RESPONSE LETTER

-- Editor-1 – Abstract too long --

Reviewer Comment	The abstract is too long and should be ~100 words-so I suggest you move the marked section until later.
Author Response	We changed the abstract according to the editor's comments on the marked manuscript.
Excerpt From Revised Manuscript	[Abstract] From merging a number of data sources, we created an extensive map of the transcriptional regulatory network in yeast, comprising 7419 interactions connecting 180 transcription factors (TFs) with their target genes. We integrated this network with gene-expression data, relating the expression profiles of TFs and target genes. We found that genes targeted by the same TF tend to be co-expressed, with the degree of co-expression increasing if genes share more than one TF. Moreover, shared targets of a TF tend to have similar cellular functions. In contrast, the expression relationships between the TFs and their targets are much more complicated, often exhibiting time-shifted or inverted behavior.

-- Editor-2 – Reference style --

Reviewer Comment	The reference articles need titles
Author Response	We changed the reference style accordingly.
Excerpt From Revised Manuscript	[Reference] ***1*** Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. [see comments.]. *Science* 270, 467-70 ***2*** Chee, M. et al. (1996) Accessing genetic information with high-density DNA arrays. *Science* 274, 610-4 …

-- Editor-3 – Other microarray datasets --

Reviewer Comment	Referee 1 makes the point you have only looked at the cell-cycle. Would it be an idea to extend the analysis to one other condition?
Author Response	1. The referee and the editor are right in that some transcription factors only activate (or repress) their targets in response to specific conditions (sporulation, heat-shock, etc). In this sense, we should extend the analysis to other conditions. However, we have shown that the regulators do not have significant enrichment of correlated relationships with their targets; instead they tend to have delayed relationships. In order to detect these more subtle temporal relationships, it requires that (1) the expression dataset should be a time-series dataset with a considerable number of time points; (2) the interval between time points should be uniform and relatively small for a good resolution. These are also the

requirements to use the local clustering method. Unfortunately, most of the microarray datasets under these conditions are not time-series datasets or have big intervals and few time points. Therefore, we cannot analyze the relationships between the regulators and their targets within these datasets. We added this explanation in the text.

2. We did extend our analysis of co-regulated genes to other microarray datasets. We used the Pearson correlation coefficient method to determine the correlated relationships between co-regulated target genes within these datasets. We found that, although the specific LOD values are different, the general trend remains the same as our previous results from the cell cycle dataset, i.e. (1) in general, co-regulated genes tend to be co-expressed; (2) co-expression is much better when genes are co-regulated by more than one regulators. We added these new results in the text.

Excerpt From Revised Manuscript

[Page 3]

1.2 Gene expression dataset

We obtained expression profiles of yeast genes through two complete cell cycles.[11] Between the expression profiles of pairs of genes, we used a local clustering method to calculate four types of temporal relationships as diagramed in Fig. 1b[12]: correlated, time-shifted, inverted, and inverted time-shifted. To find these relationships, expression levels must be assessed over a time-course, with many measurements, at small and uniform intervals. Most available datasets do not satisfy these conditions, being only suitable for simple correlation calculations (*ie* co-expression); thus, we can only conduct detailed analysis on the cell-cycle dataset. Nevertheless, similar overall results are observed in other microarray datasets.

[Page 5]

Similar results are observed for other expression datasets[3,13-17] (Table 2).

[Table 1]

Table 1. Summary of transcription regulatory network dataset.

	motifs†	SIM	MIM	FFL	ALL
	# TFs	119	118	97	188
	# targets	1754	986	511	3416
# TF-target pairs	Total	1754	2781	1523	7419
	Activation‡	37	50	19 - 33§	144
	Repression‡	12	34	23 - 10§	79
LOD values for co-expressed target pairs£	Stress response	0.44*	3.55*	0.59	0.88*
	Sporulation	0.03	0.25	0.08	-0.05
	Diauxic shift	0.11*	1.78*	0.30*	0.30*
	DNA damage	1.24*	4.87*	1.26*	2.14*
	Cell Cycle (Spellman et al.)	0.37*	2.09*	1.62*	0.52*
	" " (Cho et al)	0.29*	2.79*	1.35*	0.93*
	" " (Zhu et al)	0.22*	2.50*	0.91*	0.64*

* LOD values with P-value smaller than 1e-05 (see supplementary Table 1)

† The abbreviation for the motifs is the same as in the caption of Figure 1A. ALL, All the TF-target pairs. There are 3 smaller motifs: Auto, 22 targets, MCL, 31 targets, RC, 119 targets. The random expectation for the number of targets is 6130, the number of yeast genes. The random expectation for the number of gene pairs in yeast is 18785385 = 6130(6129)/2, which is obtained by counting all pairs between yeast genes.

‡ Positive expression relationships (correlated and time-shifted) are considered as activation signals, while negative relationships (inverted and inverted time-shifted) are considered as repression signals. Overall, 18 regulators activate some of their targets but repress others. Note this is distinct from the number of activator relations determined experimentally (as described in §2.4 and §3.2)

§ We show the number of relations for FFL:TF1 and FFL: TF2.
£ Log odds ratios for target gene pairs having correlated profiles in different expression datasets. The local clustering method cannot be applied, so expression correlation is measured using the Pearson correlation coefficient. Co-expressed gene pairs are those in the top 1% of largest correlation coefficients.

