



# Updated regulation curation model at the *Saccharomyces* Genome Database

Stacia R. Engel  
Stanford University

# Yeast is valuable model organism for study of different types of regulation

- metal ion homeostasis by controlling metabolism genes
- cellular memory of prior nutrient environments
  - lifespan extending effects of calorie restriction
  - downregulation of nutrient signaling
- S-adenosyl-L-methionine (AdoMet)-dependent methylation disorders
  - hyperhomocysteinemia

# what do people want to know about regulation?

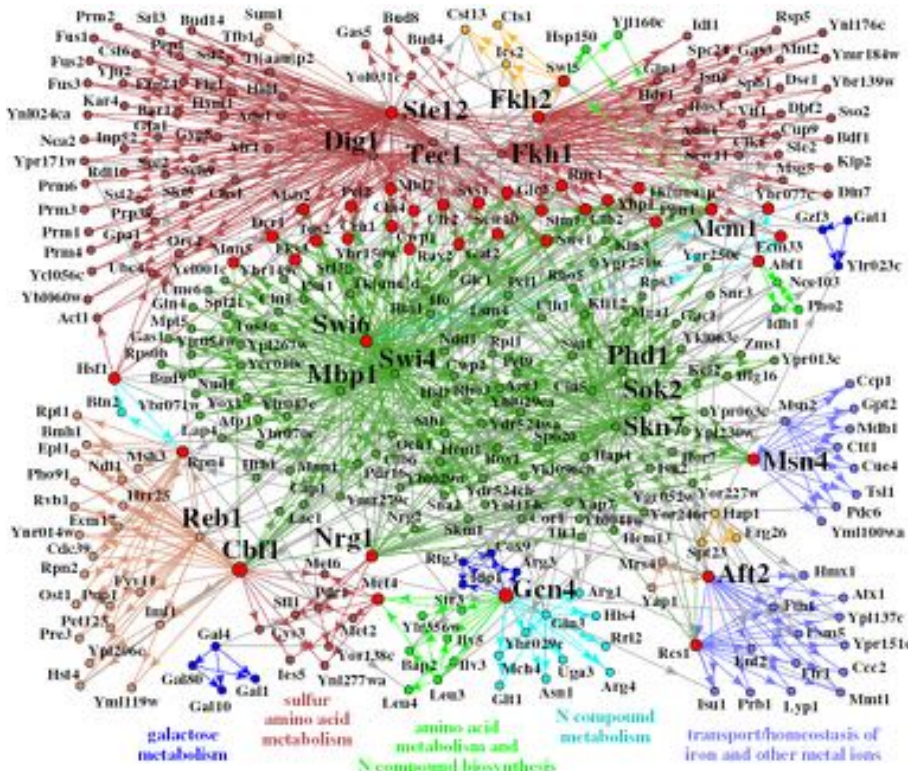
- what it is
- what it does
- how that changes in certain situations
- which agents make that happen

# Regulatory Relationships

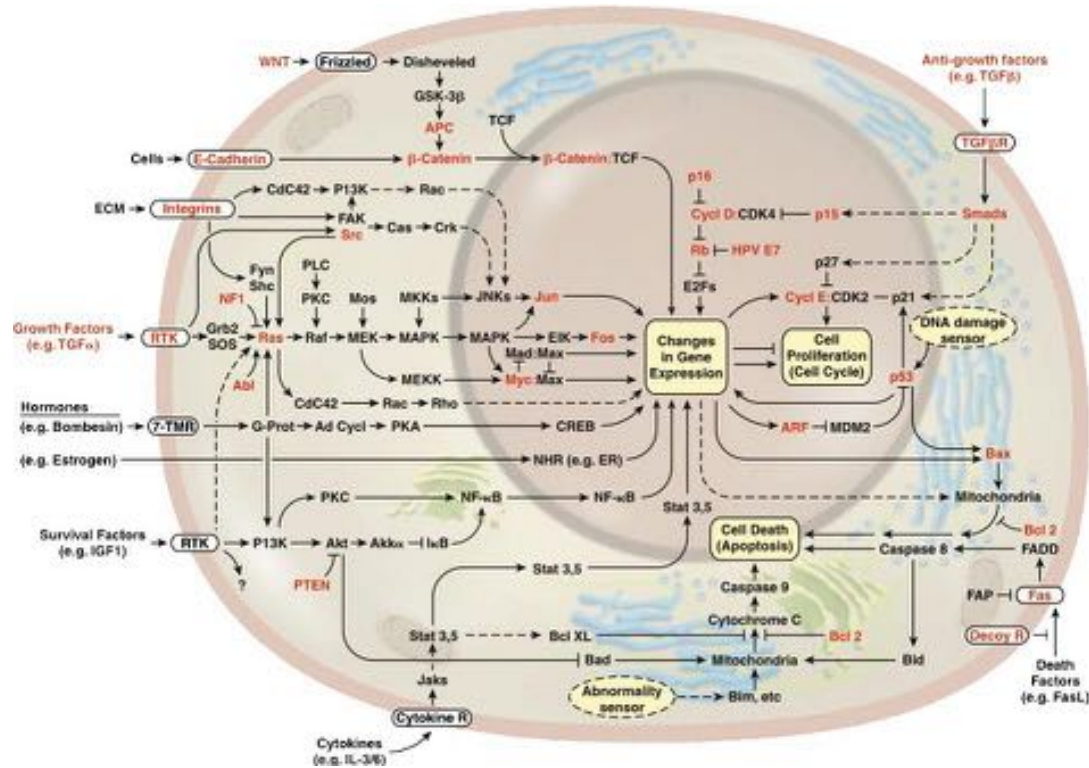
- exist between genes, proteins
  - one entity controls expression or activity of the other
- occur at many levels
  - transcription, RNA stability, RNA activity
  - translation, protein stability, protein activity



