Supplementary Results for Reconstructed dynamic regulatory maps

Supplementary website with data and software can be found at
URL: http://www.sb.cs.cmu.edu/drem

1 Dynamic maps of Amino Acid Starvation

The time series expression data for response to stress in Amino Acid (AA) starvation conditions for *Saccharomyces cerevisiae* was sampled relative to an unstressed control at five time points 0.5h, 1h, 2h, 4h, and 6h (Gasch et al, 2000). Genes were filtered if they had more than one missing data point or did not exhibit an absolute log base two fold change of at least one for at least one time point, leaving 2029 of the initial 6152 genes profiled remaining. The time series expression data was initially integrated with ChIP-chip data for 34 transcription factors (TFs) from AA starvation conditions in (Harbison et al, 2004), also listed in Supplementary Table 1.

For each gene, g, there was a binary input vector, $I_g = (I_{g_1}, ..., I_{g_{34}})$ with each element of the vector, I_{g_i}, corresponding to one of the 34 TFs. An element, I_{g_i}, was '1' if the corresponding transcription factor bound the promoter region of the gene at a p-value < 0.005 in the ChIP-chip experiment, and otherwise was '0'. If a p-value was missing the corresponding element was set to '0'.

A dynamic regulatory map produced as output by the method was presented

Supplementary Table 1: TFs with ChIP-chip data in AA Starvation

Adr1	Arg80	Arg81	Aro80	Bas1	Cad1
Cbf1	Cha4	Dal81	Dal82	Fhl1	Gat1
Gcn4	Gcr2	Gln3	Hap4	Hap5	Leu3
Met28	Met31	Met32	Met4	Mot3	Pho2
Put3	Rap1	Rcs1	Rph1	Rtg1	Rtg3
Sfp1	Sip4	Stp1	Uga3		

Table 1: Table of TFs with ChIP-chip data in AA starvation conditions from (Harbison et al, 2004)

in Figure 2a. Genes were assigned to the most likely path through the model (see supplementary methods). In total the model assigned 1270 of the 2029 genes (62.6%) to the higher path out of the first split (Supplementary Figure 1). Of the 2029 genes, 157 were predicted to be regulated by Gcn4 based on the ChIP-chip data, and of these 157, 127 were also assigned to the higher path (80.9%). If the assignments were random, based on the hypergeometric distribution the p-value of seeing 127 or more genes would be $< 2 \times 10^{-7}$. The ChIP-chip data was used to learn the model and assign genes to paths so this statistic does not provide an independent assessment of significance, but still be can used to score the relative importance of transcription factors at splits (lower score being more important). The transcription factor with the next most significant score along the higher path was Cbf1 with 101 of 131 genes (77.1%) assigned to it and a score of 2×10^{-4}. Among the 759 genes assigned to the path initially repressed there were over-enrichments of Fhl1 (133 of 165; score 2×10^{-32}), Sfp1 (54 of 57; score 2×10^{-20}), and Rap1 (89 of 128; score 3×10^{-14}) predicted regulated genes.

Supplementary Figure 1: Initial Bifurcation in Amino Acid Starvation

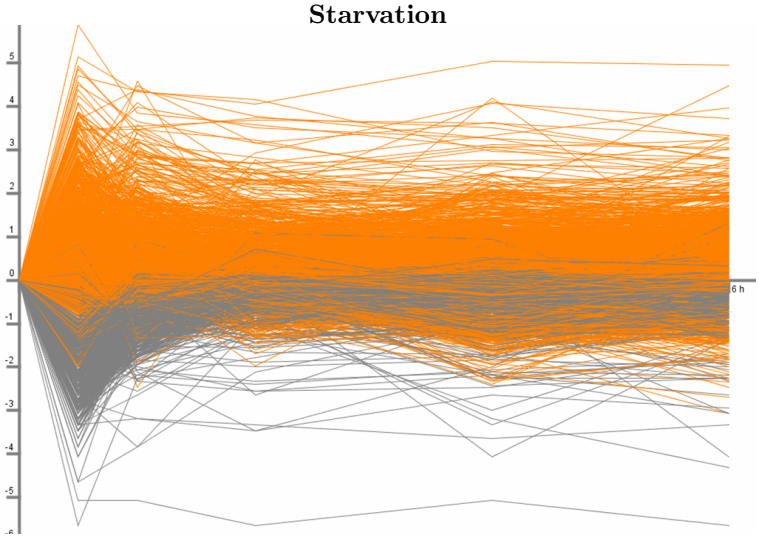

Figure 1: Plot of genes passing filter for the amino acid starvation time series. Orange genes were assigned to the higher path out of the initial split, while gray genes were assigned to the lower path.

On the split out of the higher state at the 30min time point, as with the initial split, genes assigned to the higher path tended to be bound in the ChIP-chip data by Gcn4 (split score 5×10^{-31}) or Cbf1 (2×10^{-5}). Unlike with the first split we also see an over-enrichment of genes on the higher path regulated by a number of additional transcription factors including Met32 (score 2×10^{-10}), Hap5 (6×10^{-8}), Met4 (6×10^{-7}), Dal82 (3×10^{-6}), Rtg3 (1×10^{-5}), Gln3 (6×10^{-5}), Arg81 (3×10^{-4}), and Stp1 (7×10^{-4}).

Supplementary Table 2: TFs with AA starvation ChIP-chip data appearing on the Map in Figure 2b

Arg81	Cbf1	Dal82	Fhl1	Gcn4
Gln3	Hap5	Met31	Met32	Met4
Met4	Rap1	Rtg3	Sfp1	Stp1

Table 2: Table of TFs with AA starvation ChIP-chip data appearing on the map in Figure 2b.

Supplementary Table 3: TFs without AA starvation ChIP-chip data appearing on the Map in Figure 2b

Abf1	Ino4	Mbp1
Swi4	Yap7	

Table 3: Table of TFs without AA starvation ChIP-chip data appearing on the map in Figure 2b. The activity of Ino4 in the response to AA starvation was confirmed with new ChIP-chip data.

AA starvation map based on condition specific and general binding data: The map in Figure 2b was derived with the same expression data as described above, but the 34 TFs with ChIP-chip binding data in AA conditions was augmented with an additional 75 TFs. These 75 TFs were part of the regulatory code presented in (Harbison et al, 2004). For these additional TFs, a TF was predicted to regulate the gene if the gene was bound by the transcription factor with a p-value < 0.005 in at least one condition for which a ChIP-chip experiment was performed (usually YPD media), and has a motif for the transcription factor in its promoter region that is also present in at least two of three other related yeast species (Harbison et al, 2004). In total 109 TFs were used to learn the model, 34 as before based on the AA ChIP-chip, and these additional 75. The resulting map appears in Figure 2b. Supplementary Table 2 presents those TFs appearing on the map in Figure 2b that were among the 34 TFs with AA starvation ChIP-chip data. Supplementary Table 3 presents the TFs that appeared on the map in Figure 2b, and did not have condition specific AA ChIP-chip data.

Ino4 based on the map in Figure 2b is associated with genes whose average expression level is near 0 at the one hour time point and then hold steady or rise at later time points. In this model 15 of the 15 Ino4 predicted regulated genes that passed filter and assigned to the path going through the second highest node at the one hour time point were assigned to the higher path at the two hour time point (score 9.9×10^{-4}). At the next split along this path, 13 of the 15 Ino4 genes were assigned to the higher path in the model, the brown path in Figure 2b, (split score 0.09). This path was the most associated with Ino4. Overall 422 genes of the 6152 genes on the microarray were assigned to this higher path out of two hour split. A total of 46 genes initially were predicted to

Supplementary Table 4: Ino4 Predicted Regulated Genes on Response Path

YDR497C	YER026C	YER092W	YGR196C
YHR123W	YIL119C	YJL048C	YJL141C
YJL167W	YNL169C	YNL180C	YOR316C
YOR317W			

Table 4: 13 Ino4 predicted regulated genes that appear on the main Ino4 response path. These genes are all bound in YPD media by Ino4 at a p-value < 0.005 and have a conserved motif for Ino4.

be regulated by Ino4 thus observing 13 on this path corresponds to an overall score 8×10^{-6}. These 13 Ino4 genes are listed in Supplementary Table 4 and their expression profiles are presented in Figure 3a.

Experimental results for Ino4: As discussed in the main text, Ino4's expanded regulation role in AA starvation response was experimentally verified. Table I in the main text summarizes the increase in binding of Ino4 in AA at 4hr compared to in synthetic complete + D-glucose (SCD) conditions. Supplementary Figure 2 presents experimental validation of the increased activity of Ino4 as a result of AA starvation for the second repeat (see Figure 3d for the first repeat).