[Supplementary Table 1]
Supplementary Table 1. P-values* for the LOD values in Table 1

Motif†	Stress response	Sporulation	Diauxic shift	DNA damage	Cell-cycle by Spellman et al	Cell-cycle by Cho et al	Cell-cycle by Zhu et al
SIM	2.50E-06	0.2958	4.88E-06	1.33E-11	2.28E-11	1.29E-11	2.28E-09
FFL	0.0097	0.2829	5.71E-07	5.81E-07	0	3.95E-13	0
MIM	0	0.1351	0	9.78E-13	3.73E-13	1.48E-12	3.22E-15
ALL	4.67E-11	0.9877	0	1.16E-10	5.96E-10	8.88E-10	0
Correlation coefficient Cut-off‡	0.70	0.95	0.90	0.80	0.70	0.70	0.70

* P-values are calculated by the formula given in text.
† The abbreviation for the motifs is the same as in the caption of Figure 1.
‡ Correlation coefficient cut-off is determined as the Pearson correlation coefficient, above which roughly top 1% gene pairs with the largest correlation coefficient are. The correlation coefficient cut-offs are equivalent to local clustering score of 13.

-- Editor-4 – Measurement of expression correlation --

Reviewer Comment	What is expression correlation exactly and how is it measured? Is it the correlation over time drawn from expression of gene a vs gene b?
Author Response	1. Yes, expression correlation is the correlation over time drawn from expression of gene a vs gene b. We added the explanation in the caption of Figure 1. 2. Please refer to response Editor-5.
Excerpt From Revised Manuscript	[Figure 1 caption] The local clustering method uses a dynamic programming algorithm to align the expression profiles of the genes in question. From the alignment, the method is able to determine which of the four types the relationship is and assign a clustering score measuring the significance of the relationship; for the Cho et al dataset, a score of 13 or above corresponds to a relationship significant to $p = 2.7\times10^{-3}$ (see supplementary materials).

-- Editor-5 – Explanation of expression relationships--

Reviewer Comment	I think the time courses (which you label "Supplementary material") need to be shown a new fig 1. (You will then need to re-label the figures). You need to make clearer whether all the panels show what you later call "correlation ", so do time shifted and inverted also fall into the class of "correlated"? Or are you just talking about "simultaneous" as correlated? We also need some details as to the statistical tests that are used to determine correlation (of the three types).
Author Response	1. We changed the old "supplementary material" to be the new Figure 1, as the editor suggested. 2. The "correlated" relationship is the "simultaneous"

	relationship shown in Figure 1B.1. And, only the class of "correlated" is considered as "correlation" or "co-expression". In order to prevent the confusion, we removed the term "simultaneous relationship" and named Figure 1B.1 as "correlated relationship (i.e. co-expression)". We also labeled Figure 1B.1 as "correlated (i.e. co-expression)". Moreover, we modified the text to make it clear that we only consider "correlated" relationship as "co-expression". 3. In order to make it clearer what each relationship means biologically, we showed real expression profiles of transcription factors and their targets in three examples in Figure 3. Different expression relationships between these genes within each example are discussed in detail in the manuscript. 4. The local clustering method uses a dynamic programming algorithm to align the expression profiles of the genes in question. From the alignment, the method is able to determine which type (of the four types) the relationship is and assign a clustering score as a statistical measurement of the alignment. We added the calculation details of the local clustering method in the Figure 1 caption. We also adapted the description of the method from our previous JMB paper and added it in the "Supplementary materials". The reader should be able to repeat the calculations to determine the expression relationships after she/he reads the paper and the supplementary materials.
Excerpt From Revised Manuscript	[Page 3] **1.2 Gene expression dataset** We obtained expression profiles of yeast genes through two complete cell cycles.[11] Between the expression profiles of pairs of genes, we used a local clustering method to calculate four types of temporal relationships as diagramed in Fig. 1b[12]: correlated, time-shifted, inverted, and inverted time-shifted. To find these relationships, expression levels must be assessed over a time-course, with many measurements, at small and uniform intervals. Most available datasets do not satisfy these conditions, being only suitable for simple correlation calculations (*ie* co-expression); thus, we can only conduct detailed analysis on the cell-cycle dataset. Nevertheless, similar overall results are observed in other microarray datasets. [Figure 1 caption] **Figure 1. Schematic representations of transcription regulatory motifs and temporal gene expression relationships. (A)** Depiction of the six basic regulatory motifs: ○ TF, □ target. (1) single input motif - target gene has one TF, (2) multi-input motif – target gene has multiple TFs, (3) feed-forward loop – leading TF (TF1) regulates an intermediate TF (TF2) and both regulate the target gene, (4) autoregulation – TF targets itself, (5) multi-component loop – two TFs regulate each other, and (6) regulator chain – set of TFs regulate each other in series. **(B)** Schematic of the four gene expression relationships: (1) correlated (*ie* co-expressed - genes have similar profiles), (2) time-shifted (genes have similar profiles, but one is delayed with respect to the other in the cell cycle), (3) inverted (genes have opposing profiles), and (4) inverted time-shifted. The local clustering method uses a dynamic programming algorithm to align the expression profiles of the genes in question. From the alignment, the method is able to determine which of the four types the relationship is and assign a clustering score measuring the significance of the relationship; for the Cho et al dataset, a score of 13 or above corresponds to a relationship significant to $p = 2.7\times10^{-3}$ (see supplementary materials).

[Page 39-10]

4. Examples of TF-target relationships

In Fig. 3 we examine specific regulatory networks.

4.1 SIM: ndd1 network

Ndd1, a cell cycle regulator during S and G_2/M transition[22,23], acts as the sole regulating TF for *MCM21,* kinetochore protein required for normal cell growth from late S to early M phase[24,25], and *STB5*, another transcription factor[26]. All three genes display cell cycle periodicity. *NDD1* peaks early in S and sustains high expression until G2. The targets are co-expressed and time-shifted with respect to *NDD1* by one time-point, peaking later in S.

