 | Mcm1_SFF
Clb5_6       <- (MBF | !Sic1) & !Cdc20_Cdc14
Sic1         <- (Swi5 & Cdc20_Cdc14) | !(Cln1_2 | Clb5_6 | Clb1_2)
Clb1_2       <- !(Cdh1 | Sic1 | Cdc20_Cdc14) | (Clb5_6 & Mcm1_SFF)
Mcm1_SFF     <- Clb1_2 | Clb5_6
```

Supplementary Figure S8. The logical rules for the second budding yeast cell cycle model. The rules are based on (Faure and Thieffry, 2009).



Supplementary Figure S9. Two fission yeast cell cycle networks. The left one is a simplified version of (Davidich and Bornholdt, 2008) and the right one is a modified version by Faure and Thieffry (2009). In the simplified model from (Davidich and Bornholdt, 2008), fictitious edges were removed based on the logical rules given in (Faure and Thieffry, 2009). A "Start" node in the originally published versions was also removed. In the modified model, a self-repeating Cdc2_Cdc13_A node was added to make the two models contain exactly the same set of nodes.

The 1st fission yeast cell cycle model (Davidich and Bornholdt, 2008)

```
SK           <- SK
Cdc2_Cdc13   <- Cdc2_Cdc13
Ste9         <- PP | Ste9
Rum1         <- PP | Rum1
Slp1         <- Cdc2_Cdc13_A
Cdc2_Cdc13_A <- Cdc25
Wee1_Mik1    <- PP | Wee1_Mik1
Cdc25        <- Cdc25 | Cdc2_Cdc13
PP           <- Slp1
```

Supplementary Figure S10. The logical rules for the first fission yeast cell cycle model. The rules are simplified from (Davidich and Bornholdt, 2008).

The 2nd fission yeast cell cycle model (Faure and Thieffry, 2009)

```
SK           <- SK
Cdc2_Cdc13   <- 1, if(!Ste9 & !Slp1 & !Rum1 & (!Cdc25 | Wee1_Mik1))
                2, else if(Cdc25 & !Wee1_Mik1 & !Ste9 & !Rum1 & !Slp1)
                0, otherwise
Ste9         <- !(SK | Cdc2_Cdc13) | PP
Rum1         <- !(SK | Cdc2_Cdc13) | PP
Slp1         <- Cdc2_Cdc13 > 0   ## Cdc2_Cdc13 has 3 levels
Cdc2_Cdc13_A <- Cdc2_Cdc13_A
Wee1_Mik1    <- PP | !Cdc2_Cdc13
Cdc25        <- ((Cdc2_Cdc13==1) & !PP) | (Cdc2_Cdc13==2)
PP           <- Slp1
```

Supplementary Figure S11. The logical rules for the second fission yeast cell cycle model. Based on (Faure and Thieffry, 2009).

Data generation by model simulation

Using the four models, we simulated noisy data to cover all state transitions. For each model, trajectories were obtained by exhaustive simulation in our GLN modeling software. During exhaustive simulation, each unvisited state of the model is used as the initial state and the simulation run stops until an attractor cycle is detected. This process is iterated until all states have been visited.

The first budding yeast cell cycle model (Li et al., 2004) has 1803 trajectories each of 2-8 time points. The second budding yeast cell cycle model (Faure and Thieffry, 2009) has 1965 trajectories each of 2-13 time points. The first fission yeast cell cycle model (Davidich and Bornholdt, 2008) has 496 trajectories each of 2-7 time points. The second fission yeast cell cycle model (Faure and Thieffry, 2009) has 718 trajectories each of 2-8 time points.

Each yeast gene network, also called a circuit, represents dynamic interactions among 9 or 10 cell cycle genes. The gold standard, or groundtruth, for the simulation studies is conserved and differential interactions in the two network pairs. Each pair differs in either wiring or control logic, but may still share some identical interactions. Altogether there are 20 pairs of interactions: 13 differential and 7 conserved in the two network pairs. For example, interactions at Swi5 are conserved in the first pair of budding yeast networks because Swi5 shares the same logical rule; interactions at Cdc25 are differential in the pair of fission yeast networks because Cdc25 uses different rules across the two networks. These interactions, as the gold standard, are checked against those obtained from either method in performance evaluation.

The input to comparative analysis is time courses or trajectories obtained from running each model using the logical rules. For each circuit, we generated various numbers of simulated and noisy trajectories, each lasting 2-13 time points, to cover all states of the networks. Then we added different levels of noise to each trajectory using the noise model defined in Eq. (S28).

