**TITLE:** Dynamical Systems Modeling of Six Database-Derived Gene Regulatory Networks and Associated Random Networks Identifies Key Regulators Controlling the Early Response to Cold Shock in *Saccharomyces cerevisiae*

**INTRODUCTION:**

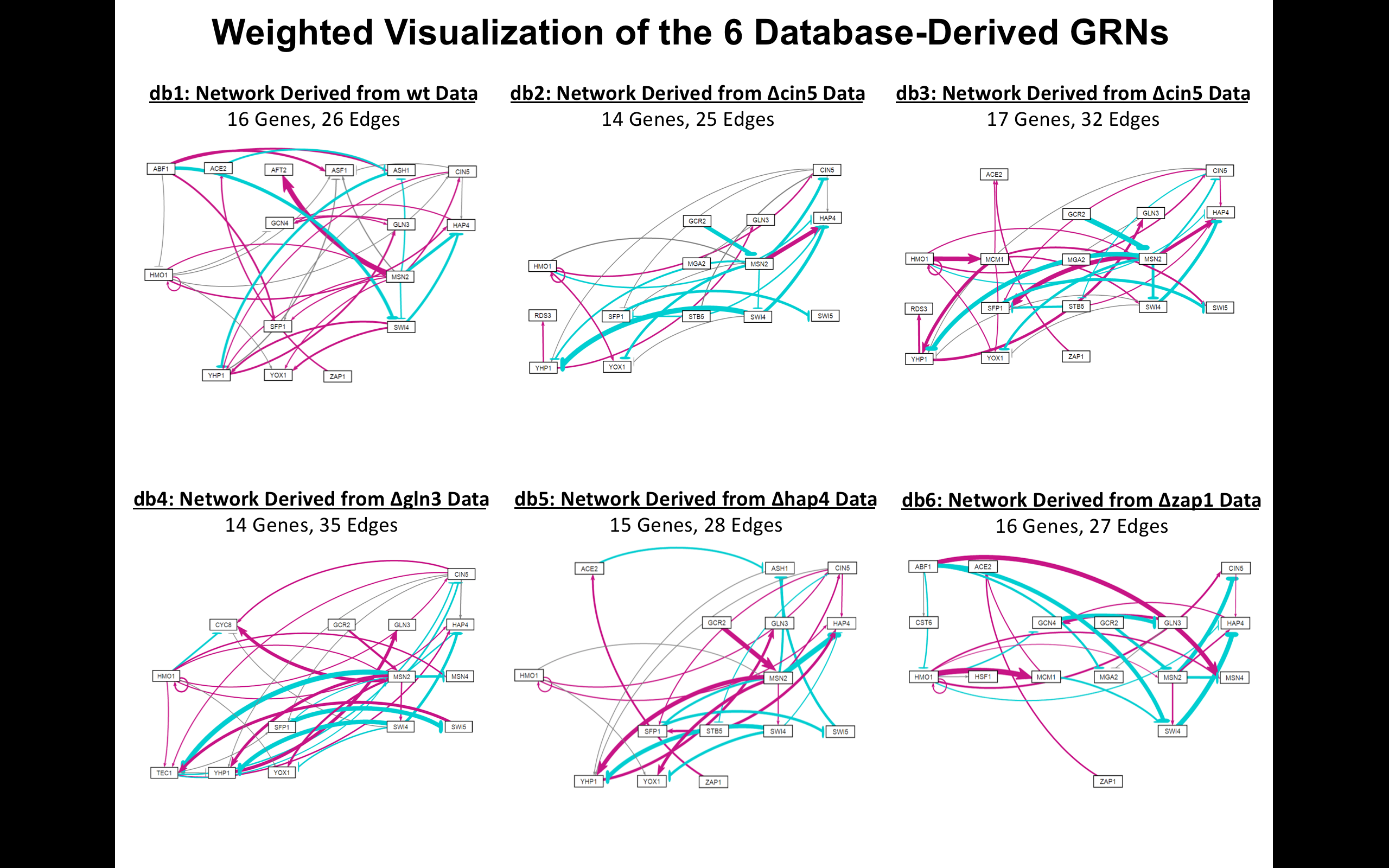
* *Saccharomyces cerevisiae* is a model eukaryotic organism for systems biology that has been extensively studied.
  + It has its own model organism database, the Saccharomyces Genome Database, and its study has precipitated numerous advances in genomics research (Cherry et al., 2011).
* Organisms must compensate for environmental perturbations throughout their lifetime, which is achieved through the modulation of gene expression.
  + Microarray experiments assessing the response of *S. cerevisiae* to varying environmental stressors have characterized a group of approximately 900 genes that consistently show similar expression profiles, referred to collectively as the environmental stress response, or ESR (Gasch et al., 2000).
  + Temperature stress is a ubiquitous environmental stressor that is easy to study in the lab, as has been done extensively in the case of *S. cerevisiae* and heat shock (Ingolia et al., 1982) (Morano et al., 2012) (Verghese et al., 2012). However, the transcriptional response to cold shock in yeast has received less attention.
* The documentation of cold-sensitive phenotypes in yeast points to the existence of a specific response to cold temperature stress (Hampsey, 1997).
  + Early low temperature experiments distinguished the cold-shock response from the near-freezing response, which is MSN2/4-dependent and shows significant overlap with the environmental stress response (Kandror et al. 2004).
  + Further experiments distinguished an early cold shock response occurring within 2 hours of cold temperature exposure from a late cold response occurring after 12 hours, which also resembles the ESR (Schade et al., 2004).
  + Microarray experiments have pointed to distinct molecular mechanisms governing the response to cold shock in *S. cerevisiae*.
    - Approximately 25% of the genome was is involved in the transcriptional response to low temperatures, and nearly one-third of all up-regulated genes have been observed to encode ribosomal proteins (Sahara et al., 2002).
    - Due to mRNA and ribosome instability in early cold shock, there is a need for the upregulation of ribosome biogenesis to compensate for impaired translation (Al-Fageeh and Smales, 2006).
    - This cold ribosome adaptation hypothesis represents one mechanism thought to be involved in the early response to cold shock (Aguilera et al., 2007).
* Gene expression changes are controlled by transcription factors, which themselves are regulated by other transcription factors. Together, networks of transcription factors controlling the expression of target sets of genes form gene regulatory networks (GRNs), which control global transcriptional responses in the cell.
  + Through genome-wide location analysis and motif analysis, a 141 node GRN controlling the global transcriptional response in yeast has been proposed (Lee et al., 2002).
  + More recently, a global genetic interaction network was developed in an effort to computationally model overall cellular function (Costanzo et al., 2016).
* The Dahlquist Lab has previously described the use of GRNmap, a MATLAB software package that uses mass balance ordinary differential equations to model gene regulatory networks and estimate associated parameters from noisy, temporally sparse microarray data (Dahlquist et al., 2015).
  + The primary aim of this investigation was to analyze existing cold shock microarray data with GRNmap to identify and model the gene regulatory network controlling the early cold shock response in *Saccharomyces cerevisiae*.
  + Secondary aims of this study included validating GRNmap and developing semi-automated tools for the post-hoc analysis of GRNs modeled by GRNmap.