4.2 MIM: forkhead network

Ndd1 is recruited to G_2/M-transition-specific promoters by Fkh1 and Fkh2, two forkhead transcription activators[22,23,27]. Collectively, these three TFs regulate Dbf2, a kinase needed for cell-cycle regulation[28], and HDR1 (function unknown). The expression profiles of the three TFs are only loosely correlated and peak at different points from early S to late G_2. The targets are time-shifted with respect to *FKH1* by two time-points and peak at the G_2/M transition. The local clustering scores show that their expression profiles are better correlated than in the preceding SIM example (Supplementary Table 3).

4.3 FFL: mbp1/swi4 network

In a feed-forward-loop, Mbp1 (a cell-cycle regulator controlling DNA replication and repair[6,29]) is the leading TF, Swi4 (a cell-cycle regulator controlling cell-wall and membrane synthesis[6,29]) is the intermediate TF, and *SPT21* (a TF involved in histone expression[30]) and *YML102C-A* (function unknown) are the target genes. The profiles of the intermediate TF and target genes are correlated and peak sharply in G_1. In contrast, the leading TF displays an inverted relationship, which highlights its involvement as a target repressor. (Previous studies have shown Mbp1 acts as an activator for ~50% its targets during the G_1/S transition and as a repressor for ~10% of its targets later in the cycle[6,7,29].)

[Figure 3 caption]
Figure 3. Expression profiles of example regulatory networks during the cell cycle. ○ TF, □ target. |→| indicates a time-shift relationship. The inset describes the TF and target genes involved in the example. **(A)** Single input motif, **(B)** multi-input motif, and **(C)** feed-forward loop.

[Supplementary materials]
Determination of the expression relationships using local clustering method (excerpt from Jiang et al, JMB, 314:1053-1066)
"We use a degenerate dynamical programming algorithm to find time-shifted and inverted correlations between expression profiles. The algorithm does not allow gaps between consecutive time points in the current version. However, there are some obvious extensions, which we explore later in the discussion section.
…
Supplementary Figure 2. "Three examples showing simultaneous (A), time-delayed (B), and inverted (C) relationships in the expression profiles. Note there are only 8 time points for each profile, while in the real yeast cell-cycle data there are 17 time points. Also, the expression ratio is not normalized, whereas in the real data each profile is normalized so that the averaged expression ratio is 0 and the standard deviation is 1. The thick segments of the expression profiles are the matched part. (D) The corresponding matrix **E** for the expression profile shown in (A). The corresponding matrix **D** is not shown because in this case the match score (the maximal score) is from **E** and not **D**. The numbers outside the border of the matrix are the expression ratio shown in (A). The black cell contains the overall match score *S* for these two expression profiles, and the light gray cells indicate the path of the optimal alignment between the expression profiles. The path starts from the match score and ends at the first encountered 0. (E) The corresponding matrix **E** for the expression profile shown in (B). Note the time-shifted relationship and how the length of the overall alignment can be shorter than 8 positions. (F) The corresponding matrix **D** for the expression profiles shown in (C). The matrix **E** is not shown because the best match score is not from this matrix in this case."

-- Editor-6 – Explanation of FFL --

Reviewer Comment	You need to expand on what a "feed forward loop" is. Is it where transcription factor R1 regulates R2 and then both R1 and R2 bind us of the gene of interest?
Author Response	Yes, a "feed forward loop" is where R1 regulates R2 and both of them together regulate the target. 1. In order to make it clear what each network motif is, we added Figure 1A to show a schematic diagram for each motif. We also explained what they are in detail in the figure caption. 2. For the three main motifs (SIM, MIM, and FFL) that are the focus of the manuscript, we also showed three real biological examples in Figure 3 to further illustrate their topological and biological meanings. 3. For each example, we added a diagram as an inset to show what the motif is and why the example falls into this category.
Excerpt From Revised Manuscript	[Page 3] **1.1 TF-target regulatory network** We compiled a yeast regulation dataset from merging the results of genetic, biochemical and ChIp-chip experiments[4,5,7,10]. It contains 7,419 TF-target pairs from 180 TFs and 3,474 target genes (Table 1). Regulatory networks can be simplified into six basic motifs (Fig. 1a)[9,10]. Here, we focus on the single input motif (SIM), multi-input motif (MIM) and feed-forward loop (FFL) as the data for the remaining motifs are too sparse. [Figure 1 caption] **Figure 1. Schematic representations of transcription regulatory motifs and temporal gene expression relationships. (A)** Depiction of the six basic regulatory motifs: ○ TF, □ target. (1) single input motif - target gene has one TF, (2) multi-input motif – target gene has multiple TFs, (3) feed-forward loop – leading TF (TF1) regulates an intermediate TF (TF2) and both regulate the target gene, (4) autoregulation – TF targets itself, (5) multi-component loop – two TFs regulate each other, and (6) regulator chain – set of TFs regulate each other in series. **(B)** Schematic of the four gene expression relationships: (1) correlated (*ie* co-expressed - genes have similar profiles), (2) time-shifted (genes have similar profiles, but one is delayed with respect to the other in the cell cycle), (3) inverted (genes have opposing profiles), and (4) inverted time-shifted. The local clustering method uses a dynamic programming algorithm to align the expression profiles of the genes in question. From the alignment, the method is able to determine which of the four types the relationship is and assign a clustering score measuring the significance of the relationship; for the Cho et al dataset, a score of 13 or above corresponds to a relationship significant to $p = 2.7\times10^{-3}$ (see supplementary materials). [Figure 3 caption] **Figure 3. Expression profiles of example regulatory networks during the cell cycle.** ○ TF, □ target. |→| indicates a time-shift relationship. The inset describes the TF and target genes involved in the example. **(A)** Single input motif, **(B)** multi-input motif, and **(C)** feed-forward loop.