Supplementary Figure 2: Validation of Ino4 Response in Amino Acid Starvation

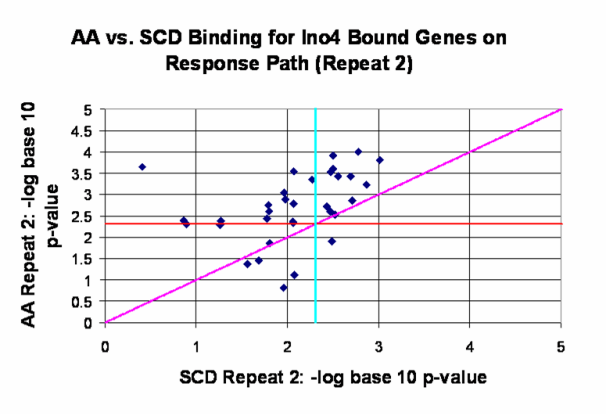

Figure 2: Comparison of binding p-values for SCD media and AA starvation for genes assigned to the main path determined by DREM to be regulated by Ino4 in the second repeat (see Figure 3d for the other repeat). The plot for the negative log base 10 of the binding p-value for genes that were bound at the 0.005 p-value significance level in one or more of the four experiments and are on the identified Ino4 response path. The horizontal and vertical lines represent a 0.005 p-value significance level. Anything to the right of the vertical line is significant in the SCD experiment. Anything above the horizontal line is significant in the AA starvation experiment. Anything above the diagonal line had a lower p-value in the AA starvation experiment.

2 Dynamic map of Hydrogen Peroxide

The expression data for the Hydrogen Peroxide (H_2O_2) stress condition was obtained from (Gasch et al, 2000). The time points used were sampled at 10m, 20m, 30m, 50m, 60m, 80m, 100m, 120m, and 160m relative to an unstressed control. Genes were filtered if they had more than one missing data point, or their expression change did not exceed an absolute value of one at any time point leaving 2170 genes. For the static input of TF-gene regulation predictions, ChIP-chip data from H_2O_2 moderate concentrations, which matched the expression data conditions, was used for 28 TFs (Supplementary Table 5). A TF was predicted to regulate a gene if it bound the genes promoter region with a p-value < 0.005. These TFs were combined with an additional 76 TFs. For these additional TFs without H_2O_2 moderate concentrations condition ChIP-chip data, a TF was predicted to regulate a gene if if it bound the gene in at least one ChIP-chip experiment (usually in YPD media) with a p-value < 0.005 and had a conserved motif in two other yeast species (Harbison et al, 2004).

Supplementary Table 5: TFs with ChIP-chip data from the Hydrogen Peroxide condition

Aft2	Cin5	Fkh2	Hap4	Hsf1
Mal33	Mbp1	Mot3	Msn2	Msn4
Nrg1	Pdr1	Pho2	Put3	Rcs1
Reb1	Rim101	Rox1	Rph1	Rpn4
Rtg3	Sfp1	Skn7	Xbp1	Yap1
Yap6	Yap7	YJL206C		

Table 5: Table of TFs with ChIP-chip data from the moderate concentration Hydrogen Peroxide condition in (Harbison et al, 2004). The moderate concentration conditions of Hydrogen Peroxide correspond to the conditions of the expression data of (Gasch et al, 2000).

In the regulatory map output for the H_2O_2 experiment (Supplementary Figure 3) we see at the 10 minute time point genes on the higher path tended to be regulated by Yap7 (score 4×10^{-17}), Yap1 (1×10^{-13}), Rpn4 (10^{-11}), Msn2 (2×10^{-10}), Gcn4 (4×10^{-56}), Skn7 (2×10^{-5}), Msn4 (2×10^{-5}), Aft2 (2×10^{-5}), and Put3 (8×10^{-5}) genes. When modeling the distribution of genes assigned to the higher state at 10m, we see their distribution at 20m can be modeled by two well separated Gaussians. Conditioned on the set of genes assigned to the path going through the higher state at 10m, we still see an enrichment of Msn2, Yap1, Msn4, Yap7, Cin5, Skn7, and Hap1 regulated genes assigned to this higher state all with a split score < 0.001. On the most repressed path which reached its minimum at 20 minutes there was 607 genes, including 107 of the 155 ribosomal genes on the microarray (p-value $< 2 \times 10^{-73}$). This same set of genes is also enriched for Fhl1 and Rap1 predicted regulated genes, 52 of 94 Fhl1 (overall score 1×10^{-28}) and 70 of 176 Rap1 genes were on this path (5×10^{-15}).

Supplementary Figure 3: Dynamic Map of Hydrogen Peroxide Response

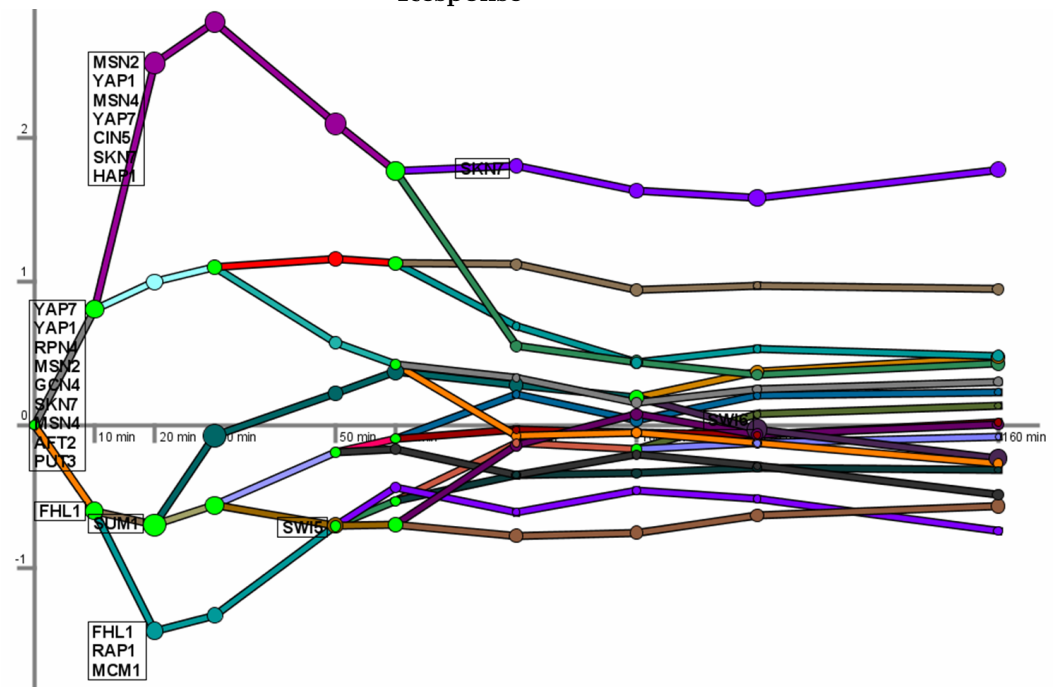

Figure 3: Temporal map derived by DREM for the Hydrogen Peroxide stress experiment of (Gasch et al, 2000). TF labels for a path out of a split appear if their split score is less than 0.001 (see supplementary methods) and are ranked order by most significant score. The top of a box of TF labels is aligned with the top of the circle representing the next state on the path out of the split. The area of a circle is proportional to the standard deviation of the Gaussian distribution of the associated state.

We also observe two interesting bifurcation events happening later in the time course. Around 60m we see a split in the expression data where many of the most highly induced genes begin to rapidly lower in expression level. Of the 91 genes assigned to the highest path 45 are assigned to the path which drops rapidly after 60m while 46 were assigned to a path for which the genes continue to have a higher expression level. We observe that 15 of 16 Skn7 predicted regulated genes (score 9×10^{-5}), 16 of 21 Msn2 regulated genes (score 0.007), and 13 of 13 genes regulated by both Skn7 and Msn2 appear on the highest path. Also at this split we observe that 8 of the 91 genes were annotated to the GO category cell homeostasis, and all 8 were assigned to the higher path (uncorrected split p-value < 0.0031, and corrected p-value < 0.04). We note that of these 8 genes, 6 of them, Sod1, Trx2, Tsa1, Isu2, YDR453C, and YOR226C are predicted to be regulated by Skn7 and of these 6 all but Sod1 are also predicted to be regulated by Msn2. Another bifurcation occurs around 100m where genes with a similar initial repression and then recovery diverge at the last two time points. Between 120m and 160m we see genes on one of these paths recover higher than their baseline level, this path represents 92 of the 307 genes assigned into the split. Of these genes we see an enrichment of genes regulated by a number of cell cycle transcription factors, Swi6 (8 of 9; score 4×10^{-4}), Fkh2 and Ndd1 (6 of 7; score 3.4×10^{-3}), Mcm1 (8 of 11; score 4×10^{-3}). Supplementary Tables 6 and 7 summarize the TFs appearing on the map in Supplementary Figure 3 based on condition specific specific ChIP-chip and without it respectively.