**MATERIALS AND METHODS:**

* **Statistical Analysis of Microarray Data and Generation of Hypothesis Networks**
  + GitHub Issue: <https://github.com/kdahlquist/GRNmap/issues/143>
  + Steps 6, 10, and 11 of the Dahlquist Lab microarray data analysis workflow will be summarized: [www.openwetware.org/wiki/Dahlquist:Microarray\_Data\_Analysis\_Workflow](http://www.openwetware.org/wiki/Dahlquist:Microarray_Data_Analysis_Workflow)
  + The YEASTRACT database was used to identify regulators of the genes exhibiting significant B&H expression changes in the microarray data (Teixeira et al., 2017)
* **Generation of GRNmap Input Workbooks**
  + GitHub Issue: <https://github.com/kdahlquist/GRNmap/issues/245>
  + The protocol for formatting input workbooks for GRNmap will be summarized: <https://github.com/kdahlquist/GRNmap/wiki/How-to-format-the-input-file-for-GRNmap-v1.4-and-above>
    - Degradation rates and derived production rates were adapted from Neymotin et al., 2014.
    - The expression data used is publically accessible from the following link: <https://github.com/kdahlquist/DahlquistLab/blob/master/GEO_submission/Dahlquist_GA_dual_ch_w_platf_20160613.xls?raw=true>
* **Network Modeling and Parameter Estimation Using GRNmap**
  + The GRNmap software was used for modeling and parameter estimation for gene regulatory networks in *S. cerevisiae* (Dahlquist et al., 2015).
    - The mass balance ordinary differential equations used by GRNmap and the penalized least squares approach will be described.
    - Running GRNmap in MATLAB will be described.
  + GRNmap download link: <http://kdahlquist.github.io/GRNmap/downloads>
* **Visualization of Weighted Networks Using GRNsight**
  + The GRNsight web app was used for visualization of individual transcription factors, the regulatory relationships connecting them, and the weight of these regulatory relationships (Dahlquist et al., 2016).
  + GRNsight link: <http://dondi.github.io/GRNsight/index.html>
* **Post Hoc Statistical Analysis of Network Parameters**
  + The following post hoc network analyses were automated through the development of R scripts of custom Excel workbooks:
    - R script—generation of degree distribution charts from input adjacency matrices (GitHub Issue: <https://github.com/kdahlquist/GRNmap/issues/230>)
    - R script—generation of L-curve plots (GitHub Issue: <https://github.com/kdahlquist/GRNmap/issues/351>)
      * L-curve plots can be used to select assess the efficiency of regularization methods and the selection of regularization parameters (Hansen and O’Leary, 1993)
    - R script—generation of random networks with specified numbers of nodes and edges (GitHub Issue: <https://github.com/kdahlquist/GRNmap/issues/223>)
    - Custom Excel workbook—statistical analysis of edge weights and heat map generation (GitHub Issue: <https://github.com/kdahlquist/GRNmap/issues/315>)
  + Multiple regression analyses and weighted degree distribution charts were generated in SPSS.
    - Multiple regression protocol: <https://openwetware.org/wiki/Analyzing_GRNmap_Output_Workbooks_Using_Multiple_Regression_and_SPSS>
      * GitHub Issue: <https://github.com/kdahlquist/GRNmap/issues/330>
    - Weighted degree distribution charts protocol: <https://openwetware.org/wiki/Generating_Distribution_Charts_and_Cumulative_Plots_for_GRNmap_Weight_Values_in_SPSS>
      * GitHub Issue: <https://github.com/kdahlquist/GRNmap/issues/315>

**Results:**

**I. Comparative Analysis of the Six Database-Derived GRNs**

****

**Figure 1 (precursor).** Weighted visualizations of the six database-derived networks, db1-db6, were produced in GRNsight.

**db1-db6.pdf**

**Figure 2.** L-curve comparison of db1-db6.

**Table 1.** GRNmap network parameter outputs for db1-db6.

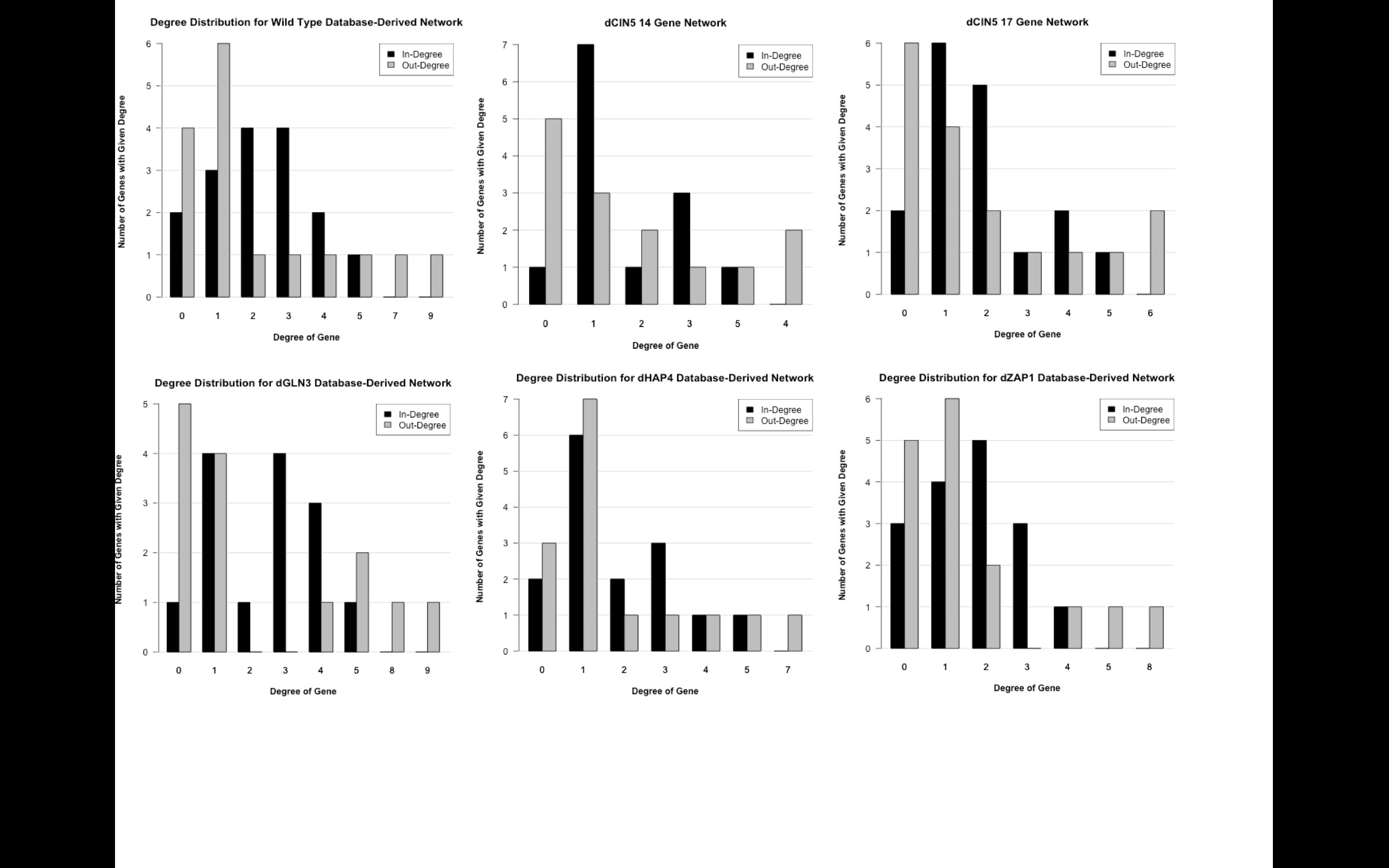
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** | **db1** | **db2** | **db3** | **db4** | **db5** | **db6** |
| Penalty Term | 2.5923 | 1.8830 | 1.7570 | 1.7895 | 2.3443 | 1.8276 |
| LSE | 0.8194 | 0.6634 | 0.7864 | 0.6994 | 0.6919 | 0.8602 |
| minLSE | 0.5768 | 0.4885 | 0.5449 | 0.5379 | 0.4851 | 0.6156 |
| LSE:minLSE Ratio | 1.4206 | 1.3580 | 1.4100 | 1.3000 | 1.4263 | 1.3973 |
| Iteration Count | 109,718 | 53,862 | 118,921 | 78,124 | 62,139 | 76,769 |

* LSE:minLSE ratios for db1-db6 ranged from 1.3000 (db4) to 1.4263 (db5), indicating that GRNmap consistently modeled the dynamics of small GRNs well.
* The database-derived network with the lowest LSE:minLSE ratio was db4 (1.3000), indicating that this network best modeled the regulatory dynamics observed in the cold shock experiment from which the network was derived.
  + db4 contains the lowest number of genes (14) and the highest number of edges (35).
  + db4 was the only database-derived network to include the transcription factors CYC8 and TEC1 (Table 2), both of which demonstrated interesting connectivity to the network. CYC8 exhibited an in-degree of 4 and an out-degree of 0, while TEC1 showed an in-degree of 4 and out-degree of 5. The high betweenness centrality of TEC1 (24.3333) trailed only behind that of MSN2 in this network (26.8333), both of with exceeded the betweenness centrality measures of the nodes in the five other database-derived networks.
    - Removal of nodes with high betweenness centrality most disrupts the communications within a GRN, as these nodes lie along the largest number of shortest paths between other nodes in the network. MSN2 is present in all six database-derived networks and consistently exhibits a high betweenness centrality. However, TEC1 is only included in db4. If TEC1 is part of the “true” GRN controlling the response to cold shock in yeast, we would expect its exclusion from a network to substantially hinder the GRN’s ability to model cold shock experiment microarray data.
      * However, of TEC1 inclusion in this network is likely due to the fact that db4 was derived from a ∆gln3 deletion strain of *S. cerevisiae*. Deletion of gln3 is known to abolish pseudohyphal growth; and the transcription factor TEC1—which is known to regulate pseudohyphal growth—was substantially down-regulated during the ∆gln3 deletion strain cold shock experiment.
      * If TEC1 is not related to the response to cold shock, then its exclusion from other networks would not be relevant. Instead, its inclusion in db4 could merely have reduced the resulting LSE:minLSE ratio for db4 by increasing the number of parameters (35 edges) available within a 14-gene GRN (compared to 25 edges in db2, another 14-gene GRN).
* db5 exhibited the highest LSE:minLSE ratio (1.4263). Thus, it poorly modeled the cold shock microarray data from which it was derived compared to the other database-derived networks.
  + db5 contains an intermediate number of genes (15) and edges (28).