-- Editor-7 – Question on expression level --

Reviewer Comment	P3. you talk about co-expressed. Does it depend on how highly expressed-where the "co" is likely to be more reliable-rather than transcripts which just go up a bit?
Author Response	The editor is right in that when the gene's expression level is too low, the measurements become relatively noisy and therefore not

	suitable for the analysis. In our analysis, we did not select highly-expressed genes explicitly. But, we believe that genes with very low expression levels are excluded from our analysis implicitly during our calculation and our conclusions are not affected by the gene's expression level. There are four main reasons: 1. All the microarray experiments (both Affymetrix and cDNA microarray) have their own detection mechanisms. The suspicious spots will be flagged and excluded from the final results. Furthermore, using the same set of the Affymetrix chips, Wodicka et al (Nature Biotechnology, 15:1359-1367) found that 20% yeast genes are expressed in a very low level (< 0.1 copy per cell). In our analysis, however, 95% genes that have significant relationships with other genes are highly expressed (> 0.1 copy per cell). This confirms that our analysis is based on the true biological signal, rather than the noise of the measurements. 2. It is well known that different genes have quite different expression levels. When a transcription factor activates its targets in the cell, the expression of these targets will all increase, but to different individual levels. Therefore, it is the relative shape, not the absolute level, of the expression profile that is more interesting and more related to our analysis. 3. Using the local clustering method, the clustering score between a pair of genes are essentially the sum of the products of their normalized expression levels at each time point. Therefore, only highly-expressed genes with significant changes in the expression levels could have a local clustering score higher than the cut-off (13) used in the analysis. 4. Using Mann-Whitney U-test (i.e. non-parametric T-test), the average expression level of the genes included in the analysis is indeed much higher (P-value < 1e-16) than others.

-- Editor-8 – Explanation of LOD --

Reviewer Comment	P4. The log odds ratio needs explaining. Is it the log of the ratio of co-expression of two genes connected by having (say) MIM-to well what? A pair of genes chosen at random? The average co-expression of possible pairs of genes? And what is the category "All" in the x axis of (new) fig 2? In panel 1a "All" seems to be about 2, this means that something is 100 times (if using $\log_{10}$) as likely as what to occur by what (chance?). How come no enrichment has a LOD~1 (2^{nd} par) surely if the odds ratio is 1 then LOD is 0?
Author Response	1. In general, the LOD values are calculated by the formula:

	$$\mathrm{LOD} = \ln[\frac{P(relationship \mid regulation)}{P(relationship)}]$$ where *P(relationship | regulation)* is the possibility for gene pairs with certain regulatory relationships to have a specific expression or functional relationship. *P(relationship)* is the possibility for gene pairs randomly chosen from the dataset to have the corresponding expression or functional relationship. Because the LOD values are the log of the odds ratios, positive LOD values mean that the possibility is higher that random expectation, and vice versa. Detailed description of how to calculate each individual LOD value is given in the supplementary materials. So, in Figure 2A, the LOD value of bar "MIM" is the log of the ratio of co-expression of two genes co-regulated by the same MIM to that of "a pair of genes chosen at random" (the editor's first guess is correct). 2. The category "All" includes all gene pairs co-regulated by at least one common transcription factor. For instance, if gene A is regulated by regulators 1 & 2 and gene B is regulated by regulators 2 & 3, the gene pair "A-B" won't be included in the category "MIM", because they belong to two different MIMs. However, the "A-B" pair is included in the category "All", because they share one common regulator – regulator 2. 3. We defined the LOD value here to be the natural logarithm of the odds ratio. So, in Figure 2A, "All" is 1.3, which means that the probability for the gene pairs sharing at least one common regulator to be co-expressed is about four ($e^{1.3} \sim 4$) times higher than that of the gene pairs chosen at random. 4. The editor is absolutely right. No enrichment should have a LOD value of 0. The probability for gene pairs co-regulated by the same FFLs to have the same function is still significantly higher than gene pairs chosen at random. However, compared with other categories (especially "SIM"), the functional enrichment of FFL is the lowest. This is what we meant by saying "FFL motifs do not display much functional enrichment (LOD ~ 1)." We carefully revised this sentence to remove any ambiguity.
Excerpt From Revised Manuscript	[Page 3-4] **1.3 Statistical formalism** We use several statistics to quantify the significance of our observations. The *p*-value is the probability that an observation (*eg* co-expression of target genes) would be made by chance, and is calculated using the cumulative binomial distribution: $$\mathrm{P}(c \geq c_o) = \sum_{c=c_o}^{N} [\frac{N!}{N!(N-c)!}] p^c (1-p)^{N-c}$$ *N* is the total number of possible gene pairs in the data, c_o is the number of observed pairs with a specific relationship (*ie* from expression or function), and *p* is the probability of finding a gene pair with the same relationship randomly (picking from the entire genome). The log odds ratio (LOD) is the enrichment a particular relationship in the presence of regulation