Supplementary Table 6: TFs with ChIP-chip data from the Hydrogen Peroxide Condition Appearing on Map

Aft2	Cin5	Msn2
Msn4	Put3	Rpn4
Skn7	Yap1	Yap7

Table 6: Table of TFs with ChIP-chip data in moderate concentration Hydrogen Peroxide appearing on the map in Supplementary Figure 3

Supplementary Table 7: TFs without ChIP-chip data from the Hydrogen Peroxide condition appearing on Map

Fhl1	Gcn4	Hap1	Mcm1
Rap1	Sum1	Swi5	Swi6

Table 7: Table of TFs without ChIP-chip data in the moderate concentration Hydrogen Peroxide condition appearing on the map in Supplementary Figure 3

3 Dynamic Map of Heat Shock

We also analyzed the stress response under heat shock conditions (Gasch et al, 2000). In particular we present results here for the Heat-1 experiment sampled at 5m, 10m, 15m, 20m, 30m, 40m, 60m, and 80m. Genes were filtered if they had more than one missing data point, or their expression change did not exceed an absolute value of one at any time point leaving 2399 genes. For the static input data we used the requirement of binding at a p-value of < 0.005 in any ChIP-chip experiment and a conserved motif in two other yeast (Harbison et al, 2004) except for Hsf1 for which we used the list of Hsf1 targets published in (Hahn et al, 2004), derived from 20 heat shock binding experiments. In the heat shock experiment map (Figure 4a) the DREM map shows that Hsf1, Msn4, Msn2, and Skn7 are the transcription factors associated with genes assigned to the higher path out of the first split, their split scores are 2×10^{-28}, 5×10^{-11}, 4×10^{-7}, and 9×10^{-6} respectively. Additionally Hap4, Rpn4, Aft2, and Sko1 are associated with the higher path out of the first split with scores in the range of 0.0001-0.001. Msn2, Msn4, and Hsf1 are also seen on the map regulating genes which were assigned to the highest expressed paths in the model. The ribosomal transcription factors Fhl1, Rap1, and Sfp1 as well as the cell cycle transcription factors, Mbp1 and Swi4, are associated with genes repressed in some of the initial splits. See the main text for a discussion of the cell-cycle response in this condition. Supplementary Figure 4 presents a comparison of the expression of cell cycle genes (Spellman et al, 1998) along the path most associated with the cell cycle with ribosome genes on the path most associated with the ribosome.

Supplementary Figure 4: Comparison of Ribosome and Cell Cycle Genes Response in Heat Shock Conditions

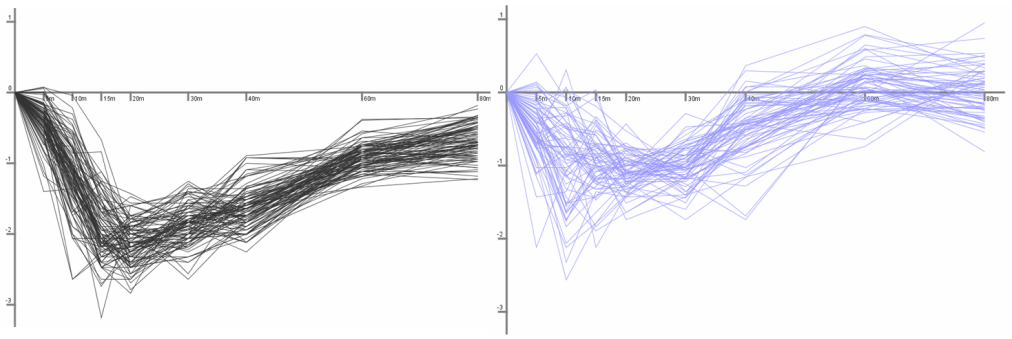

Figure 4: Comparison of expression pattern of GO ribosome genes assigned to the main ribosomal path (left) and cell cycle genes (Spellman et al, 1998) assigned to the main cell cycle path (right). The ribosome genes were repressed more steeply than cell cycle genes, and also unlike the cell cycle did not recover to their initial level.

4 Dynamic map of DTT

The next stress condition analyzed is the DTT-2 stress condition of (Gasch et al, 2000). The time series expression data was sampled at 0m, 5m, 15m, 30m, 60m, 120m, 240m, and 480m and all expression values were normalized against the 0m value. Genes were filtered if they had more than one missing data point, or their expression change did not exceed an absolute value one at any time point leaving 2457 genes. In the static input there was 102 TFs, and for a TF to be predicted to regulate a gene the TF was required to bind the promoter region of the gene with a p-value < 0.005 in a ChIP-chip experiment from at least one condition and have a conserved motif for the TF in two other yeast (Harbison et al, 2004). In this experiment we see that Fhl1, Rap1, and Sfp1 are associated with the lower path in a series of splits (Supplementary Figure 5). The path is severely repressed at 240m, after which there is a substantial recovery. The GO p-value enrichment for the category cytosolic ribosome along this path is $< 10^{-152}$. The map shows that Aft2 (score 3×10^{-4}) and Hsf1 (6×10^{-4}) are the most significant TFs associated with genes on the higher path out of a bifurcation that occurs after 30m on this activated path. This higher path is enriched for GO categories such as carbohydrate metabolism (p-value $< 8 \times 10^{-10}$) and cell wall ($< 5 \times 10^{-7}$). The map indicates another bifurcation event begins to occurs between 60m and 120m. Msn4 (score 4×10^{-7}) and Msn2 (7×10^{-4}) are most associated with the highest path out of this 60m split. This higher path has overall enrichment of GO response to stress genes of (p-value $< 4 \times 10^{-6}$) and even conditioned on the set of genes going into the split at 60min the enrichment is still significant (p-value uncorrected $< 6 \times 10^{-4}$ and corrected < 0.04).