Regulatory Network = sum of regulatory relationships



# Goal: model regulatory network of a eukaryotic cell



# Recent update of data model

- regulation by chromatin modifiers
  - *in addition to transcription factors*
- address regulation at protein level
  - *in addition to transcription*
- post-translational modifications
- expert curation of focused, small-scale studies
  - *after previously seeding with large-scale screens*



# Curation Strategy

- survey yeast literature to find, record regulatory relationships
- leverage existing tags to prioritize papers
  - added during literature triage
- leverage existing annotations to prioritize genes
  - Gene Ontology
    - ‘cellular protein modification process’ (GO:0006464)
    - ‘regulation of gene expression’ (GO:0010468)
  - Post-translational modifications





# Record results from same experiment in different areas...

- functional data
  - GO annotations
- protein modifications
  - PTM annotations
- regulation data
  - regulatory relationships



*...to present as complete of a picture of biology as possible!*

# How are the data recorded?

- regulator-target pair
- type of regulation
- role of regulator
  - *broadly classified based on GO annotations*
- direction of regulation
- type of experiment
- publication
- cellular situation under which regulation occurs



# Updated curation data model

Data type	Description	Value	
Regulator	Gene identifier	Gene name or systematic name	
Regulator type	Controlled vocabulary	Transcription factor Chromatin modifier Protein modifier	<i>Future:</i> <i>RNA-binding protein</i> <i>RNA modifier</i>
Target	Gene identifier	Gene name or systematic name	
Direction	Controlled vocabulary	Positive, Negative, [null]	
Regulation of	Controlled vocabulary	Transcription Protein activity	<i>Future:</i> <i>RNA stability</i> <i>Protein stability</i>
Happens during	Controlled vocabulary; Situation under which the regulation occurs	Defined subset of Gene Ontology terms: descendent terms of Biological Process 'cellular response to stimulus' GO:0051716 and 'cell cycle phase' GO:0022403	
Annotation method	Controlled vocabulary	Manually curated High-throughput	
Evidence	Type of experiment used to test for and/or demonstrate the regulatory relationship	Evidence & Conclusion Ontology (ECO) term	
Strain background	Controlled vocabulary	Strain name	
Reference	Publication in which the regulatory relationship is described	PubMed ID	

## YOX1 / YML027W Regulation

Regulation Help

YOX1 encodes a helix-turn-helix transcriptional repressor of the homeodomain family. Yox1p restricts transcription of genes containing early cell cycle boxes (ECBs) to M-G1 phase. Yox1p is an ancillary transcription factor that binds alongside Mcm1p on promoters. Examples include important cell cycle genes such as *SWI4* and *CLN3*, themselves rate-limiting for cell cycle progression, and essential for the control of subsequent waves of gene expression at G1-S. Yox1p binds upstream of genes involved in spindle function, DNA synthesis and repair, and also associates with a small portion of intergenic regions adjacent to tRNA genes. Removal of Yox1p from promoters allows gene expression at M-G1. YOX1 is under transcriptional regulation by SBF, and Mcm1p controls the expression of *SWI4* which encodes a component of SBF, thereby setting up a negative feedback loop. The transcription of the *CDC20* APC (anaphase promoting complex) activator is repressed by Yox1p and kept very low during S phase. Upon mitotic onset, Yox1p abundance declines and the *CDC20* gene is actively transcribed. Cdc28p negatively regulates YOX1 transcription, thereby keeping intracellular levels of Yox1p low, and promoting the transcription of *CDC20*. Yox1p is an unstable protein, so that suppression of its transcription results in rapid reduction of intracellular protein levels.