	with respect to random expectation for the occurrence of the relationship: $$\mathrm{LOD} = \ln[\frac{P(relationship \mid regulation)}{P(relationship)}]$$ *P(relationship | regulation)* is the probability for gene pairs with certain regulatory relationship (*eg* TF=>target) to have a specific expression or functional relationship (*eg* correlated expression). *P(relationship)* is the probability for randomly selected gene pairs to have the same expression or functional relationship. When we report this together with p-values, we use the following notation {log p-value,LOD value}. [Page 4] Interestingly, FFL motifs display the smallest enrichment {-11,1.5}. [Supplementary materials] **Calculation of the LOD values** ***Figure 2A*** $$\mathrm{LOD} = \ln[\frac{P(co\text{-}exp \mid co\text{-}reg)}{P(co\text{-}exp)}]$$ where *P(co-exp | co-reg)* is the possibility for genes co-regulated by a certain motif to be co-expressed (i.e. correlated), which is calculated as the percentage of correlated pairs between all possible pairs of co-regulated genes. *P(co-exp)* is the possibility for gene pairs randomly chosen from the dataset to be co-expressed, which is calculated as the percentage of correlated pairs between all possible gene pairs in Cho's dataset. ***Figure 2B*** $$\mathrm{LOD} = \ln[\frac{P(same-function \mid co-reg)}{P(same-function)}]$$ where *P(same-function | co-reg)* is the possibility for gene pairs co-regulated by a certain motifs to have the same functions. *P(same-function)* is the possibility for gene pairs randomly chosen from the dataset to have the same functions. … ***Table 1*** $$\mathrm{LOD} = \ln[\frac{P(co\text{-}exp \mid co\text{-}reg)}{P(co\text{-}exp)}]$$ where all the calculations are very similar to those in Figure 2A, except that the expression relationships between gene pairs are determined using Pearson correlation coefficient in different microarray datasets. All the possibilities in the analysis are calculated in the same way as in Figure 2A.

-- Editor-9 – Explanation of FFL --

Reviewer Comment	You need to expand on what a "feed forward loop" is. Is it where transcription factor R1 regulates R2 and then both R1 and R2 bind us of the gene of interest?
Author Response	Please refer to response Editor-6.

-- Ref1-1 – Other microarray datasets --

Reviewer Comment	The authors used only expression data for the yeast cell cycle. I see no reason why it should be so. For example, it is well known that some regulators in yeast are

	activated during sporulation, under stressful conditions, etc … Therefore, using only one expression dataset reduce the possibility to find positive expression correlation between regulator and their target genes.
Author Response	Please refer to response Editor-3.

-- Ref1-2 – Explanation of local clustering method --

Reviewer Comment	The authors should better explain the biological implication of concepts like co-expression, inverted and time shifted profiles. After all, general readers might be unfamiliar with their JMB paper on the subject.
Author Response	1. The goal of this paper is to discover the biological meanings of these expression relationships within the regulatory networks. After careful analysis, we were able to show that (i) co-regulated target genes tend to be co-expressed; (ii) The regulators show delayed relationships with their targets, although the enrichment is not very clear. This is summarized in "conclusion" section. 2. We showed the real expression profiles of the regulators and their targets within three different motifs in Figure 3, and discussed in detail the biological implication of the expression relationships in the context of the different regulatory relationships. 3. Please refer to response Editor-5.
Excerpt From Revised Manuscript	[The changes are the same as in response Editor-5.]

-- Ref1-3 – Number of connections with different regulatory modes --

Reviewer Comment	Can the authors provide a rough estimate on the number of positive/negative regulatory connections? Does it happen that a single regulatory gene activates one gene but represses another?
Author Response	1. There are 144 positive regulatory connections; 2. There are 79 negative regulatory connections; 3. 18 regulators activate some of their targets but repress others; 4. We also surveyed the number of positive/negative regulatory connections within different motifs. The results are recorded in Table 1.
Excerpt From Revised Manuscript	[Table 1] Table 1. Summary of transcription regulatory network dataset.

	motifs†	SIM	MIM	FFL	ALL
	# TFs	119	118	97	188
	# targets	1754	986	511	3416
# TF-target pairs	Total	1754	2781	1523	7419
	Activation‡	37	50	19 - 33§	144
	Repression‡	12	34	23 - 10§	79
LOD values for co-expressed target pairs£	Stress response	0.44*	3.55*	0.59	0.88*
	Sporulation	0.03	0.25	0.08	-0.05
	Diauxic shift	0.11*	1.78*	0.30*	0.30*
	DNA damage	1.24*	4.87*	1.26*	2.14*
	Cell Cycle (Spellman et al.)	0.37*	2.09*	1.62*	0.52*
	" " (Cho et al)	0.29*	2.79*	1.35*	0.93*
	" " (Zhu et al)	0.22*	2.50*	0.91*	0.64*

* LOD values with P-value smaller than 1e-05 (see supplementary Table 1)

† The abbreviation for the motifs is the same as in the caption of Figure 1A. ALL, All the TF-target pairs. There are 3 smaller motifs: Auto, 22 targets, MCL, 31 targets, RC, 119 targets. The random expectation for the number of targets is 6130, the number of yeast genes. The random expectation for the number of gene pairs in yeast is 18785385 = 6130(6129)/2, which is obtained by counting all pairs between yeast genes.

‡ Positive expression relationships (correlated and time-shifted) are considered as activation signals, while negative relationships (inverted and inverted time-shifted) are considered as repression signals. Overall, 18 regulators activate some of their targets but repress others. Note this is distinct from the number of activator relations determined experimentally (as described in §2.4 and §3.2)

§ We show the number of relations for FFL:TF1 and FFL: TF2.

£ Log odds ratios for target gene pairs having correlated profiles in different expression datasets. The local clustering method cannot be applied, so expression correlation is measured using the Pearson correlation coefficient. Co-expressed gene pairs are those in the top 1% of largest correlation coefficients.

-- Ref1-4 – Number of AND gates --

Reviewer Comment	Would it be possible to investigate the number of AND gates (two transcription factors are needed simultaneously to activate a single target gene)?
Author Response	1. Determined by the expression data, there are 3 AND gates; 2. We cut out this section as we felt it was too complex to describe in detail.