Supplementary Figure 5: Dynamic Map of the DTT Response

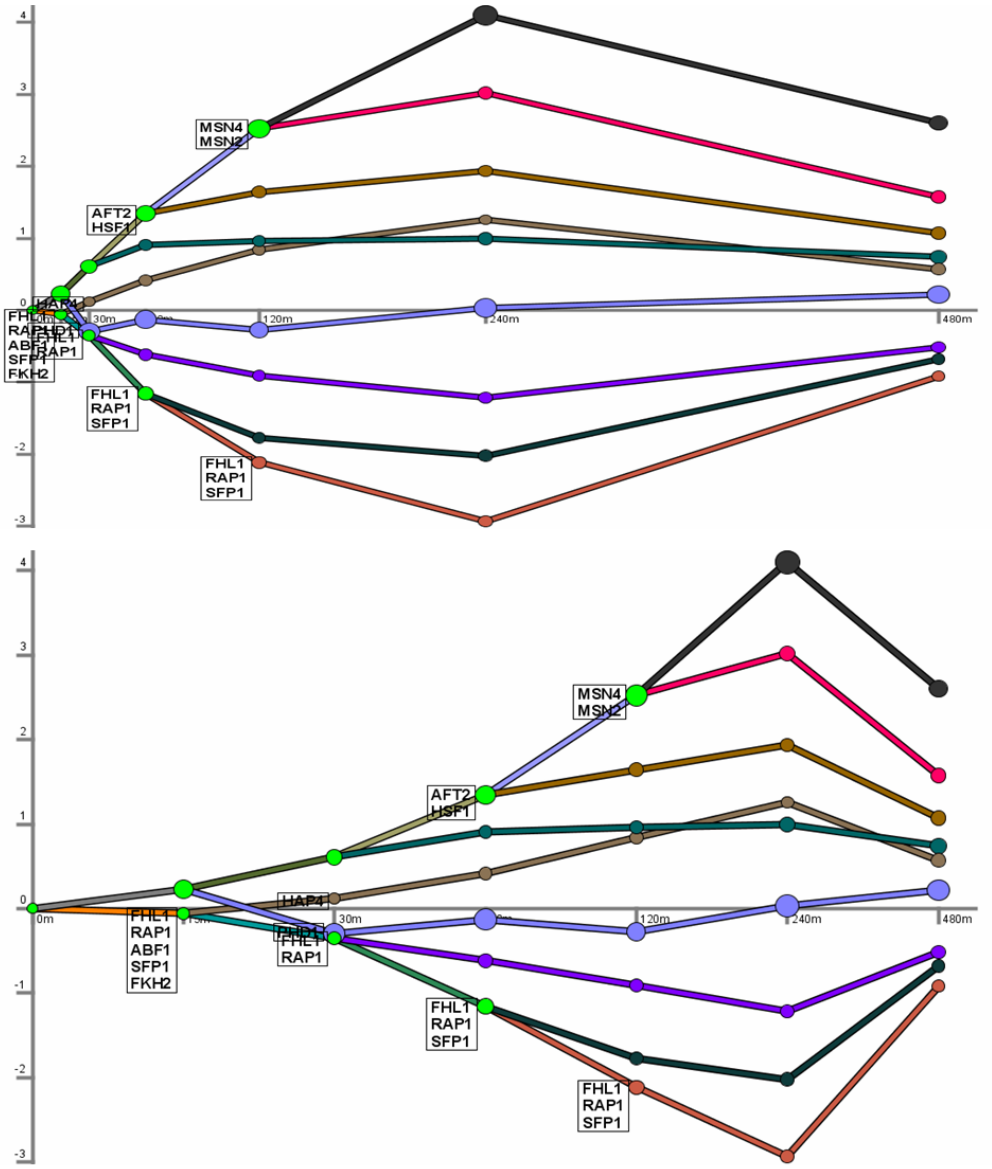

Figure 5: Temporal map derived by DREM for the DTT experiment of (Gasch et al, 2000). Top and bottom are same map except in the top the spacing between time points is proportional to the actual sampling rate while in the bottom map they are uniformly spaced. TF labels appear if their split score is less than 0.001 (see Supplementary Methods) and are ranked order by most significant score. The top of a box of TF labels is aligned with the top of the circle representing the next state on the path out of the split. The area of a circle is proportional to the standard deviation of the Gaussian distribution of the associated state.

5 Dynamic Map of Cold Shock

The dynamic map for the cold shock experiment, in which the temperature was shifted from 37°C to 25°C, can be seen in Supplementary Figure 6. The map was inferred based on the expression data of (Gasch et al, 2000) sampled at 0m, 15m, 30m, 45m, 60m, and 90m. Genes were filtered if they had more than one missing data point, or their expression change did not exceed an absolute value of one at any time point leaving 2551 genes. The static input of TF-gene regulation predictions was the same used in the DDT-2 example. In this map Fhl1, Rap1, Sfp1, and Abf1 are all associated with genes along the higher path out of the initial split. While in contrast Fhl1, Rap1, and Sfp1 appeared on the heat shock map on the down regulated paths. We also see the inverse placement of Msn4 now appearing on a down regulated path. Genes regulated by Mbp1 (split score 6×10^{-6}), Swi6 (4×10^{-4}), and Swi4 (0.0024) are enriched along the higher path of a split between 45min and 60min. This higher path reaches above baseline levels and is enriched for cell cycle genes (p-value $< 10^{-8}$) based on (Spellman et al, 1998). In this condition also between 45m and 60m DREM detected a bifurcation event along the activated path (see Supplementary Figure 7). Conditioned on the set of genes going into the split, all the genes in Figure 7, the set of gray genes is most enriched for structural constituent of ribosome (p-value $< 10^{-52}$). The pink path is most enriched for nucleobase, nucleoside, nucleotide, and nucleic acid metabolism (p-value $< 10^{-21}$).

Supplementary Figure 6: Dynamic Map of the Cold Shock Response

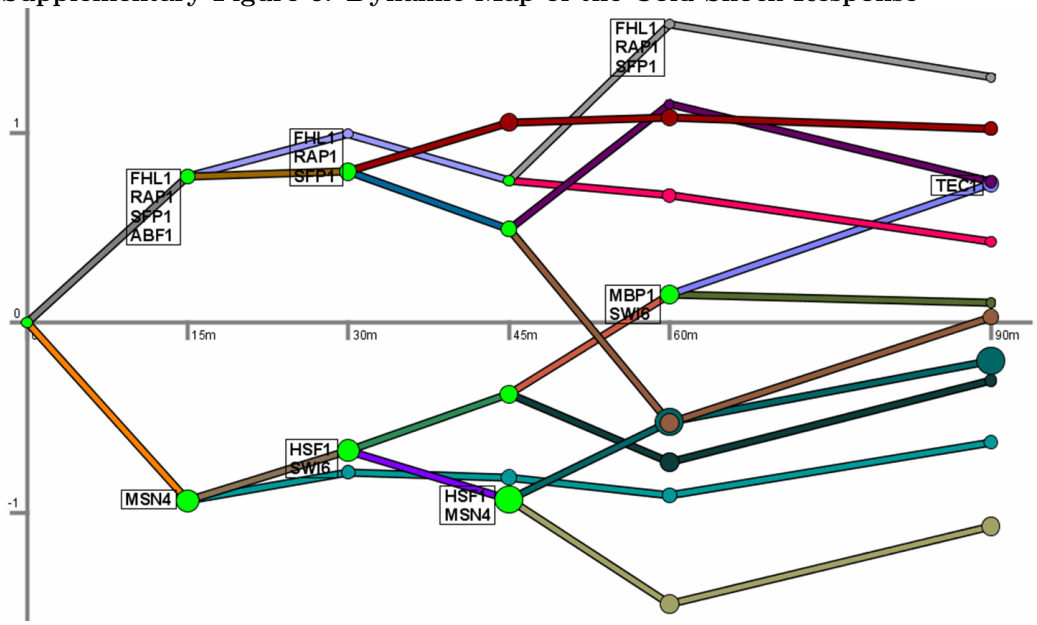

Figure 6: Temporal map derived by DREM for the cold shock experiment (Gasch et al, 2000). TF labels appear if their split score is less than 0.001 (see Supplementary Methods) and are ranked order by most significant score. The top of a box of TF labels is aligned with the top of the circle representing the next state on the path out of the split. The area of a circle is proportional to the standard deviation of the Gaussian distribution of the associated state.

Supplementary Figure 7: Bifurcation Event During the Cold Shock Response

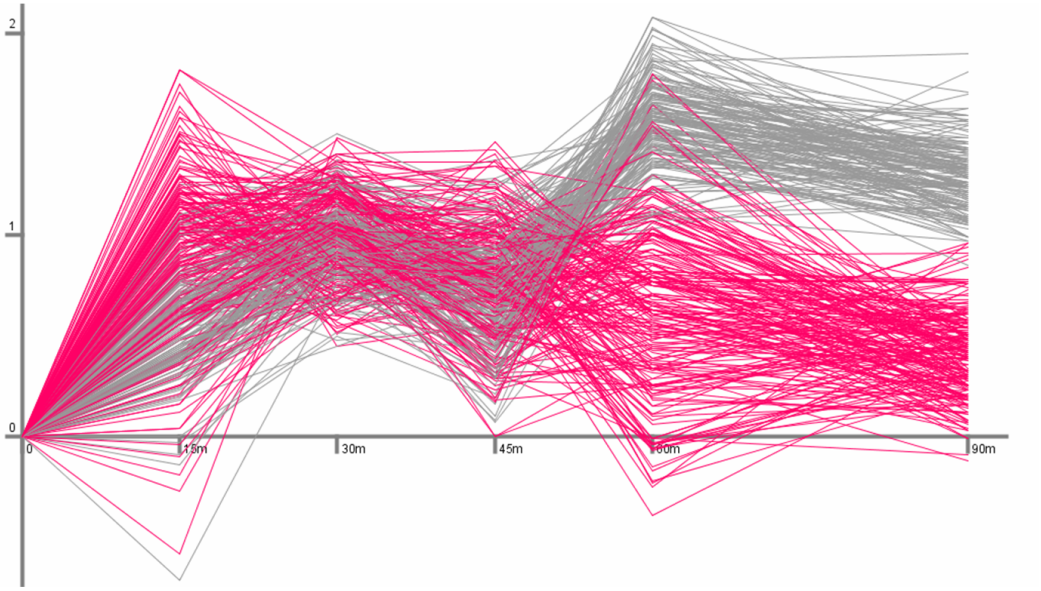

Figure 7: Bifurcation event in cold shock. Gray genes are enriched for ribosome genes while pink genes are enriched for the GO category of nucleobase, nucleoside, nucleotide, and nucleic acid metabolism genes.

6 Methyl-methanesulfonate response and the role of Gcn4

The time series gene expression data for methyl-methanesulfonate (MMS) was obtained from (Gasch et al, 2001) and was sampled at 0m, 5m, 15m, 30m, 45m, 60m, 90m, and 120m. Genes were filtered if they had more than one missing data point, or their expression change did not exceed an absolute value of 0.6 at any time point leaving 2227 genes. The ChIP-chip data was obtained from (Workman et al, 2006) and consisted of 30 TFs from 60min into the MMS condition, and an additional 105 TFs profiled in YPD media. The YPD media ChIP-chip data was based on data originally published in (Lee et al, 2002). The procedure used here to convert the ChIP-chip data into a static input vector was different than used for the previous conditions. The static input vector for a gene had 135 elements. For 105 of the elements, corresponding to the 105 YPD media ChIP-chip experiments, a '1' was encoded if the gene was bound by the TF in the ChIP-chip experiment at a p-value < 0.005 otherwise a '0' was encoded. No motif constraints were used here. For the remaining 30 elements, corresponding to the 30 MMS ChIP-chip experiments, a '1' was encoded if the gene was bound by the TF in the ChIP-chip binding data at a p-value < 0.005, otherwise a '0' was encoded. Some of the TFs appeared twice in the input, once based on their YPD media ChIP-chip data and the other time based on their MMS ChIP-chip data.