**Table 2.** Gene list for db1-db6. Factors that were deleted in tested strains of *Saccharomyces cerevisiae* are highlighted in yellow. Counts of the number of networks in which a gene is represented are displayed on the right.

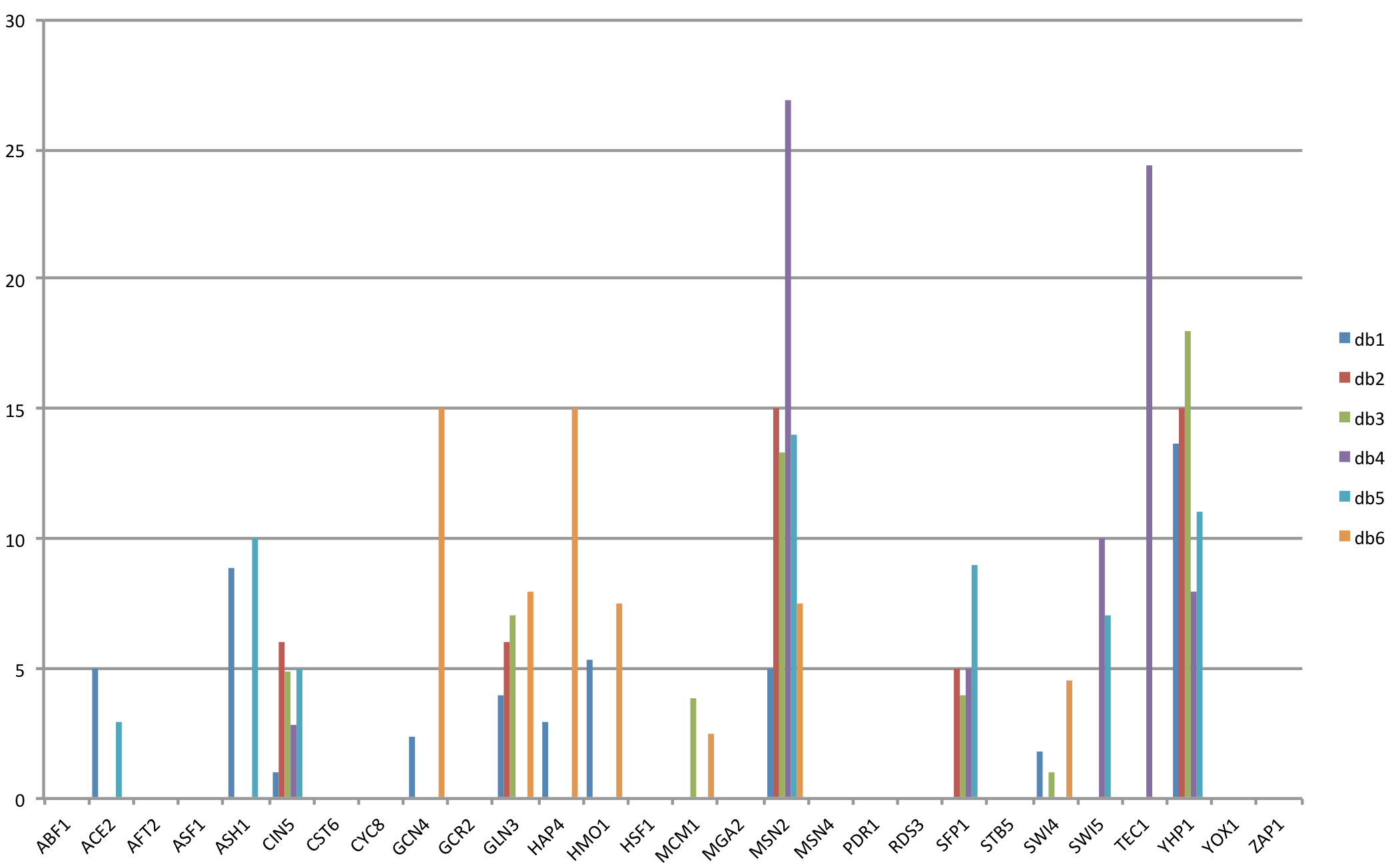
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **All Genes** | **db1** | **db2** | **db3** | **db4** | **db5** | **db6** | **Count** |
| ABF1 | ABF1 |  |  |  |  | ABF1 | 2 |
| ACE2 | ACE2 |  | ACE2 |  | ACE2 | ACE2 | 4 |
| AFT2 | AFT2 |  |  |  |  |  | 1 |
| ASF1 | ASF1 |  |  |  |  |  | 1 |
| ASH1 | ASH1 |  |  |  | ASH1 |  | 2 |
| CIN5 | CIN5 | CIN5 | CIN5 | CIN5 | CIN5 | CIN5 | 6 |
| CST6 |  |  |  |  |  | CST6 | 1 |
| CYC8 |  |  |  | CYC8 |  |  | 1 |
| GCN4 | GCN4 |  |  |  |  | GCN4 | 2 |
| GCR2 |  | GCR2 | GCR2 | GCR2 | GCR2 | GCR2 | 5 |
| GLN3 | GLN3 | GLN3 | GLN3 | GLN3 | GLN3 | GLN3 | 6 |
| HAP4 | HAP4 | HAP4 | HAP4 | HAP4 | HAP4 | HAP4 | 6 |
| HMO1 | HMO1 | HMO1 | HMO1 | HMO1 | HMO1 | HMO1 | 6 |
| HSF1 |  |  |  |  |  | HSF1 | 1 |
| MCM1 |  |  | MCM1 |  |  | MCM1 | 2 |
| MGA2 |  | MGA2 | MGA2 |  |  | MGA2 | 3 |
| MSN2 | MSN2 | MSN2 | MSN2 | MSN2 | MSN2 | MSN2 | 6 |
| MSN4 |  |  |  | MSN4 |  | MSN4 | 2 |
| RDS3 |  | RDS3 | RDS3 |  |  |  | 2 |
| SFP1 | SFP1 | SFP1 | SFP1 | SFP1 | SFP1 |  | 5 |
| STB5 |  | STB5 | STB5 |  | STB5 |  | 3 |
| SWI4 | SWI4 | SWI4 | SWI4 | SWI4 | SWI4 | SWI4 | 6 |
| SWI5 |  | SWI5 | SWI5 | SWI5 | SWI5 |  | 4 |
| TEC1 |  |  |  | TEC1 |  |  | 1 |
| YHP1 | YHP1 | YHP1 | YHP1 | YHP1 | YHP1 |  | 5 |
| YOX1 | YOX1 | YOX1 | YOX1 | YOX1 | YOX1 |  | 5 |
| ZAP1 | ZAP1 |  | ZAP1 |  | ZAP1 | ZAP1 | 4 |

* MSN2 and SWI4 were included in each of the six database-derived networks based on empirical evidence. CIN5, GLN3, HAP4, and HMO1 were manually included in all six GRNs, as deletion strain data was obtained from yeast in which these genes were knocked out.
* AFT2 and ASF1 only appear in db1.
  + Neither regulates other genes in the network (out-degrees of 0).
* CYC8 and TEC1 only appear in db4.
  + CYC8 does not regulate other genes in the network (out-degree of 0), but TEC1 regulates the expression of five other transcription factors in the network.
* CST6 and HSF1 only appear in db6.
  + Neither regulates other genes in the network (out-degrees of 0).
* SFP1, YHP1, and YOX1 are only *excluded* in db6.
  + db6 exhibited the second lowest LSE:min LSE ratio (1.4206).