-- Ref1-5 – Number of different types of FFLs --

Reviewer Comment	The Lee et al paper gives an estimate on the number but not the type of feedback loops. What is the fraction of positive and negative feedback loops, and why?
Author Response	1. In the Lee et al paper, they discussed in detail feedforward loops (FFL), not feedback loops. Therefore, there might be a typo in the referee's comment. 2. We investigated the number of different types of FFLs determined from the expression data and added a paragraph together with supplementary Table 2 to discuss the results; 3. We also summarized the number of different types of regulatory connections between R1's and their targets and those between R2's and their targets. Table 1b is added in to show the results;

Excerpt From Revised Manuscript	[Page 7] FFL motifs present the most interesting and complex relationships. The leading TFs in the motif (denoted TF1) generally have negative relationships with the target genes -- i.e. inverted {-2,0.82} or inverted time-shifted {-10,2.0}. The intermediate TFs (TF2) exhibit all four types of relationships; The most common arrangement (55% of FFLs, supplementary table 2) is where the leading TF has a negative relationship with the target and the intermediate TF has a positive one (*ie* correlated or time-shifted). (Note, however, there are only 11 FFLs for which both TF1 and TF2 have significant expression relationships with the targets.) [Supplementary Table 2] **Supplementary Table 2. Number of FFLs with different regulatory relationships between the regulators and their targets determined from the expression data** **Type of FFLs** / **# of FFLs** **TF1-target** / **TF2-target** **P*** / **P** / 3 **P** / **N** / 2 **N** / **P** / 6 **N** / **N** / 0 * P: positive relationships between the TFs and their targets; N: negative relationships between the TFs and their targets. [Other changes are the same as in response Ref1-3]

-- Ref1-6 – Another perspective on the poor co-expression of genes co-regulated by SIMs --

Reviewer Comment	The authors claim that co-expression is more tightly regulated when more than one transcription factors (TF) are involved. But what if the yeast genome contains other, yet unidentified TFs. In this case, single input motifs of two genes would be part of a multiple - possibly only partially overlapping – motifs. The authors should at least mention this problem.
Author Response	We agree with the referee and modified the text accordingly.
Excerpt From Revised Manuscript	[Page 5] The differences in enrichment (*ie* LOD values) indicate that expression is much more tightly regulated when multiple TFs are involved. However, with >100 yeast transcription factors yet to be investigated[18], unidentified TF-target relationships will probably alter the classification of SIM target genes to MIM or FFL networks in the future.

-- Ref2-1 – More concise title --

Reviewer Comment	The title could be made more concise as follows: 'Correlation of gene expression in regulatory network motifs'
Author Response	We've changed the title to make it more concise. The title now is: *Genome-wide analysis of gene expression relationships in transcriptional regulatory networks*
Excerpt From Revised Manuscript	[Title] **Genome-wide analysis of gene expression relationships in transcriptional regulatory networks**

-- Ref2-2 – More regulatory data --

Reviewer Comment	These network motifs were found by analysis of ChIp-chip experiments described Lee et al. (2002). It would make sense to expand the data set to all known transcriptional regulatory relationships, such as those in the TRANSFAC database.
Author Response	The referee made an excellent point here. Now, we added in not only the TRANSFAC database, but also ChIp-chip experiments by Horak et al. (2002) and transcriptional regulatory data by Guelzim et al. (2002). Therefore, our calculations have taken into account all available transcriptional regulatory data, which consist of 180 transcription factors and 3474 genes (> half of yeast genome) in total. Compared with our previous dataset (106 regulators and 2416 genes), new information is more than half. And, the results remain the same.
Excerpt From Revised Manuscript	[Abstract] From merging a number of data sources, we created an extensive map of the transcriptional regulatory network in yeast, comprising 7419 interactions connecting 180 transcription factors (TFs) with their target genes. We integrated this network with gene-expression data, relating the expression profiles of TFs and target genes. We found that genes targeted by the same TF tend to be co-expressed, with the degree of co-expression increasing if genes share more than one TF. Moreover, shared targets of a TF tend to have similar cellular functions. In contrast, the expression relationships between the TFs and their targets are much more complicated, often exhibiting time-shifted or inverted behavior. [Page 3] **1.1 TF-target regulatory network** We compiled a yeast regulation dataset from merging the results of genetic, biochemical and ChIp-chip experiments[4,5,7,10]. It contains 7,419 TF-target pairs from 180 TFs and 3,474 target genes (Table 1). Regulatory networks can be simplified into six basic motifs (Fig. 1a)[9,10]. Here, we focus on the single input motif (SIM), multi-input motif (MIM) and feed-forward loop (FFL) as the data for the remaining motifs are too sparse.

-- Ref2-3 – Multiple parallel FFLs? --

Reviewer Comment	In section 1., it says target genes part of a feed forward loop also display significant correlation. Is this referring to a situation where multiple genes are regulated by the same specific (R2) and general (R1) transcription factor, in other words multiple parallel feed forward loops?
Author Response	1. Yes, they are multiple parallel feed forward loops. 2. Please refer to response Editor-6.
Excerpt From Revised Manuscript	[The changes are the same as in response Editor-6]

-- Ref2-4 – Summary table --

Reviewer Comment	A table giving the number of SIMs, MIMs and FFLs studied, and the percentages that have the specific features described, would be useful.
Author Response	We agree with the referee and changed the manuscript accordingly.