Supplementary Figure 8a shows the portion of the map above the x-axis. This map indicates that the transcription factors Gcn4, Hsf1, Cad1, and Yap1, are associated with regulating genes increasing in expression around 15min. After 15min there is a bifurcation event in which some of the highly expressed genes at 15min continue to increase in expression while others stop increasing or start to decrease. Gcn4 based on a ChIP-chip experiment in YPD media is associated with the path of genes which did not increase in expression after 15min (split score 6×10^{-5} and overall score 7×10^{-19}). Genes on this path were enriched for GO categories such as amino acid metabolism (p-value $< 4 \times 10^{-19}$). The path that continues to increase after 15min is enriched for response to stress at a p-value of $< 10^{-7}$. Table 8 presents experimental results of ChIP-chip data showing for two repeats that Gcn4 indeed is more active in regulating genes at 15min into MMS compared to in YPD media. Also in one of the repeats a comparison was also made between MMS at 15min and 60min, and it was found that MMS was more active at 15min than 60min. Supplementary Figure 8b shows the portion of the map below the x-axis. The map predicts that a number of cell cycle transcription factors such as Ace2, Fkh2, Mcm1, Ndd1, Swi4, Swi5, and Swi6 can be used to explain some of the bifurcation events involving repressed genes.

Experimental results for Gcn4: Supplementary Table 8 presents results of genome wide binding experiments for Gcn4 in MMS at 0, 15, and 60 minutes following induction of stress. As can be seen, Gcn4 drastically expands the set

of genes it regulates in response to MMS. The increase is most noticeable at the 15m time point with up to four fold increase in the number of bound genes compared to YPD media and 10% increase over the 60m time point.

Supplementary Table 8: Gcn4 Binding in MMS Experimental Results

Gcn4 Binding Experiment	Intergenic Region < 0.001	Genes < 0.001	Intergenic Region < 0.005	Genes < 0.005
YPD Repeat 1	21	26	34	44
MMS 15min Repeat 1	106	155	151	221
YPD Repeat 2	23	26	38	45
MMS 15min Repeat 2	112	159	165	235
MMS 60min Repeat 2	107	154	150	212

Table 8: Summary of binding in Gcn4 genome wide binding experiments. Table gives number of intergenic regions bound and number of associated genes at both the 0.001 and 0.005 significance levels for two repeats. Only intergenic regions with assigned genes are included in the intergenic region counts.

Supplementary Figure 8: Dynamic Map of MMS Response

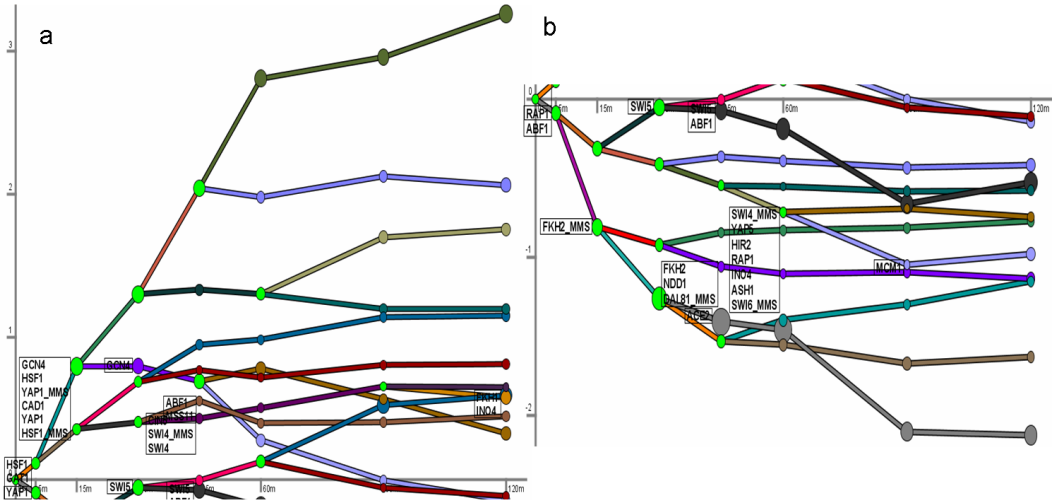

Figure 8: Temporal map derived by DREM for the MMS experiment of (Gasch et al, 2001). (a) Portion of map above the x-axis. (b) Portion of map below the x-axis. Those TF labels without 'MMS' are based on ChIP-chip experiments in YPD media, and those with the 'MMS' label are based on ChIP-chip experiments in the MMS condition. TF labels appear if their split score is less than 0.001 (see Supplementary Methods) and are ranked order by most significant score. The top of a box of TF labels is aligned with the top of the circle representing the next state on the path out of the split. The area of a circle is proportional to the standard deviation of the Gaussian distribution of the associated state.

7 Network Motifs

Network motifs correspond to recurring graph structures in regulatory and other networks (Milo et al, 2002). In the past these have either been studied in static graphs in one or more conditions (Luscombe et al, 2004) or independently analyzed for their dynamic properties (Prill et al, 2005). The ability to order the activation time of factors allows us to study these motifs in the general context of a dynamic response network. It also allows us to divide the set of factors into primary (controlling initial activation) and secondary (later time points) and test whether there is a difference in the utilization of these motifs based on the status of the factors involved. For this analysis we only used condition specific binding data. Having established Cbf1 and Gcn4 as primary activators of Amino Acid (AA) starvation response that are still active at the second time point we looked at the other transcription factors in Cbf1 and Gcn4 Feed Forward Loops (FFLs). Based on the AA ChIP-chip data we found support for six of these factors for Gcn4 (Arg81, Leu3, Met4, Put3, Rtg3, and Gln3), and one for Cbf1 (Met4) regulating a significant number of genes in a FFL (supplementary Figures 9 and 10). To investigate the effect of the FFL structure with Gcn4 or Cbf1 as the activators we compared the expression levels of genes bound by these FFLs with expression levels of genes regulated in another common network motifs; the Multiple Inputs motif with Gcn4 or Cbf1 (MIs, see Supplementary Figure 9) and the Single Input (SI) motif (that is, genes that are not regulated by any of the other 33 TFs with ChIP-chip data in AA). We compared the mean expression level of genes controlled by FFL, MI, and SI motifs. We found that the genes regulated in a FFL with Gcn4 and Cbf1 had the highest expression, MI regulated genes had the second highest expression levels and SI regulated genes had the lowest expression levels. The fact that these genes could keep their higher expression level in a FFL for a longer duration is consistent with a behavior previously noted (Mangan & Alon, 2002) of a common type of FFL, called a Coherent Type 1 OR, where the regulated gene is expressed at a higher level even after the initial signal is turned off. While FFLs seem to play an important role for regulating genes controlled by the two primary activators, FFLs involving other factors, including secondary factors, did not always exhibit such a behavior. Of the 8 additional factors controlling genes in a FFL, 5 did not display major differences between the expression values of these genes and genes controlled by MIs and SIs (Supplementary Figures 9 and 10).

Supplementary Figure 9: Motif Expression Comparison in Amino Acid Starvation Response