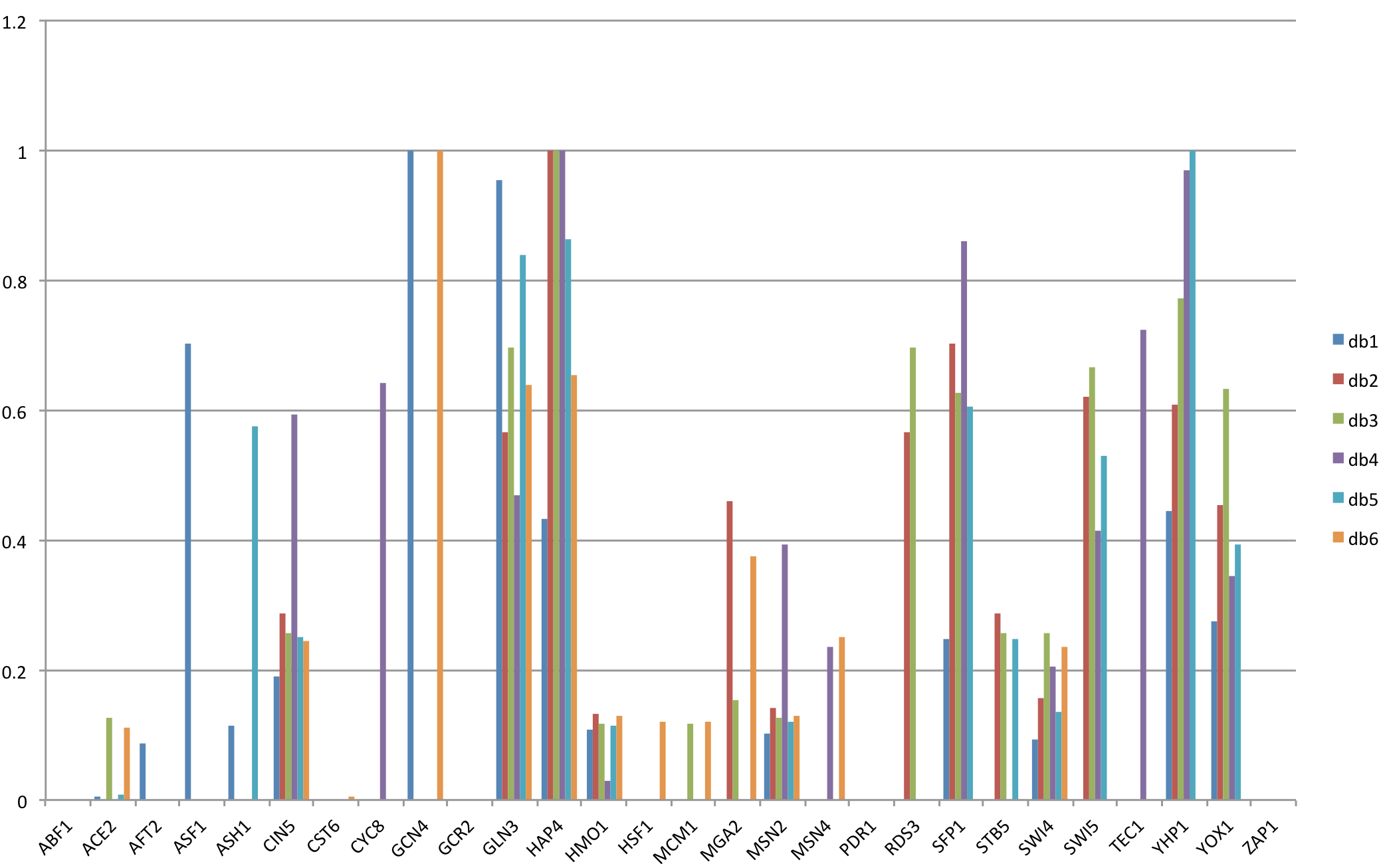
****

**Figure 3 (precursor).** Degree distribution charts for db1-db6.

* Comparison of graph statistics in db1-db6 (include? If so, need to add Gephi to methods)



**Figure 4 (precursor).** Comparison of betweenness centrality for individual nodes in db1-db6.



**Figure 5 (precursor).** Comparison of eigen centrality for individual nodes in db1-db6.

**Table 3.** Characterization of weighted edges in db1-db6.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **db1** | **db2** | **db3** | **db4** | **db5** | **db6** |
| # Activation | 19 | 6 | 18 | 16 | 14 | 13 |
| # Repression | 7 | 11 | 10 | 12 | 10 | 11 |
| # Small Influence | 10 | 8 | 4 | 7 | 4 | 3 |
|  |  |  |  |  |  |  |
| **SUM:** | 8.1432 | -13.2390 | 1.5688 | -2.5750 | 1.2985 | -5.8876 |
| **SUM(activation):** | 23.4929 | 6.9575 | 18.8606 | 17.1864 | 17.7493 | 13.5913 |
| **SUM(repression):** | -15.3497 | -20.1966 | -17.2608 | -19.7614 | -16.4508 | -19.4790 |
| **SUM(abs):** | 38.8426 | 27.1541 | 36.1524 | 36.9479 | 34.2000 | 33.0703 |
| **MAX:** | 5.9424 | 3.0815 | 3.1810 | 2.4940 | 3.2167 | 3.7631 |
| **MIN:** | -3.3601 | -4.2947 | -3.6707 | -4.0281 | -4.0374 | -3.5141 |
| **MAX(abs):** | 5.9424 | 4.2947 | 3.6707 | 4.0281 | 4.0374 | 3.7631 |
| **“Small Influence” Threshold:** | 0.2971 | 0.2147 | 0.1835 | 0.2014 | 0.2019 | 0.1882 |
| **MEAN:** | 0.2262 | -0.5296 | 0.0490 | -0.0736 | 0.0464 | -0.2181 |
| **STDEV:** | 1.6215 | 1.5085 | 1.5597 | 1.5026 | 1.6433 | 1.6812 |
| **MEDIAN:** | 0.3310 | -0.1214 | 0.2323 | 0.1139 | 0.2144 | 0.1276 |

* In **db1**, there was **substantially** **more activation than repression** of genes in the network.
  + Edge counts: 19 activation vs. 7 repression
  + Sum of activation edge weights = 23.4929
  + Sum of repression edge weights = -15.3497
* In **db2**, there was **substantially** **more repression that activation** of genes in the network.
  + Edge counts: 6 activation vs. 11 repression
  + Sum of activation edge weights = 6.9575
  + Sum of repression edge weights = -20.1966
* db1 and db2 exhibited a large number of small influence regulatory relationships (~30%), which we define as edges with regulatory weights < 5% of the maximum edge weight in the network. This contrasts to db3, db5, and db6, where < 15% of edges were classified as small influence.
  + In db1, the greatest activation relationship exhibited an edge weight of 5.9424, which far exceeded the weight of any other activation relationship in db2-db6 (next highest: 3.7631, db6).
    - The regulatory relationship in question is MSN2 🡪 AFT2.
    - Note: AFT2 only appears in db1, where MSN2 is its only input.
  + In db2, the greatest repression relationship exhibited an edge weight of -4.2947, which was the highest magnitude repression relationship of any network.
    - The regulatory relationship in question is SWI4 🡪 YHP1.
    - Note: This relationship appears in three other networks, although with lower magnitudes (and as a weak activation relationship in the case of db1).
  + These outliers contribute to the observed increase in “small influence” edges in db1 and db2. They might also influence the large imbalance of activation and repression in these two networks.
* In **db3-db5**, there were more activation than repression regulatory relationships, although the **overall magnitudes of activation and repression in the networks are similar**.
  + Activating edges: 18, 16, 14
  + Repressing edges: 10, 12, 10
  + Sum of activation edge weights = 18.8606, 17.1864, 17.7493
  + Sum of repression edge weights = -17.2608, -19.7614, -16.4508
* In **db6**, there were similar numbers of activation and repression regulatory relationships, although the **overall magnitude of repression exceeded that of activation in the network**.
  + Edge counts: 13 activation vs. 11 repression
  + Sum of activation edge weights = 13.5913
  + Sum of repression edge weights = -19.4790
* For db2-db5, the edge weights with the largest magnitudes cluster together in a relatively tight range: 3.6707-4.2947.