	1. We summarized the number of all six motifs and other statistics on the yeast regulatory networks in Table 1. 2. We added the specific LOD values and P values for each feature in the main text. 3. We added the specific percentages for most features in the main text. The percentage of each feature can also be calculated easily from its LOD value.
Excerpt From Revised Manuscript	[Table 1] Table 1. Summary of transcription regulatory network dataset.

	motifs†	SIM	MIM	FFL	ALL
	# TFs	119	118	97	188
	# targets	1754	986	511	3416
# TF-target pairs	Total	1754	2781	1523	7419
	Activation‡	37	50	19 - 33§	144
	Repression‡	12	34	23 - 10§	79
LOD values for co-expressed target pairs£	Stress response	0.44*	3.55*	0.59	0.88*
	Sporulation	0.03	0.25	0.08	-0.05
	Diauxic shift	0.11*	1.78*	0.30*	0.30*
	DNA damage	1.24*	4.87*	1.26*	2.14*
	Cell Cycle (Spellman et al.)	0.37*	2.09*	1.62*	0.52*
	" " (Cho et al)	0.29*	2.79*	1.35*	0.93*
	" " (Zhu et al)	0.22*	2.50*	0.91*	0.64*

* LOD values with P-value smaller than 1e-05 (see supplementary Table 1)
† The abbreviation for the motifs is the same as in the caption of Figure 1A. ALL, All the TF-target pairs. There are 3 smaller motifs: Auto, 22 targets, MCL, 31 targets, RC, 119 targets. The random expectation for the number of targets is 6130, the number of yeast genes. The random expectation for the number of gene pairs in yeast is 18785385 = 6130(6129)/2, which is obtained by counting all pairs between yeast genes.
‡ Positive expression relationships (correlated and time-shifted) are considered as activation signals, while negative relationships (inverted and inverted time-shifted) are considered as repression signals. Overall, 18 regulators activate some of their targets but repress others. Note this is distinct from the number of activator relations determined experimentally (as described in §2.4 and §3.2)
§ We show the number of relations for FFL:TF1 and FFL: TF2.
£ Log odds ratios for target gene pairs having correlated profiles in different expression datasets. The local clustering method cannot be applied, so expression correlation is measured using the Pearson correlation coefficient. Co-expressed gene pairs are those in the top 1% of largest correlation coefficients.

[Page 5]
First, we investigate expression relationships between genes targeted by the same TFs. Overall, 3.3% of target gene pairs are co-expressed, which is four times greater than random expectation {-12,1.3} (Fig. 2a, bar-ALL). We detect few inverted or time-shifted relationships (§2.4).

The level of correlation is very dependent on the type of regulatory network motif (Fig. 2a). Genes targeted by individual TFs (SIM) are not strongly correlated: just 1.3% of target pairs are co-expressed though this is significantly higher than expected {-11,0.29}. Correlation is stronger for genes targeted by multiple, common TFs: 24.4% of MIM target pairs {-12,3.2} and 5.0% of FFL targets exhibit co-expression {-12,1.6}. Similar results are observed for other expression datasets[3,13-17] (Table 1).

-- Ref2-5.1 – Explanation of AND-gate --

Reviewer Comment	In section 3., the behaviour of feed forward loops has to be clarified. First, the authors must explain what an AND-gate is.
Author	We cut out this section as we felt it was too complex to describe

Response	in detail.

-- Ref2-5.2 – Inhibition vs. Activation --

Reviewer Comment	Next, they should specify the different combinations of inhibition and activation that occur in the (mostly coherent?) FFLs.
Author Response	1. We added in Table 1 to summarize the number of different combinations of inhibition and activation that occur in all three motifs we analyzed; 2. Furthermore, we analyzed the expression relationships between the known activators (or repressors) and their targets. These known activation and inhibition regulations are determined by genetic or biochemical methods. We found that (1) Co-expression of co-activated gene pairs is much better than that of co-repressed gene pairs; (2) The probability of co-activated gene pairs having negative expression relationships is lower than random expectation; (3) Activators tend to have correlated relationships with their targets; while inhibitors tend to have inverted relationships with their targets.
Excerpt From Revised Manuscript	[Table 1] Table 1. Summary of transcription regulatory network dataset.

	motifs†	SIM	MIM	FFL	ALL
	# TFs	119	118	97	188
	# targets	1754	986	511	3416
# TF-target pairs	Total	1754	2781	1523	7419
	Activation‡	37	50	19 - 33§	144
	Repression‡	12	34	23 - 10§	79
LOD values for co-expressed target pairs£	Stress response	0.44*	3.55*	0.59	0.88*
	Sporulation	0.03	0.25	0.08	-0.05
	Diauxic shift	0.11*	1.78*	0.30*	0.30*
	DNA damage	1.24*	4.87*	1.26*	2.14*
	Cell Cycle (Spellman et al.)	0.37*	2.09*	1.62*	0.52*
	" " (Cho et al)	0.29*	2.79*	1.35*	0.93*
	" " (Zhu et al)	0.22*	2.50*	0.91*	0.64*

* LOD values with P-value smaller than 1e-05 (see supplementary Table 1)

† The abbreviation for the motifs is the same as in the caption of Figure 1A. ALL, All the TF-target pairs. There are 3 smaller motifs: Auto, 22 targets, MCL, 31 targets, RC, 119 targets. The random expectation for the number of targets is 6130, the number of yeast genes. The random expectation for the number of gene pairs in yeast is 18785385 = 6130(6129)/2, which is obtained by counting all pairs between yeast genes.

‡ Positive expression relationships (correlated and time-shifted) are considered as activation signals, while negative relationships (inverted and inverted time-shifted) are considered as repression signals. Overall, 18 regulators activate some of their targets but repress others. Note this is distinct from the number of activator relations determined experimentally (as described in §2.4 and §3.2)

§ We show the number of relations for FFL:TF1 and FFL: TF2.

£ Log odds ratios for target gene pairs having correlated profiles in different expression datasets. The local clustering method cannot be applied, so expression correlation is measured using the Pearson correlation coefficient. Co-expressed gene pairs are those in the top 1% of largest correlation coefficients.