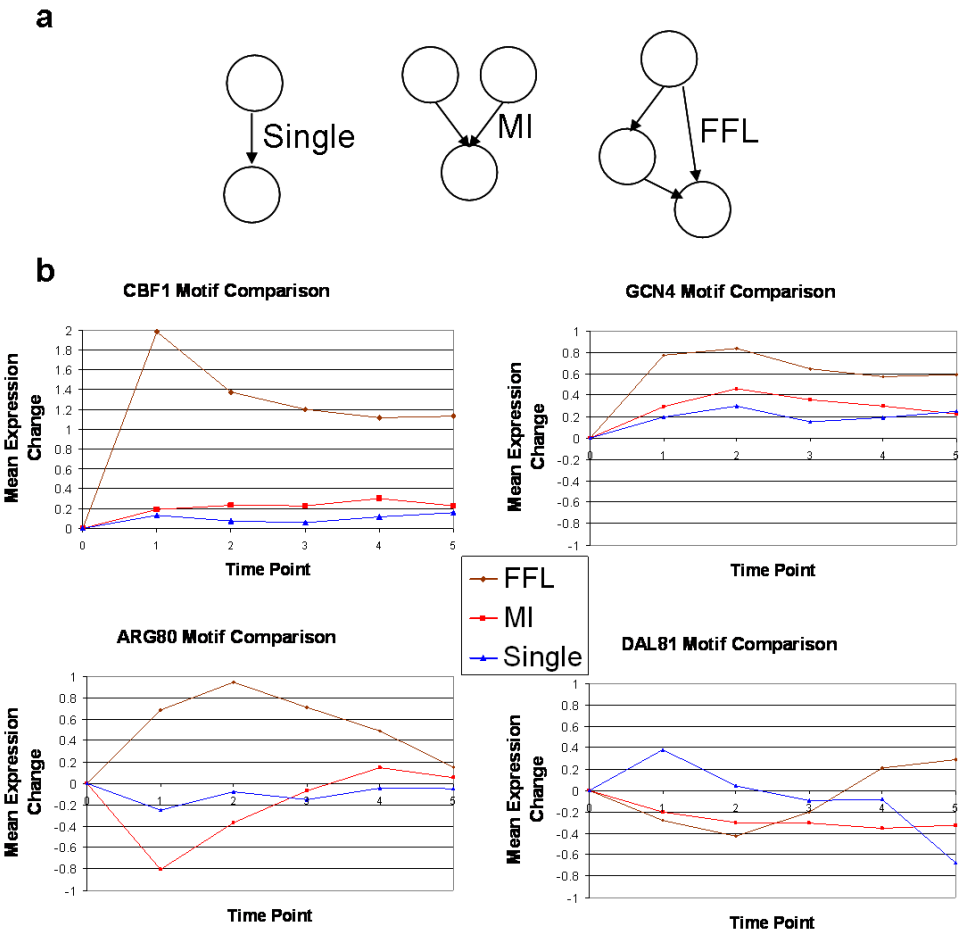

Figure 9: (a) Motif diagrams for Multiple Input (MI), Feed Forward Loop (FFL), and Single Input (SI). (b) Comparison of mean expression patterns of genes regulated in a FFL, MI and not an FFL, or SI for Cbf1, Gcn4, Arg80, and Dal81. Arg and Met complex TFs were considered to regulate genes even if not bound themselves if another TF in its complex bound the gene. Only TF pairs with significant intersection of target genes (number of genes ≥ 5 and intersection p-value < 0.005) were included as active FFLs.

Supplementary Figure 10: Motif Expression Comparison in Amino Acid Starvation Response (continued)

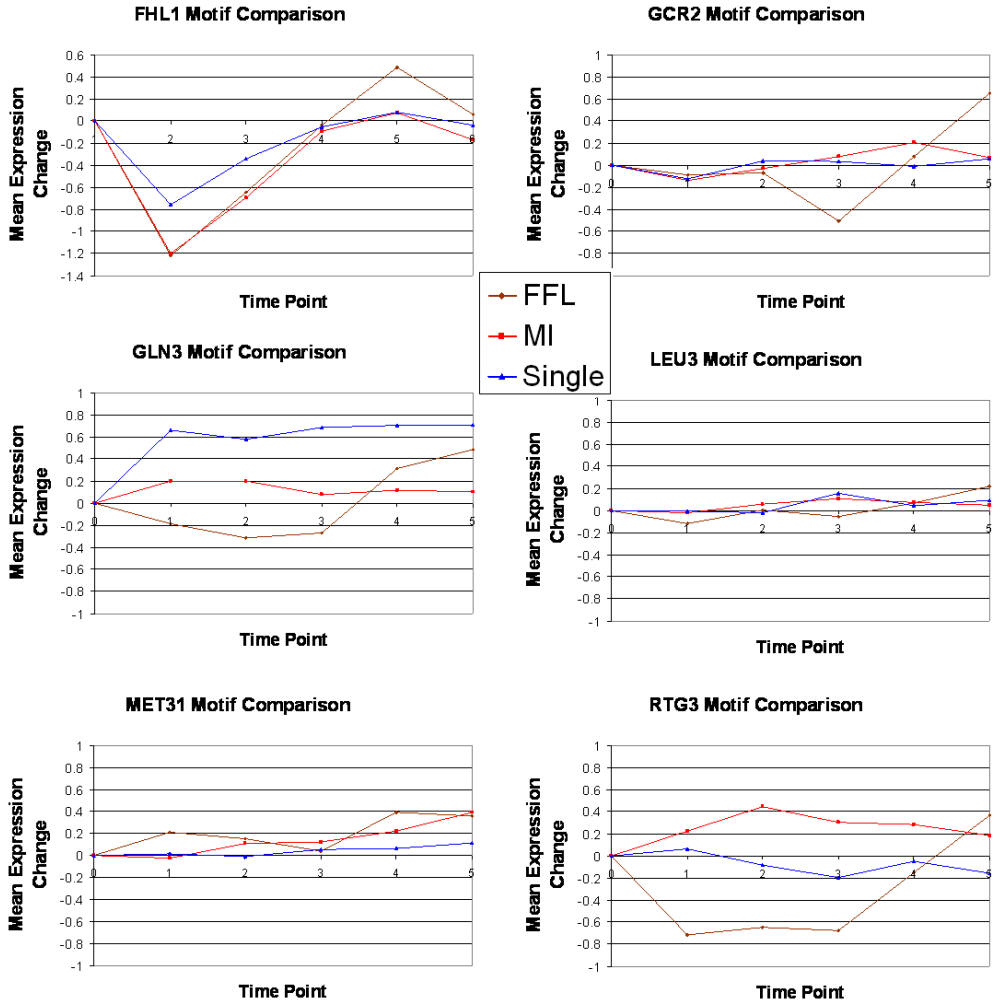

Figure 10: Comparison of mean expression patterns of genes regulated in a FFL, MI not an FFL, or SI for Fhl1, Gcr2, Gln3, Leu3, Met31, and Rtg3 under the same criteria as in the previous figure.

8 Condition specific regulation in response to stress

Below are the supplementary tables showing the percentage overlap between genes regulated in YPD media and the Amino Acid and Heat stress condition for Primary TFs (Supplementary Table 9) and Secondary TFs (Supplementary Table 10).

Supplementary Table 9: Percentage of Regulated Gene Overlap between YPD Media and Stress Condition for Primary TFs

TF and Condition	Top 100 Intersection %	Binding 0.005 Intersection %
Cbf1 AA	64	20
Gcn4 AA	75	33
Fhl1 AA	75	68
Rap1 AA	48	86
Sfp1 AA	21	6
Hsf1 Heat*	63	63
Msn2 Heat	3	0
Skn7 Heat	48	80

Table 9: Regulated gene intersection for primary TFs between YPD media and condition specific binding. Top 100 is the percentage of genes bound in the top 100 in both conditions. Binding at 0.005 is the percentage of the genes bound in both YPD media and the stress condition at < 0.005 p-value out of those that were bound in the stress condition. Binding data for all TFs is from (Harbison et al, 2004) except for Hsf1 which is from (Hahn et al, 2004). (*) For Hsf1 since equivalent binding p-values is not available, the overlap is only computed based on the intersection of genes associated with the main Hsf1 regulated genes identified from (Hahn et al, 2004) and using an equivalent number from the unstressed condition.

Supplementary Table 10: Percentage of Regulated Gene Overlap between YPD Media and Stress Condition for Secondary TFs

TF and Condition	Top 100 Intersection %	Binding 0.005 Intersection %
Arg81 AA	17	27
Dal82 AA	5	4
Gln3 AA	8	7
Hap5 AA	9	5
Met32 AA	24	35
Met4 AA	20	15
Rtg3 AA	14	13
Stp1 AA	21	12

Table 10: Regulated gene intersection for Secondary TFs between YPD media and condition specific binding. Top 100 is the percentage of genes bound in the top 100 in both conditions. Binding at 0.005 is the percentage of the genes bound in both YPD media and the stress condition at < 0.005 p-value out of those that were bound in the stress condition. Binding data for all TFs is from (Harbison et al, 2004).

9 The contributions of the dynamic and static data sources

Supplementary Figure 11 shows that randomizing the static input data, in this case the Amino Acid (AA) ChIP-chip data (Supplementary Figure 11), leads to maps with few or no TF labels. Supplementary Figures 12 and 13 compare models learned based on HMMs that use time series data only and models learned using an IOHMM which considers time series data and available ChIP-chip or motif data. These two figures show that a more coherent biological model is learned when considering both data sources jointly.