**Table 4.** Heat map comparing the edge weights of regulatory relationships in db1-db6.

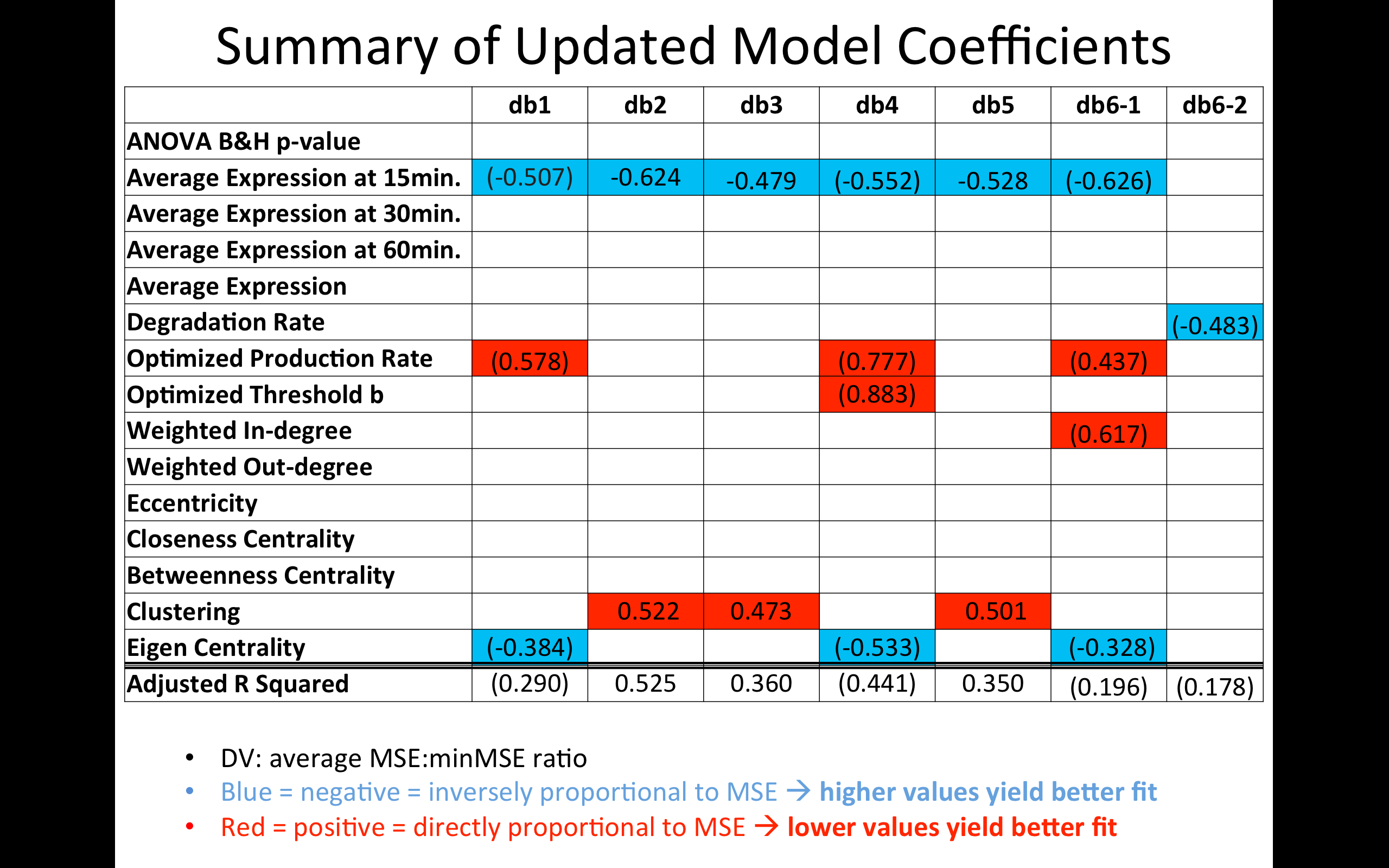
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **db1** | **db2** | **db3** | **db4** | **db5** | **db6** |  | **db1** | **db2** | **db3** | **db4** | **db5** | **db6** |
| ABF1->ASF1 |  |  |  |  |  |  | MCM1->SWI4 |  |  |  |  |  |  |
| ABF1->ASH1 |  |  |  |  |  |  | MCM1->SWI5 |  |  |  |  |  |  |
| ABF1->CST6 |  |  |  |  |  |  | MCM1->YHP1 |  |  |  |  |  |  |
| ABF1->HMO1 |  |  |  |  |  |  | MCM1->YOX1 |  |  |  |  |  |  |
| ABF1->MGA2 |  |  |  |  |  |  | MSN2->AFT2 |  |  |  |  |  |  |
| ABF1->MSN4 |  |  |  |  |  |  | MSN2->ASF1 |  |  |  |  |  |  |
| ABF1->SFP1 |  |  |  |  |  |  | MSN2->ASH1 |  |  |  |  |  |  |
| ABF1->SWI4 |  |  |  |  |  |  | MSN2->CIN5 |  |  |  |  |  |  |
| ACE2->ASH1 |  |  |  |  |  |  | MSN2->CYC8 |  |  |  |  |  |  |
| ASH1->YHP1 |  |  |  |  |  |  | MSN2->HAP4 |  |  |  |  |  |  |
| CIN5->ASF1 |  |  |  |  |  |  | MSN2->MSN4 |  |  |  |  |  |  |
| CIN5->CYC8 |  |  |  |  |  |  | MSN2->SFP1 |  |  |  |  |  |  |
| CIN5->HAP4 |  |  |  |  |  |  | MSN2->SWI4 |  |  |  |  |  |  |
| CIN5->SFP1 |  |  |  |  |  |  | MSN2->TEC1 |  |  |  |  |  |  |
| CIN5->STB5 |  |  |  |  |  |  | MSN2->YHP1 |  |  |  |  |  |  |
| CIN5->TEC1 |  |  |  |  |  |  | MSN2->YOX1 |  |  |  |  |  |  |
| CIN5->YHP1 |  |  |  |  |  |  | SFP1->SWI5 |  |  |  |  |  |  |
| GCN4->GLN3 |  |  |  |  |  |  | STB5->HAP4 |  |  |  |  |  |  |
| GLN3->GCN4 |  |  |  |  |  |  | STB5->SFP1 |  |  |  |  |  |  |
| GLN3->MGA2 |  |  |  |  |  |  | SWI4->CYC8 |  |  |  |  |  |  |
| GCR2->MSN2 |  |  |  |  |  |  | SWI4->HAP4 |  |  |  |  |  |  |
| HAP4->GCN4 |  |  |  |  |  |  | SWI4->YHP1 |  |  |  |  |  |  |
| HMO1->ASF1 |  |  |  |  |  |  | SWI4->YOX1 |  |  |  |  |  |  |
| HMO1->CIN5 |  |  |  |  |  |  | SWI5->ASH1 |  |  |  |  |  |  |
| HMO1->CYC8 |  |  |  |  |  |  | SWI5->TEC1 |  |  |  |  |  |  |
| HMO1->GCN4 |  |  |  |  |  |  | TEC1->CIN5 |  |  |  |  |  |  |
| HMO1->HAP4 |  |  |  |  |  |  | TEC1->HAP4 |  |  |  |  |  |  |
| HMO1->HMO1 |  |  |  |  |  |  | TEC1->MSN2 |  |  |  |  |  |  |
| HMO1->HSF1 |  |  |  |  |  |  | TEC1->SFP1 |  |  |  |  |  |  |
| HMO1->MCM1 |  |  |  |  |  |  | TEC1->YHP1 |  |  |  |  |  |  |
| HMO1->MSN2 |  |  |  |  |  |  | YHP1->ASF1 |  |  |  |  |  |  |
| HMO1->MSN4 |  |  |  |  |  |  | YHP1->GLN3 |  |  |  |  |  |  |
| HMO1->TEC1 |  |  |  |  |  |  | YHP1->RDS3 |  |  |  |  |  |  |
| HMO1->YOX1 |  |  |  |  |  |  | ZAP1->ACE2 |  |  |  |  |  |  |
| MCM1->ACE2 |  |  |  |  |  |  |  |  |  |  |  |  |  |