Setup of three comparative analysis methods

On each pair of trajectory collections from two networks, we ran reconstruct-then-compare, differential correlation, and CPχ^2. The common setup for the three methods is as follows: a maximum of 6 parents per child was allowed; self-regulation was enabled; a Markovian order of one was specified. The decision-making parameter was the α-level (or maximum false positive rate) for both CPχ^2 and reconstruct-then-compare. The parent selection is made based on maximum statistical significance of the chi-square statistics. Key parameters are summarized in **Supplementary Table S3**.

Besides the common setup defined above, the specific setup to each method are summarized as follows:

Reconstruct-then-compare (RTC). Here, each generalized truth table is reconstructed based on the p-value of chi-squares of the contingency table under each condition; then truth-tables are compared to determine conserved or differential interactions. Given two generalized truth tables T_1, T_2, parent sets π_1, π_2, p-values p_1, p_2, and

Supplementary Table S3. Key parameters for comparative analysis of yeast cell cycle models. They are used by all three methods – reconstruct-then-compare, differential correlation, and CPχ^2.

Parameter	Max # parents	Markovian order	Different topology	Self-cycle
Value	6	1	Allowed	Allowed

type I error α, the decision on the comparative interaction type t is made by

$$t = \begin{cases} \text{Differential} & \text{if } (\pi_1 \neq \pi_2 \text{ or } T_1 \neq T_2) \text{ and } (p_1 \text{ or } p_2 \leq \alpha), \\ \text{Conserved} & \text{otherwise,} \end{cases} \tag{S31}$$

where the condition for differential implies truth tables must be different and at least one of them significant.

Differential correlation. This method uses differential correlation to evaluate heterogeneity across contingency tables under multiple conditions. Parents of a child are selected based on p-value of chi-squares of each condition. The option used in the program is `-Y BY_EACH_COND`, the same as the setup for CPχ^2 (next). The smallest superset of parents across all K conditions are used to compute correlation coefficient r_k with the child under condition k, such that these correlation coefficients are comparable. The differential correlation across K conditions is defined by

$$dc = \frac{1}{2K} \sum_{k=1}^{K} |r_k - \bar{r}| \tag{S32}$$

where $\bar{r}$ is the average correlation coefficients over the K conditions. When $K = 2$, $dc = |r_1 - r_2|/2$.

Given Δ_0, parent sets π_1, π_2, the decision on comparative interaction type t is made by

$$t = \begin{cases} \text{Differential} & \text{if } dc \geq \Delta_0, \\ \text{Conserved} & \text{otherwise.} \end{cases} \tag{S33}$$

Comparative chi-square analysis (CPχ^2). Parents of a child are selected based on p-value of chi-squares of each condition. The option used in the program is `-Y BY_EACH_COND`. Instead of comparing truth tables as in reconstruct-then-compare, contingency tables are obtained to measure heterogeneity and homogeneity by the p-values p_d and p_c. Given p_d, parent sets π_1, π_2, and type I error α, the decision on comparative interaction type t is made by

$$t = \begin{cases} \text{Differential} & \text{if } p_d \leq \alpha, \\ \text{Conserved} & \text{otherwise.} \end{cases} \tag{S34}$$

Performance evaluation

We describe three metrics to evaluate the performance of a comparative analysis method. Let t denote the interaction type and π_1, π_2 are parent sets for an interaction under two conditions, respectively. For a given interaction at a node, the first input (t, π_1, π_2) is the detection result from a comparative analysis method, and the second input is (t^*, π_1^*, π_2^*) from the ground truth obtained directly from the original models.

The benchmarking metric is defined as

True positive if $t = t^* =$ Differential, $\pi_1 \subset \pi_1^*, \pi_2 \subset \pi_2^*$,

False positive if $t =$ Differential, $t^* =$ Conserved,

True negative if $t = t^* =$ Conserved,

False negative if ($t =$ Conserved, $t^* =$ Differential) or ($t = t^* =$ Differential, ($\pi_1 \not\subset \pi_1^*$ or $\pi_2 \not\subset \pi_2^*$)).

3.5 *Drosophila* microarray data set

In our *Drosophila* microarray studies, 10 pupal wings from animals expressing E2F1 and its co-factor DP in the dorsal wing, Cabut in the dorsal wing, or controls were dissected at time points 24 and 36 h APF. Gene overexpression was driven specifically in the dorsal wing using transgenic flies containing a wing specific Gal4 transcriptional activator (*Apterous*-Gal4) and Gal4 responsive E2F1 and Cabut transgenes (Brand and Perrimon, 1993). RNA was isolated using standard techniques (Trizol) (Buttitta et al., 2010), and cDNA synthesis was performed with one subsequent round of T7-dependent linear RNA amplification using the commercially available Message AmpTM kit from Ambion as described in (Reeves and Posakony, 2005). Amplified RNA was labeled in a subsequent cDNA synthesis reaction according to Nimblegen protocols and hybridized to Nimblegen 4-plex 60-mer *Drosophila* expression arrays (www.nimblegen.com). Hybridizations were repeated 4 times with independently obtained biological replicates to ensure maximal confidence in data reproducibility.