Supplementary Figure 11: Dynamic Maps for Amino Acid Starvation Response with Randomized ChIP-chip Input

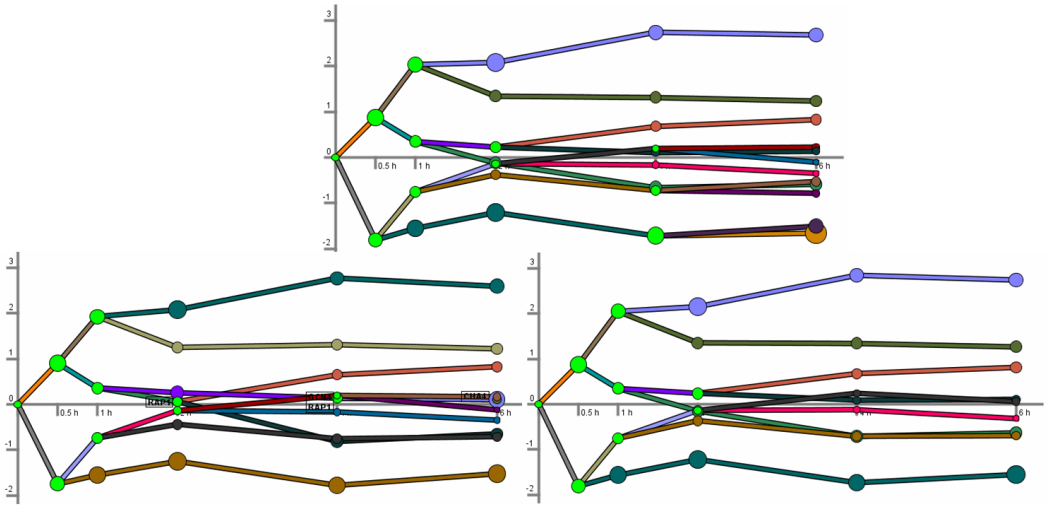

Figure 11: Three maps learned by DREM for AA starvation time series and randomized AA ChIP-chip binding data. In the randomization procedure the number of genes passing filter that a TF bound to was held fixed, but the set of genes it bound to was otherwise selected randomly. The randomization was applied independently for each TF. The entire randomization procedure was repeated three times to generate three different randomized input data sets. The above image shows the maps inferred by DREM using these randomized input data sets with TFs appearing if their split score was lower than 0.001 as in Figure 2a. In two cases there were no TF labels appearing on the map, while in the third case there was three unique TFs (Gcn4, Rap1, and Cha4) appearing (one TF appeared at two splits). In contrast in the map of Figure 2a there are 15 unique TFs appearing on the map. In addition, the temporal assignments differed significantly between the true and randomized data. While the true ChIP-chip data agreed well with the expression data in terms of the roles known TFs play in regulating this response, there was no such agreement for the randomized data. This demonstrates the compatibility of the ChIP-chip data and the time series data.

Supplementary Figure 12: Comparison of TF scores in an IOHMM vs. HMM

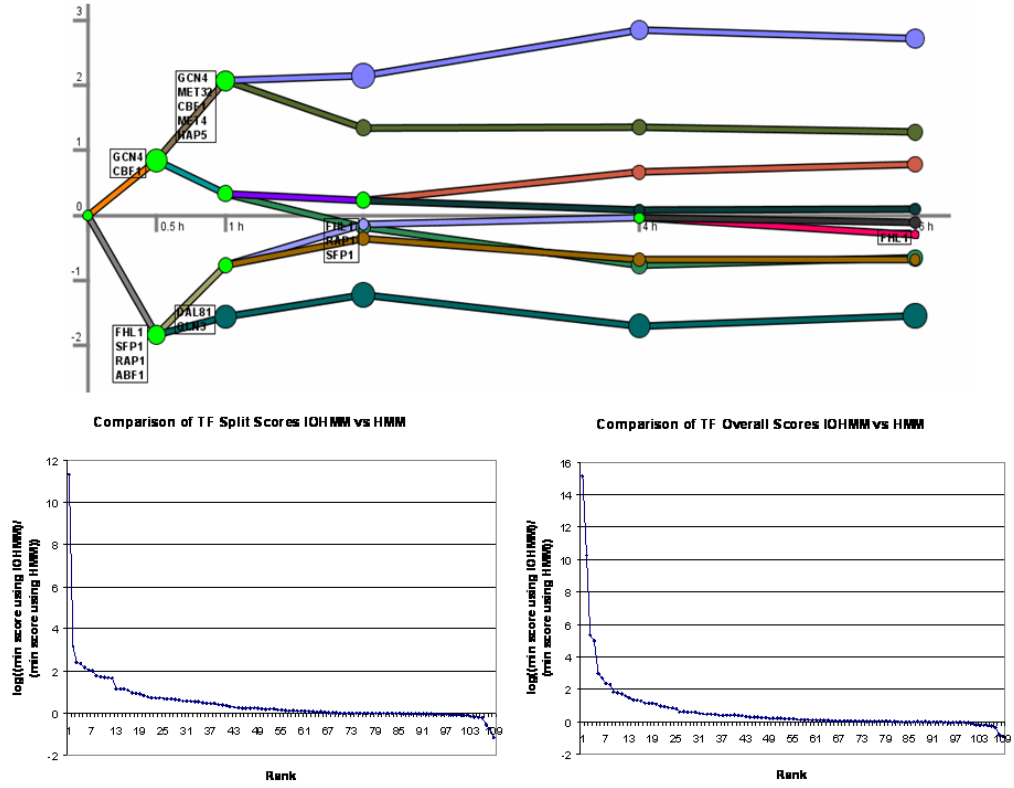

Figure 12: (Top) Map learned with an HMM that only uses the AA time series expression as input, and post-processing labeling the splits based on the same set of 109 TF used for Figure 2b. Compared to Figure 2b there are fewer relevant transcription factor (TF) assigned, especially for later time points. (Bottom Left) The rank order of the log-ratio of the minimum split score for a TF anywhere in the model using an IOHMM compared to an HMM. The scores are derived for each TF based on the median of 21 runs using different initial random seeds. The scores above 0 imply a stronger association for a split path in an IOHMM and those below 0 represent a stronger association in an HMM. (Bottom Right) Similarly the rank order of the log-ratio of the minimum overall score for a TF anywhere in the model using an IOHMM compared to a regular HMM. The graph shows for almost all TFs its minimum split and overall score is either better or about the same in an IOHMM as compared to an HMM. This is especially true for most of the top 15 factors which DREM determined to be associated with this response.

Supplementary Figure 13: Comparison of GO enrichment in an IOHMM vs. HMM

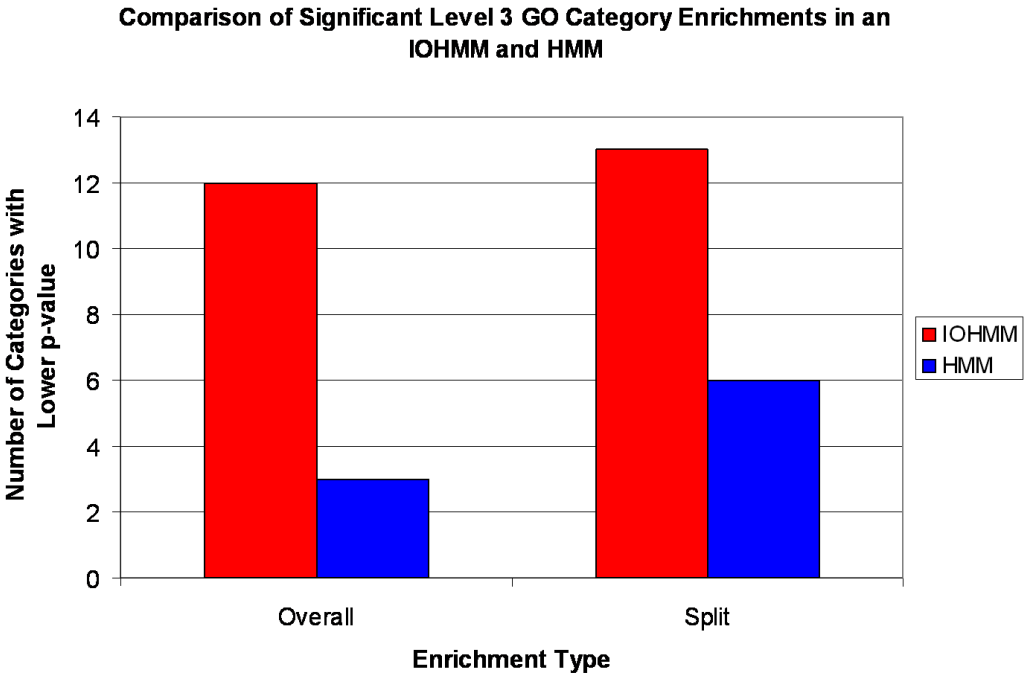

Figure 13: A comparison of level 3 Gene Ontology (GO) categories significantly enriched in paths of an IOHMM or a HMM using the same data as in Supplementary Figure 12. The comparison is restricted to GO categories with a minimum p-value < 0.001 in either model. The p-value for each category was computed based on the median of 21 runs with different initial random seeds. For computing overall enrichments for a path the base set of genes are all genes on the microarray, while in split enrichments the base set of genes are all genes going into the immediate previous split. The figure shows that more relevant GO categories had lower p-values with an IOHMM compared to an HMM.