**Maximum: 1**

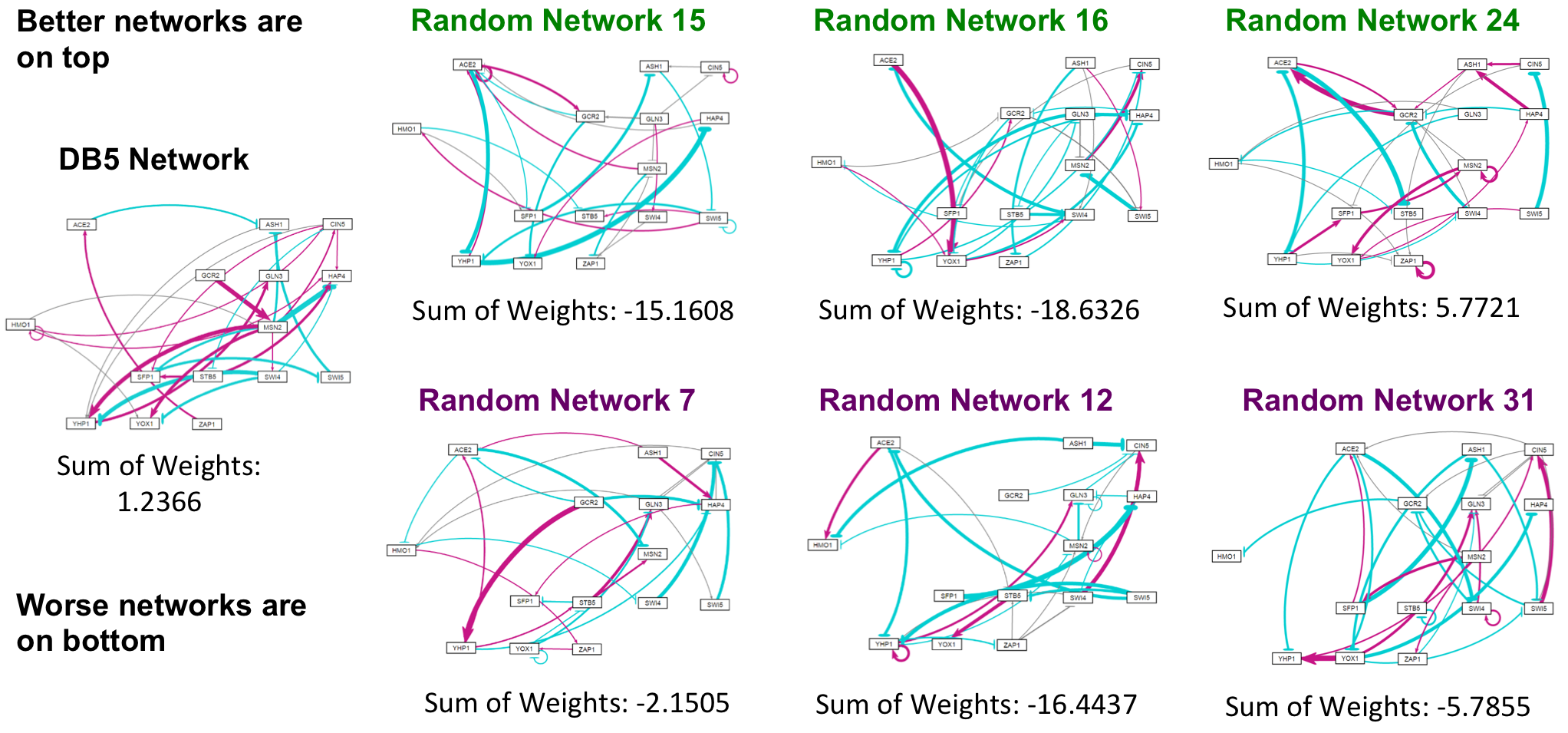
**Minimum: -1**

* The following regulatory relationships were conserved across the database-derived networks, always appearing as activation only or repression only when present:
  + Activation, CIN5🡪HAP4 (6 networks)
  + Activation, HMO1🡪CIN5 (6 networks)
  + Activation, HMO1🡪HMO1 (6 networks)
  + Activation, HMO1🡪MSN2 (6 networks)
  + Activation, HMO1🡪YOX1 (5 networks)
  + Activation, YHP1🡪GLN3 (5 networks)
  + Activation, ZAP1🡪ACE2 (4 networks)
  + Repression, SWI4🡪HAP4 (6 networks)
  + Repression, SFP1🡪SWI5 (4 networks)
  + The conserved activation relationships among CIN5, GLN3, HMO1, and YHP1 form the following regulatory chain: HMO1 🡪 CIN5 🡪 YHP1 🡪 GLN3.
* The following high magnitude edge weights appeared in the database-derived networks both as activation and repression relationships:
  + GCR2🡪MSN2 (2 Activation, 3 Repression)
    - Note: MSN2 is the only connection to GCR2 in these networks.
  + MSN2🡪CIN5 (2 Activation, 4 Repression)
  + MSN2🡪HAP4 (3 Activation, 3 Repression)
  + MSN2🡪SFP1 (3 Activation, 2 Repression)
  + MSN2🡪SWI4 (3 Activation, 3 Repression)
  + MSN2🡪YHP1 (3 Activation, 2 Repression)
  + MSN2🡪YOX1 (3 Activation, 2 Repression)
  + SWI4🡪YHP1 (2 Activation, 3 Repression)
  + SWI4🡪YOX1 (1 Activation, 2 Repression)
  + MCM1🡪SWI4 (1 Activation, 1 Repression)
  + All but one of these regulatory relationships involve either MSN2 or YHP1, which both exhibit high betweenness centralities.
* Multiple regression analysis to identify determinants of node MSE:minMSE ratio in db1-db6