NimbleScan software was used to scan the arrays and for quantile normalization (all arrays were normalized together). Gene calls were generated using the Robust Multichip Average (RMA) algorithm. Statistically significant changes in gene expression were determined using ANOVA (adjusted $p < 0.05$).

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002; Barrett et al., 2011) and are accessible through GEO Series accession number GSE30484 (`http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30484`).

References

Albertson, R., Cao, J., Hsieh, T.-s., and Sullivan, W. (2008). Vesicles and actin are targeted to the cleavage furrow via furrow microtubules and the central spindle. *Journal of Cell Biology*, 181(5):777–790. 6

Barrett, T., Troup, D. B., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., and et al. (2011). NCBI GEO: archive for functional genomics data sets–10 years on. *Nucleic Acids Research*, 39(Database issue):D1005–D1010. 19

Basto, R., Lau, J., Vinogradova, T., Gardiol, A., Woods, C. G., Khodjakov, A., and Raff, J. W. (2006). Flies without centrioles. *Cell*, 125(7):1375–1386. 7, 8

Bergmann, A., Yang, A. Y.-P., and Srivastava, M. (2003). Regulators of IAP function: coming to grips with the grim reaper. *Current Opinion in Cell Biology*, 15(6):717–724. 7, 8

Bhattacharya, A. and Steward, R. (2002). The *Drosophila* homolog of NTF-2, the nuclear transport factor-2, is essential for immune response. *EMBO Reports*, 3(4):378–383. 6

Boulton, S. J., Brook, A., Staehling-Hampton, K., Heitzler, P., and Dyson, N. (2000). A role for Ebi in neuronal cell cycle control. *The EMBO Journal*, 19(20):5376–5386. 5

Brand, A. H. and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development*, 118:401–415. 19

Brandt, A., Krohne, G., and Großhans, J. (2008). The farnesylated nuclear proteins KUGELKERN and LAMIN B promote aging-like phenotypes in *Drosophila* flies. *Aging Cell*, 7(4):541–551. 8

Brumby, A. M., Zraly, C. B., Horsfield, J. A., Secombe, J., Saint, R., Dingwall, A. K., and Richardson, H. (2002). *Drosophila* cyclin E interacts with components of the brahma complex. *The EMBO Journal*, 21(13):3377–3389. 6

Buttitta, L., Katzaroff, A. J., and Edgar, B. A. (2010). A robust cell cycle control mechanism limits E2F-induced proliferation of terminally differentiated cells *in vivo*. *J Cell Biology*, 189(6):981–996. 5, 19

Chatr-Aryamontri, A., Breitkreutz, B.-J., Heinicke, S., Boucher, L., Winter, A., Stark, C., Nixon, J., Ramage, L., Kolas, N., Lara, O., Reguly, T., Breitkreutz, A., Sellam, A., Chen, D., Chang, C., Rust, J., Livstone, M., Oughtred, R., Dolinski, K., and Tyers, M. (2013). The BioGRID interaction database: 2013 update. *Nucleic Acids Research*, 41(Database issue):D816–D823. 5, 7

Davidich, M. I. and Bornholdt, S. (2008). Boolean network model predicts cell cycle sequence of fission yeast. *PLoS ONE*, 3(2):8. 14, 16, 17

De Nooij, J. C., Letendre, M. A., and Hariharan, I. K. (1996). A cyclin-dependent kinase inhibitor, dacapo, is necessary for timely exit from the cell cycle during *Drosophila* embryogenesis. *Cell*, 87(7):1237–1247. 6

Dobbelaere, J., Josué, F., Suijkerbuijk, S., Baum, B., Tapon, N., and Raff, J. (2008). A genome-wide RNAi screen to dissect centriole duplication and centrosome maturation in drosophila. *PLoS Biology*, 6(9):e224. 7, 8

Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, 30(1):207–210. 19

Faure, A. and Thieffry, D. (2009). Logical modelling of cell cycle control in eukaryotes: a comparative study. *Mol. BioSyst.*, 5:1569–1581. 14, 15, 16, 17

Frolov, M., Stevaux, O., Moon, N., Dimova, D., Kwon, E., Morris, E., and Dyson, N. (2003). G1 cyclin-dependent kinases are insufficient to reverse dE2F2-mediated repression. *Genes & development*, 17(6):723–728. 5, 6

Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., Ooi, C. E., Godwin, B., Vitols, E., and et al. (2003). A protein interaction map of *Drosophila melanogaster*. *Science*, 302(5651):1727–1736. 5, 6

Giunta, K., Jang, J., Manheim, E., Subramanian, G., and Kim, M. (2002). subito encodes a kinesin-like protein required for meiotic spindle pole formation in Drosophila melanogaster. *Genetics*, 160(4):1489–1501. 5, 6

Gonzalez, A. G., Naldi, A., Sánchez, L., Thieffry, D., and Chaouiya, C. (2006). Ginsim: a software suite for the qualitative modelling, simulation and analysis of regulatory networks. *Bio Systems*, 84(2):91–100. 14

Goshima, G., Wollman, R., Goodwin, S. S., Zhang, N., Scholey, J. M., Vale, R. D., and Stuurman, N. (2007). Genes required for mitotic spindle assembly in Drosophila S2 cells. *Science*, 316(5823):417–421. 7, 8

Guruharsha, K. G., Rual, J.-F., Zhai, B., Mintseris, J., Vaidya, P., Vaidya, N., Beekman, C., Wong, C., Rhee, D. Y., Cenaj, O., and et al. (2011). A protein complex network of drosophila melanogaster. *Cell*, 147(3):690–703. 5, 6

Habara, Y., Urushiyama, S., Shibuya, T., Ohshima, Y., and Tani, T. (2001). Mutation in the prp12+ gene encoding a homolog of SAP130/SF3b130 causes differential inhibition of pre-mRNA splicing and arrest of cell-cycle progression in *Schizosaccharomyces pombe*. *RNA*, 7(5):671–681. 6

Hampoelz, B., Azou-Gros, Y., Fabre, R., Markova, O., Puech, P.-H., and Lecuit, T. (2011). Microtubule-induced nuclear envelope fluctuations control chromatin dynamics in *Drosophila* embryos. *Development*, 138(16):3377–3386. 8

Hayashi, S. and Yamaguchi, M. (1999). Kinase-independent activity of Cdc2/Cyclin A prevents the S phase in the Drosophila cell cycle. *Genes to Cells*, 4:111–122. 5, 6

Hochheimer, A., Zhou, S., Zheng, S., Holmes, M., and Tjian, R. (2002). TRF2 associates with DREF and directs promoter-selective gene expression in Drosophila. *Nature*, 420:439–445. 5, 6

Iliopoulos, I., Török, I., and Mechler, B. (1997). The DnaJ60 gene of Drosophila melanogaster encodes a new member of the DnaJ family of proteins. *Biological Chemistry*, 378(10):1177–1181. 7, 8

Irwin, J. G. (1949). A note on the subdivision of chi-square into components. *Biometrika*, 36(1/2):130–134. 8, 9

Kalab, P., Pu, R. T., and Dasso, M. (1999). The ran GTPase regulates mitotic spindle assembly. *Current Biology*, 9(9):481–484. 8

Kondo, S. and Perrimon, N. (2011). A genome-wide RNAi screen identifies core components of the G2-M DNA damage checkpoint. *Science Signaling*, 4(154):rs1. 7, 8

Korenjak, M., Taylor-Harding, B., Binné, U. K., Satterlee, J. S., Stevaux, O., Aasland, R., White-Cooper, H., Dyson, N., and Brehm, A. (2004). Native E2F/RBF complexes contain Myb-interacting proteins and repress transcription of developmentally controlled E2F target genes. *Cell*, 119(2):181–193. 5

Kusano, A., Staber, C., Chan, H. Y. E., and Ganetzky, B. (2003). Closing the (Ran)GAP on segregation distortion in Drosophila. *Bioessays*, 25(2):108–115. 7

Lancaster, H. (1949). The derivation and partition of χ^2 in certain discrete distributions. *Biometrika*, 36(1-2):117–129. 8, 9

Lane, M. E., Sauer, K., Wallace, K., Jan, Y. N., Lehner, C. F., and Vaessin, H. (1996). Dacapo, a cyclin-dependent kinase inhibitor, stops cell proliferation during *Drosophila* development. *Cell*, 87(7):1225–1235. 6

Lehner, C. F. and O'Farrell, P. H. (1989). Expression and function of *Drosophila* cyclin A during embryonic cell cycle progression. *Cell*, 56(6):957–968. 6

Li, F., Long, T., Lu, Y., Ouyang, Q., and Tang, C. (2004). The yeast cell-cycle network is robustly designed. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14):4781–4786. 14, 15, 17