10 Modeling Convergence of Paths from a Split

DREM can optionally model the convergence of paths from a prior split (see Supplementary Methods). Here we revisit the DTT data of Section 4 this time allowing convergence. In Supplementary Figure 14 (top) we show a dynamic map for the DTT condition allowing the merging of paths from prior splits. In this figure the states are spaced uniformly on the x-axis and are not displayed proportional to the actual sampling rate. The yellow circle in this image represents a merged state. In this case the repressed path at 60min splits into two paths and then these paths converge again around 480min. The genes on the most repressed path, the brown path, is enriched for GO cytosolic ribosome genes (p-value $< 10^{-152}$). While genes on the dark gray path are enriched for GO ribosome biogenesis genes (p-value $< 10^{-57}$). In Supplementary Figure 14 (bottom) the genes on these two paths are plotted with the x-axis scale proportional to the actual sampling rate.

Supplementary Figure 14: Example of a Convergence of a Split

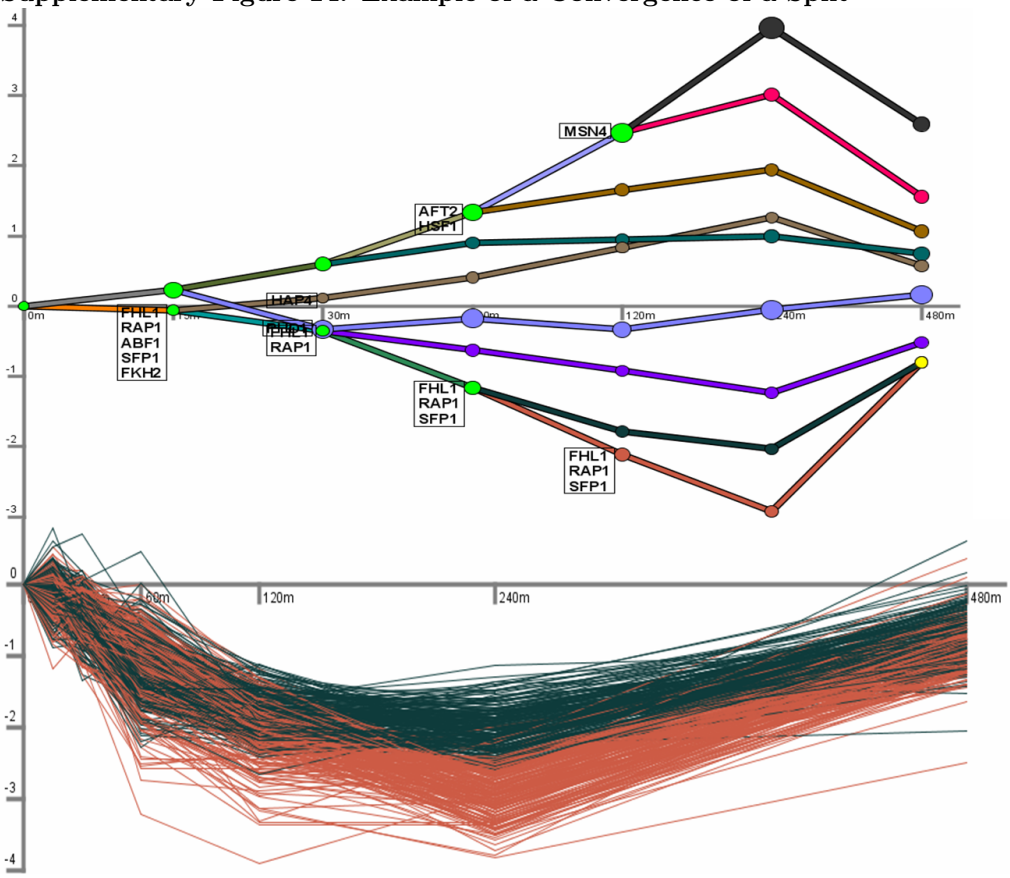

Figure 14: (Top) The DTT dynamic map with a convergence of the last split along the most repressed paths. The time points in this map are displayed spaced uniformly as opposed to the actual sampling rate. (Bottom) The time series of genes that go through the convergence state. The profiles are plotted based on the actual sampling rate.

11 Interpolation

In Supplementary Figure 15 we present a map of the data from the Amino Acid starvation example where we added a 15min time point based on interpolation. The interpolation was a linear interpolation where the value at 15min was half the value at 30min. The same TF-gene regulation predictions that was input for learning the map in Figure 2b was used here. The map in Supplementary Figure 15 shows that path controlled by the cell cycle TFs, Swi4, Swi6, and Mbp1 is initially less repressed than a path controlled by the ribosomal TFs Fhl1, Rap1, and Sfp1. This observation was not apparent from the map in Figure 2b without the interpolation.

Supplementary Figure 15: Example of Interpolating a Time Series

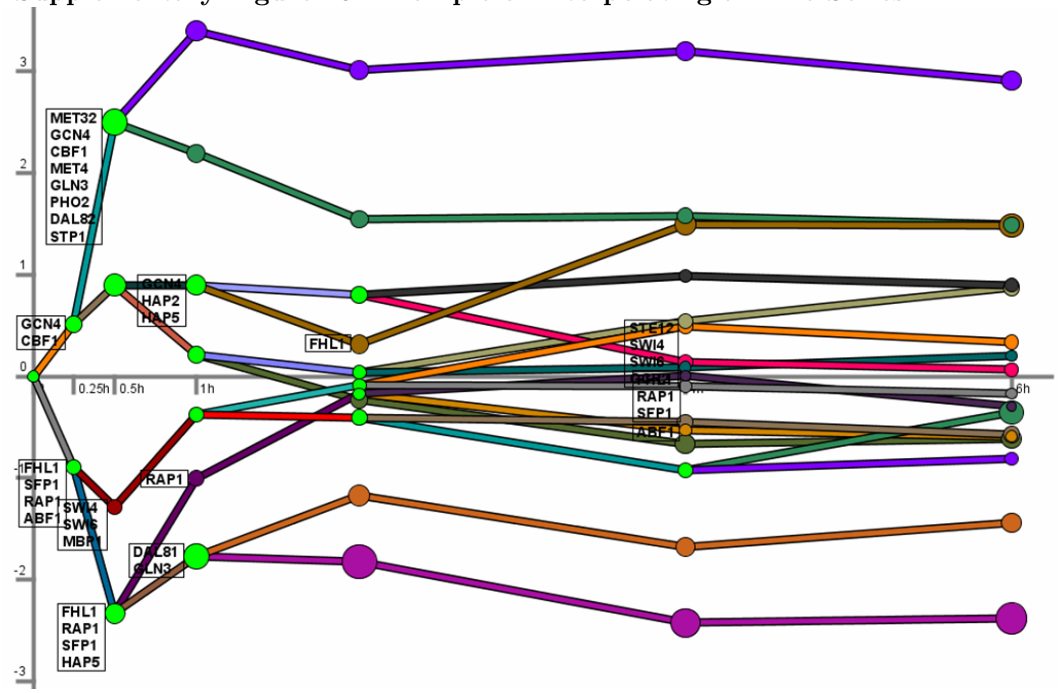

Figure 15: Above is a dynamic map for the AA starvation condition in which the values at the 15min time point were linearly interpolated as half the values at the 30min time point. TF labels that appear on the map all have a split association score of less than 0.001.

12 References

Gasch AP, Huang M, Metzner S, Botstein D, Elledge SJ, Brown PO (2001) Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. *Mol. Biol. Cell* **12**, 2987-3003.

Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**: 4241-4257.

Hahn JS, Hu Z, Thiele D, Iyer VR (2004) Genome-wide analysis of the biology of stress responses through heat. *Mol. Cell Biol.* **24**: 5249-5256.

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* 431: 99-104.

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne J, Volkert TL, Fraenkel E, Gifford DK, Young RA (2002) Transcriptional regulatory networks in Saccharomces cerevisiae. *Science* **298**: 799-804.

Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, and Gerstein M (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431**: 308-312.

Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci.* USA **100**: 11980-11985.

Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network Motifs: Simple Building Blocks of Complex Networks. *Science* **298**: 824-827.

Prill RJ, Iglesias PA, and Levchenko A (2005) Dynamic Properties of Network Motifs Contribute to Biological Network Organization, *Plos. Biol.* **3**, e343.

Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B (1998) Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol. Biol. Cell* **9**: 3273-3297.

Workman CT, Mak HC, McCuine S, Tagne JB, Agarwal M, Ozier O, Begley TJ, Samson LD, Ideker T (2006) A Systems Approach to Mapping DNA

Damage Response Pathways. *Science* **312**: 1054-1059.