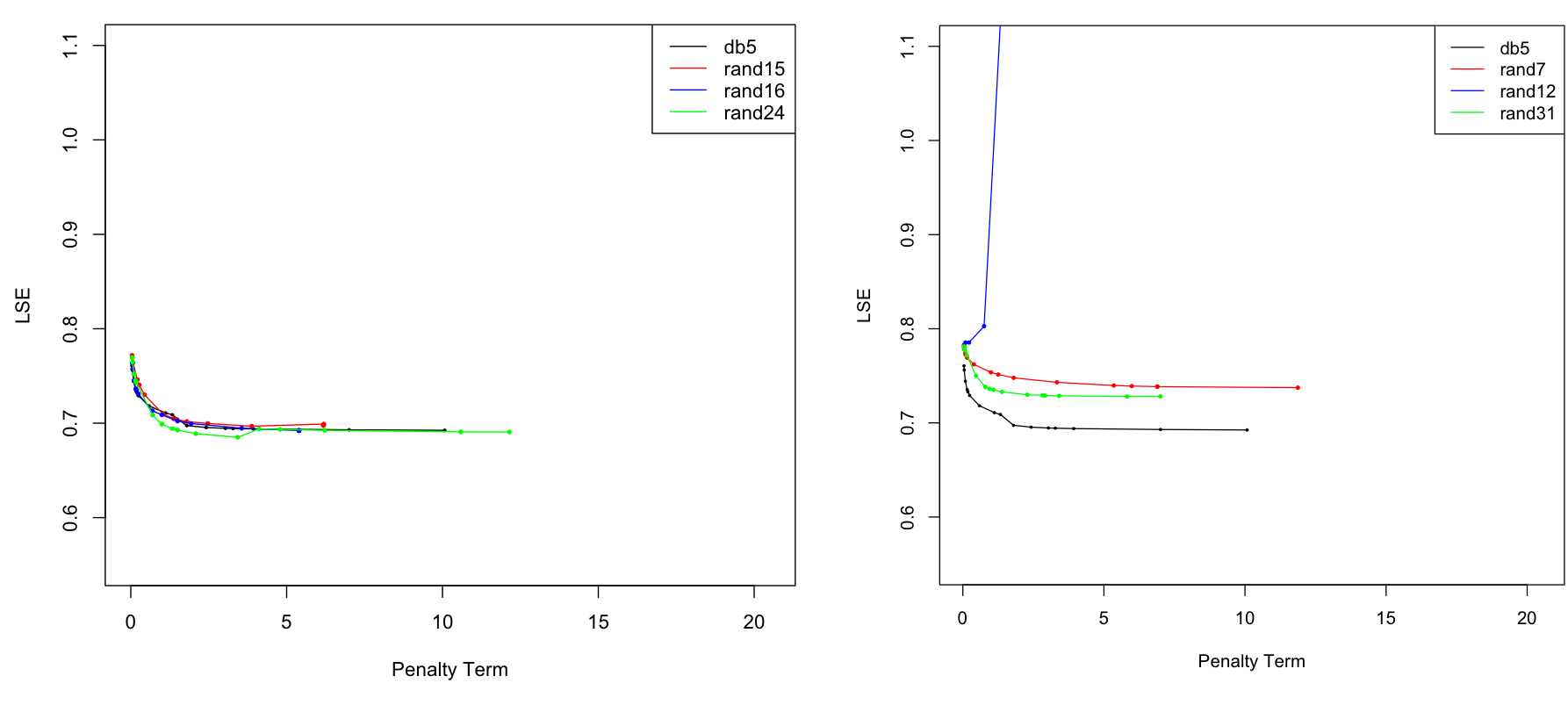


**Figure 6 (precursor).** Multivariate models of node MSE:minMSE incorporating GRNmap inputs, outputs, and network graph statistics as predictors were created for db1-db6. Model coefficients and goodness-of-fit metrics are provided, with bracketed values indicated models that fell marginally short of the p<0.05 threshold for significance.

**II. Comparison of Six Database-Derived Networks to 30 Db5-Derived Networks**

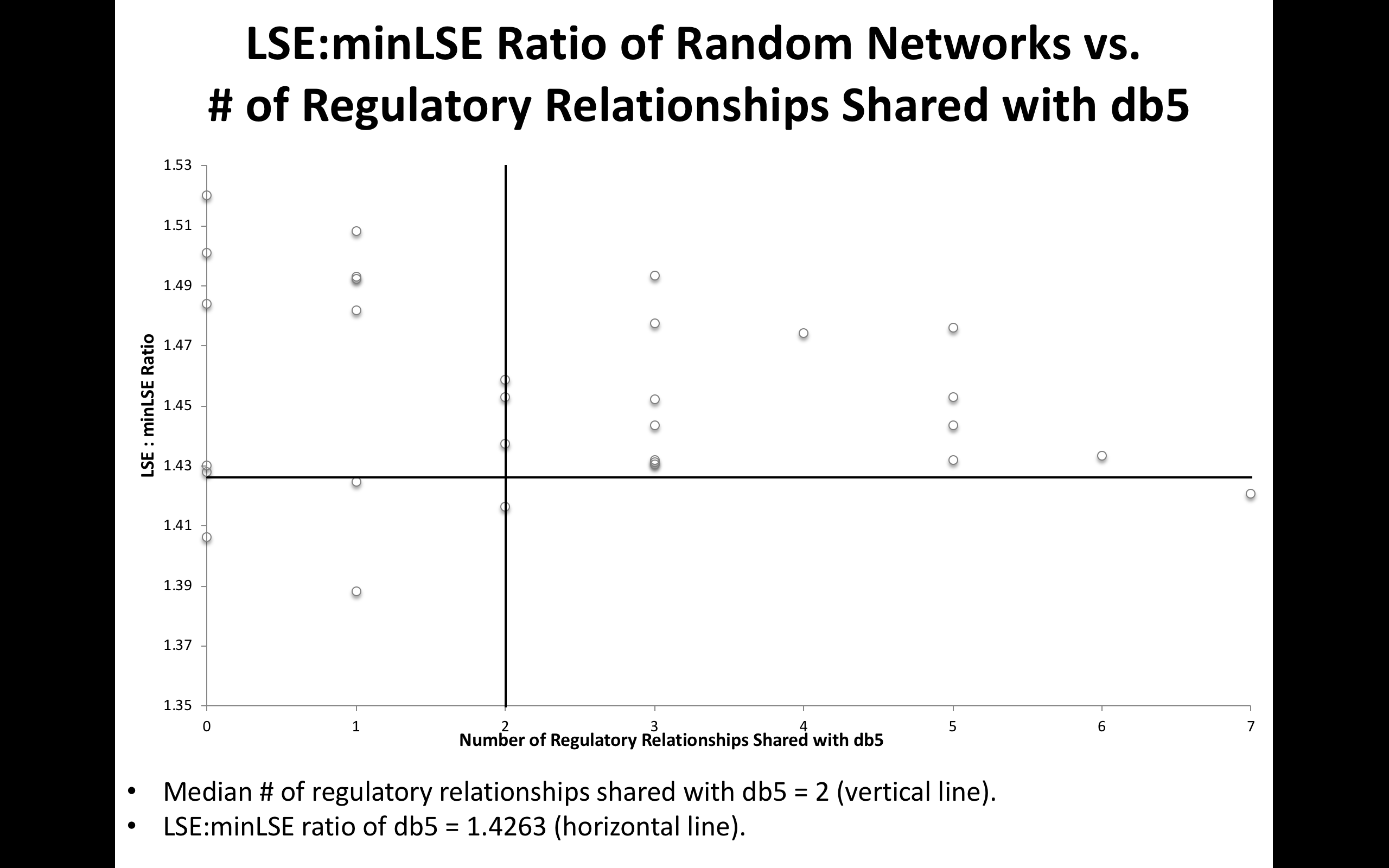


**Figure 8 (precursor)**. The weighted networks of db5 were visualized and contrasted to the three best (top row) and three worst (bottom row) performing db5-derived random networks using GRNsight.

****

**Figure 9 (precursor).** L-curve comparisons for db5 vs. the three best (left) and three worst (right) random networks.

* Common edges between db5 & 30 random networks.
  + 3 Best
    - rand15 (LSE:minLSE = 1.4063)
      * 0 shared edges.
    - rand16 (LSE:minLSE = 1.4165))
      * 2 shared activation relationships.
      * 1 shared relationship w/ opposite weights.
    - **rand24 (LSE:minLSE = 1.3880)**
      * **1 shared activation relationship**
      * **1 shared relationship w/ opposite weights.**
  + 3 Most Similar (db5 Lse:minLSE = 1.4263)
    - rand2 (LSE:minLSE = 1.4278)
      * 4 shared relationships w/ opposite weights.
    - rand3 (LSE:minLSE = 1.4302)
      * 1 shared relationship w/ opposite weights.
    - rand9 (LSE:minLSE = 1.4247)
      * 1 shared activation relationship.
  + 3 Worst
    - **rand7 (LSE:minLSE = 1.5202)**
      * **1 shared relationship w/ opposite weights.**
    - rand12 (LSE:minLSE = 1.5080)
      * 1 shared activation relationship.
      * 1 shared relationship w/ opposite weights.
    - rand31 (LSE:minLSE = 1.5009)
      * 1 shared relationship w/ opposite weights.
  + Other Notables
    - rand23 (LSE:minLSE = 1.4332)
      * 3 shared activation relationships.
      * 3 shared repression relationships.
      * 1 shared relationship w/ opposite weights.
    - rand26 (LSE:minLSE = 1.4206)
      * 3 shared activation relationships.
      * 4 shared repression relationships.

****

**Figure 7 (precursor).** LSE:minLSE Ratio of Random Networks vs. # of Regulatory Relationships Shared with db5. The vertical line indicated the median number of regulatory relationships shared with db5 (2). The horizontal line marks the LSE:minLSE ratio of db5 (1.4263).