Lim, Y.-M., Yamasaki, Y., and Tsuda, L. (2013). Ebi alleviates excessive growth signaling through multiple epigenetic functions in *Drosophila*. *Genes to Cells*. Aug 6. doi:10.1111/gtc.12088. [Epub ahead of print]. 5, 6

Mathai, A. M. and Provost, S. B. (1992). *Quadratic forms in random variables: theory and applications*. M. Dekker New York. 10

Mitsis, P. G., Kowalczykowski, S. C., and Lehman, I. R. (1993). Single-stranded DNA-binding protein from Drosophila melanogaster: Characterization of the heterotrimeric protein and its interaction with single-stranded dna. *Biochemistry*, 32(19):5257–5266. 7, 8

O'Keefe, D. D., Thomas, S. R., Bolin, K., Griggs, E., Edgar, B. A., and Buttitta, L. A. (2012). Combinatorial control of temporal gene expression in the *Drosophila* wing by enhancers and core promoters. *BMC Genomics*, 13(1):498. 6

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes – The Art of Scientific Computing*. Cambridge, third edition. 13

Qi, D., Bergman, M., Aihara, H., Nibu, Y., and Mannervik, M. (2008). *Drosophila* Ebi mediates snail-dependent transcriptional repression through HDAC3-induced histone deacetylation. *The EMBO Journal*, 27(6):898–909. 5

Ravi, D., Wiles, A. M., Bhavani, S., Ruan, J., Leder, P., and Bishop, A. J. (2009). A network of conserved damage survival pathways revealed by a genomic RNAi screen. *PLoS Genetics*, 5(6):e1000527. 8

Read, R. D., Fenton, T. R., Gomez, G. G., Wykosky, J., Vandenberg, S. R., Babic, I., Iwanami, A., Yang, H., Cavenee, W. K., Mischel, P. S., et al. (2013). A kinome-wide RNAi screen in *Drosophila Glia* reveals that the RIO kinases mediate cell proliferation and survival through TORC2-Akt signaling in glioblastoma. *PLoS Genetics*, 9(2):e1003253. 5

Reeves, N. and Posakony, J. W. (2005). Genetic programs activated by proneural proteins in the developing *Drosophila* PNS. *Dev Cell*, 8:413–25. 19

Sekelsky, J. J., Brodsky, M. H., and Burtis, K. C. (2000). DNA repair in drosophila insights from the drosophila genome sequence. *Journal of Cell Biology*, 150(2):F31–F36. 7, 8

Sheskin, D. J. (2011). *Handbook of Parametric and Nonparametric Statistical Procedures*. Chapman and Hall/CRC, 5th edition. 8, 12

Staehling-Hampton, K., Ciampa, P. J., A, B., and Dyson, N. (1999). A genetic screen for modifiers of E2F in Drosophila melanogaster. *Genetics*, 153:275–287. 5, 6

Stanyon, C. A., Liu, G., Mangiola, B. A., Patel, N., Giot, L., Kuang, B., Zhang, H., Zhong, J., and Finley, R. L. (2004). A Drosophila protein-interaction map centered on cell-cycle regulators. *Genome Biology*, 5(12):R96. 5, 6, 7

Steel, R. G., Torrie, J. H., and Dickey, D. A. (1997). *Principles and Procedures of Statistics: a Biometrical Approach.* McGraw-Hill New York, 3rd edition. 8

Taylor-Harding, B., Binné, U. K., Korenjak, M., Brehm, A., and Dyson, N. J. (2004). p55, the Drosophila ortholog of RbAp46/RbAp48, is required for the repression of dE2F2/RBF-regulated genes. *Molecular and Cellular Biology*, 24(20):9124–9136. 5, 6

Tusher, V. G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *PNAS*, 98:5116–5121.

Uttenweiler-Joseph, S., Moniatte, M., Lagueux, M., Van Dorsselaer, A., Hoffmann, J. A., and Bulet, P. (1998). Differential display of peptides induced during the immune response of drosophila: a matrix-assisted laser desorption ionization time-of-flight mass spectrometry study. *Proceedings of the National Academy of Sciences*, 95(19):11342–11347. 7, 8

Weng, L., Zhu, C., Xu, J., and Du, W. (2003). Critical role of active repression by E2F and Rb proteins in endoreplication during Drosophila development. *the The European Molecular Biology Organization Journal*, 22(15):3865–3875. 5

Zar, J. H. (2010). *Biostatistical Analysis*. Prentice Hall, New Jersey, 5th edition. 8, 12