[Page 7-8]

3.2 Relation to Regulatory-signal Type

	As in §2.4, we can measure the TF-target expression relationships when the type of regulatory signals is taken into account. Though the data is too sparse to make statistically sound conclusions, we try to make some observations. Unsurprisingly, activators are co-expressed with their targets {-2,0.63} (Fig. 2f), and comprise over 50% of TF-target pairs with significant expression relationships. We also find that repressors exhibit inverted {-2,1.1} and inverted time-shifted relationships {-2,1.2}. There are unexpected results too. Activators display significant inverted time-shifted relationships {-6,1.8} and repressors show (normal) time-shifted relationships. There are several reasons for this: A sizeable proportion of TFs (15%) act both as activators and repressors, in some cases for the same target. Furthermore, the combined effect of multiple TFs in MIM and FFL motifs can have an unpredictable effect on target expression.

-- Ref2-5.3 – Number of different types of FFLs --

Reviewer Comment	Then the number of FFLs for each combination could be given, as inferred by them from the expression data.
Author Response	Please refer to response Ref1-5.

-- Ref2-5.4 – Examples of Motifs --

Reviewer Comment	At the moment, all that is said is that there is often a direct correlation of the specific transcription factors, and an inverted time-shifted correlation with the general transcription factors for both the target and R2. If this is the most common type of FFL, then it would be nice to illustrate it with a specific example from their results.
Author Response	1. Yes, this is the most common type of FFL (i.e. R1 represses both R2 and the target. R2 activates the targets), despite the fact that the statistics are not very good; 2. A specific example (Mbp1 represses the expression of SWI4 and their targets; while the expression of SWI4 and the targets is highly correlated) is given in Figure 3C to illustrate the biological importance of our analysis;
Excerpt From Revised Manuscript	[Page 7] FFL motifs present the most interesting and complex relationships. The leading TFs in the motif (denoted TF1) generally have negative relationships with the target genes -- i.e. inverted {-2,0.82} or inverted time-shifted {-10,2.0}. The intermediate TFs (TF2) exhibit all four types of relationships; The most common arrangement (55% of FFLs, supplementary table 2) is where the leading TF has a negative relationship with the target and the intermediate TF has a positive one (*ie* correlated or time-shifted). (Note, however, there are only 11 FFLs for which both TF1 and TF2 have significant expression relationships with the targets.) [Page 9-10] ***4.3 FFL: mbp1/swi4 network*** In a feed-forward-loop, Mbp1 (a cell-cycle regulator controlling DNA replication and repair[6,29]) is the leading TF, Swi4 (a cell-cycle regulator controlling cell-wall and membrane synthesis[6,29]) is the intermediate TF, and *SPT21* (a TF involved in histone expression[30]) and *YML102C-A* (function unknown) are the target genes. The profiles of the intermediate TF and target genes are correlated and peak sharply in G_1. In contrast, the leading TF displays an inverted relationship, which highlights its involvement as a target repressor. (Previous studies have shown Mbp1 acts as an activator for ~50% its targets during the G_1/S transition and as a repressor for ~10% of its targets later in the cycle[6,7,29].) [Figure 3 caption] **Figure 3. Expression profiles of example regulatory networks during the cell cycle.** ○ TF,

	□ target. |→| indicates a time-shift relationship. The inset describes the TF and target genes involved in the example. **(A)** Single input motif, **(B)** multi-input motif, and **(C)** feed-forward loop. [Supplementary Table 2] **Supplementary Table 2. Number of FFLs with different regulatory relationships between the regulators and their targets determined from the expression data** <table><tr><th colspan="2">Type of FFLs</th><th rowspan="2"># of FFLs</th></tr><tr><th>TF1-target</th><th>TF2-target</th></tr><tr><td>P*</td><td>P</td><td>3</td></tr><tr><td>P</td><td>N</td><td>2</td></tr><tr><td>N</td><td>P</td><td>6</td></tr><tr><td>N</td><td>N</td><td>0</td></tr></table>* P: positive relationships between the TFs and their targets; N: negative relationships between the TFs and their targets.

-- Ref2-6 – Explanation of LODs --

Reviewer Comment	Figure: there should be a more complete description of how the log odds values are calculated.
Author Response	Please refer to response Editor-8.

-- Ref2-7 – Minor problems of format and language --

Reviewer Comment	Details of references and wording should be attended to more carefully, for instance the Lee et al reference in the abstract, and 'developed in our lab' in the figure legend.
Author Response	We agree with the referee and revised our manuscript very carefully. Specifically, we removed the Lee et al reference in the abstract, and 'developed in our lab' in the figure caption.
Excerpt From Revised Manuscript	[Abstract] From merging a number of data sources, we created an extensive map of the transcriptional regulatory network in yeast, comprising 7419 interactions connecting 180 transcription factors (TFs) with their target genes. We integrated this network with gene-expression data, relating the expression profiles of TFs and target genes. We found that genes targeted by the same TF tend to be co-expressed, with the degree of co-expression increasing if genes share more than one TF. Moreover, shared targets of a TF tend to have similar cellular functions. In contrast, the expression relationships between the TFs and their targets are much more complicated, often exhibiting time-shifted or inverted behavior. [Figure 1 Caption] …**(B)** Schematic of the four gene expression relationships: (1) correlated (*ie* co-expressed - genes have similar profiles), (2) time-shifted (genes have similar profiles, but one is delayed with respect to the other in the cell cycle), (3) inverted (genes have opposing profiles), and (4) inverted time-shifted. The local clustering method uses a dynamic programming algorithm to align the expression profiles of the genes in question. From the alignment, the method is able to determine which of the four types the relationship is and assign a clustering score measuring the significance of the relationship; for the Cho et al dataset, a score of 13 or above corresponds to a relationship significant to $p = 2.7\times10^{-3}$ (see supplementary materials).