**Discussion:**

* Restatement of the results and interpretation
  + (This is currently interspersed into the results section but will need to be relocated to the discussion section)
* Biological significance of transcription factors in our database-derived networks
  + HMO1
    - Interacts with transcription factors controlling transcription, ribosome biogenesis, and various stress response pathways (Berger et al., 2007)
    - TORC1-dependent regulation of ribosomal proteins and ribosome biogenesis is dependent on Hmo1 (Berger et al., 2007) (Xiao and Grove, 2009)
      * Ribosome biogenesis initial molecular response to cold shock (Aguilera et al., 2007) (Al-Fageeh and Smales, 2006)
    - Recent evidence has suggested HMO1 may function as a linker histone necessary for chromatin compaction. When upregulated, such compaction may confer resistance to environmental stress (Panday & Grove, 2016)
  + MSN2
    - Inducing the environmental stress response (Gasch et al., 2000)
  + ACE2, MCM1, SWI4, SWI5
    - Involved in control of the cell cycle in *S. cerevisiae*. Specifically, these four factors have been shown to induce expression of M/G1 transition and G1 proteins (Simon et al., 2001)
  + STB1, YHP1
    - Overexpression of either STB1 or YHP1 arrests the cell cycle at the M/G1 interval in *S. cerevisiae* (Youn et al., 2017)
  + Others not previously implicated in the cold-shock response in yeast, e.g. GLN3, HAP4, and many more!
* Interpretation of recurring motifs in the 5 database-derived networks, and comparison to motifs in the 30 db5-derived random networks
  + Characterization of motifs important to biological GRNs (Lee et al., 2002)
  + Autoregulatory negative feedback loops in gene regulatory networks provide stability in the face of environmental perturbation (Becskei and Serrano, 2000)
  + Regulator chains represent temporal circuits (Simon et al., 2001)
* Limitations
  + Microarray noise resulting from impaired translation during cold shock (Rodriguez-Vargas et al. 2002)
  + Need to consider this further.

**References:**

Aguilera, J., Randez-Gil, F., and Prieto, J.A. (2007). Cold response in *Saccharomyces cerevisiae*:

new functions for old mechanisms. *FEMS Microbiology Reviews*, *31*(3), 327-341. DOI: 10.1111/j.1574-6976.2007.00066.x.

Al-Fageeh, M.B., and Smales, C.M. (2006). Control and regulation of the cellular responses to cold

shock: the responses in yeast and mammalian systems. *Biochemical Journal*, *397*(2), 247-

259. DOI: 10.1042/BJ20060166.

Becskei, A., and Serrano, L. (2000). Engineering stability in gene networks by autoregulation.

*Nature*, *405*(6786), 590-593. DOI: 10.1038/35014651.

Berger, A.B., Decourty, L., Badis, G., Nehrbass, U., Jacquier, A., and Gadal, O. (2007). Hmo1 is

Required for TOR-Dependent Regulation of Ribosomal Protein Gene Transcription. *Molecular and Cellular Biology*, *27*(22), 8015-8026. DOI: 10.1128/MCB.01102-07.

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., ... and Fisk, D.G. (2011). Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Research*, *40*(D1), D700-D705. DOI:10.1093/nar/gkr1029.

Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., ... and Pelechano, V.

(2016). A global genetic interaction network maps a wiring diagram of cellular

function. *Science*, *353*(6306), aaf1420. DOI: 10.1126/science.aaf1420.

Dahlquist, K.D., Dionisio, J.D.N., Fitzpatrick, B.G., Anguiano, N.A., Varshneya, A., Southwick, B.J.,

and Samdarshi, M. (2016). GRNsight: a web application and service for visualizing models of small-to medium-scale gene regulatory networks. *PeerJ Computer Science*, *2*, e85. DOI: 10.7717/peerj-cs.85.

Dahlquist, K.D., Fitzpatrick, B.G., Camacho, E.T., Entzminger, S.D., and Wanner, N.C. (2015).

Parameter Estimation for Gene Regulatory Networks from Microarray Data: Cold Shock Response in *Saccharomyces cerevisiae*. *Bulletin of Mathematical Biology*, *77*(8), 1457-1492. DOI: 10.1007/s11538-015-0092-6.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., ... and Brown, P.O.

(2000). Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. *Molecular Biology of the Cell*, *11*(12), 4241-4257. DOI: 10.1091/mbc.11.12.4241.

Hampsey, M. (1997). A Review of Phenotypes in *Saccharomyces cerevisiae*. *Yeast*, *13*(12), 1099-

1133. DOI: 10.1002/(SICI)1097-0061(19970930)13:12<1099::AID-YEA177>3.0.CO;2-7.

Hansen, P.C., and O’Leary, D.P. (1993). The Use of the L-Curve in the Regularization of Discrete Ill-

Posed Problems. *SIAM Journal on Scientific Computing*, *14*(6), 1487-1503. DOI: 10.1137/0914086.

Ingolia, T.D., Slater, M.R., & Craig, E.A. (1982). *Saccharomyces cerevisiae* contains a complex multigene

family related to the major heat shock-inducible gene of *Drosophila*. *Molecular and Cellular Biology*, *2*(11), 1388-1398. DOI: 10.1128/MCB.2.11.1388.

Kandror, O., Bretschneider, N., Kreydin, E., Cavalieri, D., and Goldberg, A.L. (2004). Yeast Adapt to

Near-Freezing Temperatures by STRE/Msn2, 4-Dependent Induction of Trehalose Synthesis and Certain Molecular Chaperones. *Molecular Cell*, *13*(6), 771-781. DOI: 10.1016/S1097-2765(04)00148-0.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., ... and Zeitlinger, J.

(2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, *298*(5594), 799-804. DOI: 10.1126/science.1075090.

Morano, K.A., Grant, C.M., and Moye-Rowley, W.S. (2012). The Response to Heat Shock and Oxidative

Stress in *Saccharomyces cerevisiae*. *Genetics*, *190*(4), 1157-1195. DOI: 10.1534/genetics.111.128033.

Neymotin, B., Athanasiadou, R., and Gresham, D. (2014). Determination of in vivo RNA kinetics

using RATE-seq. *RNA*, *20*(10), 1645-1652. DOI: 10.1261/rna.045104.114.

Panday, A. and Grove, A. (2016). The high mobility group protein HMO1 functions as a linker histone in yeast. *Epigenetics & Chromatin*, *9*(1), 13. DOI: 10.1186/s13072-016-0062-8.

Rodriguez-Vargas, S., Estruch, F., and Randez-Gil, F. (2002). Gene Expression Analysis of Cold

and Freeze Stress in Baker's Yeast. *Applied and Environmental Microbiology*, *68*(6), 3024-3030. DOI: 10.1128/AEM.68.6.3024–3030.2002.

Sahara, T., Goda, T., and Ohgiya, S. (2002). Comprehensive Expression Analysis of Time-

dependent Genetic Responses in Yeast Cells to Low Temperature. *Journal of Biological Chemistry*, *277*(51), 50015-50021. DOI: 10.1074/jbc.M209258200.

Schade, B., Jansen, G., Whiteway, M., Entian, K.D., and Thomas, D.Y. (2004). Cold Adaptation in

Budding Yeast. *Molecular Biology of the Cell*, *15*(12), 5492-5502. DOI: 10.1091/mbc.E04–03–0167.

Simon, I., Barnett, J., Hannett, N., Harbison, C.T., Rinaldi, N.J., Volkert, T.L., ... and Young, R.A.

(2001). Serial Regulation of Transcriptional Regulators in the Yeast Cell Cycle. *Cell*, *106*(6), 697-708. DOI: 10.1016/S0092-8674(01)00494-9.

Teixeira, M.C., Monteiro, P.T., Palma, M., Costa, C., Godinho, C.P., Pais, P., ... and Sá-Correia, I.

(2017). YEASTRACT: an upgraded database for the analysis of transcription regulatory networks in Saccharomyces cerevisiae. *Nucleic Acids Research*, *46*(D1), D348-D353. DOI: 10.1093/nar/gkx842.

Verghese, J., Abrams, J., Wang, Y., & Morano, K. A. (2012). Biology of the Heat Shock Response and

Protein Chaperones: Budding Yeast (*Saccharomyces cerevisiae*) as a Model System. *Microbiology and Molecular Biology Reviews*, *76*(2), 115-158. DOI: 10.1128/MMBR.05018-11.

Xiao, L., and Grove, A. (2009). Coordination of Ribosomal Protein and Ribosomal RNA Gene

Expression in Response to TOR Signaling. *Current Genomics*, *10*(3), 198-205. DOI: 10.2174/138920209788185261.

Youn, J. Y., Friesen, H., Ba, A. N. N., Liang, W., Messier, V., Cox, M. J., ... and Andrews, B. (2017). Functional Analysis of Kinases and Transcription Factors in *Saccharomyces cerevisiae* Using an Integrated Overexpression Library. *G3: Genes, Genomes, Genetics*, *7*(3), 911-921. DOI: 10.1534/g3.116.038471.