Acker J, et al. (2013) Yeast RNA polymerase III transcription factors and effectors. *Biochim Biophys Acta* 1829(3-4):283-95

PMID:23063749

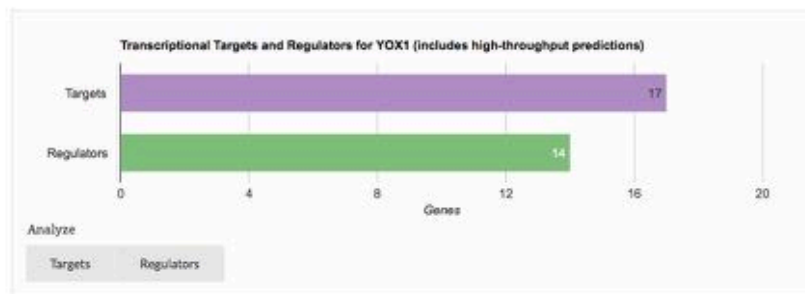
SGD Paper PubMed DOI full text of

Surana U, et al. (2012) Staging a recovery from mitotic arrest: Unusual ways of Cdk1. *Bioarchitecture* 2(2):33-37 PMID:22754627

SGD Paper PubMed PMC full text of

McInerney CJ (2011) Cell cycle regulated gene expression in yeasts. *Adv Genet* 73:51-85 PMID:21310204

SGD Paper PubMed DOI full text of



## Domains and Classification

8 entries for 8 domains

from InterPro  
([www.ebi.ac.uk/interpro](http://www.ebi.ac.uk/interpro))

Filter table

Protein Coordinates	Accession ID	Description	Source	No. of Genes with Domain
12-385	PTHR24326		PANTHER	4
12-385	PTHR24326-SF430		PANTHER	2
172-244	G3D5A:1.10.10.60		Gene3D	36
172-235	SSF46489	Homeobox domain-like	SUPERFAMILY	31
174-234	P550071	HOMEBOX_2; Homeobox domain	PROSITE	9
176-238	SM00389	HOX; Homeobox domain	SMART	9
177-233	PF00046	Homeobox; Homeobox domain	Pfam	7
209-232	P500027	HOMEBOX_1; Homeobox, conserved site	PROSITE	4

Showing 1 to 8 of 8 entries 10 records per page 1

Download (.txt)

## DNA Binding Site Motifs



as predicted by YeTFaSco  
([yetfasco.ccbr.utoronto.ca](http://yetfasco.ccbr.utoronto.ca))

YOX1 /  
YML027W

Regulation  
Overview

Domains and  
Classification

DNA Binding  
Site Motifs

Targets

Shared GO  
Processes  
Among  
Targets

Regulators

Regulation  
Network

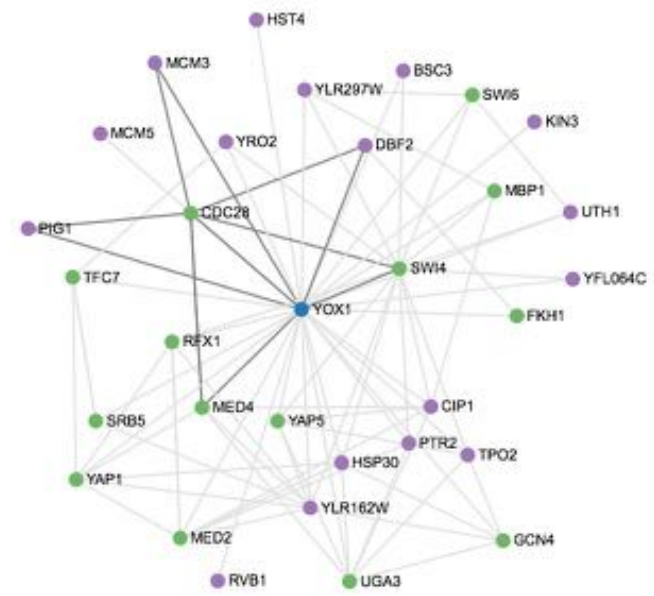
High-throughput 15 entries for 14 genes

Filter table							
Regulator	Regulator Type	Direction	Regulation Of	Happens During	Evidence	Strain Background	Reference
CDC28	protein modifier		protein activity		quantitative mass spectrometry evidence	S288C	Holt LJ, et al. (2009) PMID:19779198
FKH1	transcription factor		transcription	cellular response to heat	chromatin immunoprecipitation-chip evidence	S288C	Venters BJ, et al. (2011) PMID:21321985
GCN4	transcription factor		transcription	cellular response to boron-containing substance levels	microarray RNA expression level evidence	S288C	Ulusik I, et al. (2011) PMID:22134689
MBP1	transcription factor		transcription		chromatin immunoprecipitation-chip evidence	W303	MacIsaac KD, et al. (2006) PMID:16322208
MED2	transcription factor		transcription	cellular response to heat	chromatin immunoprecipitation-chip evidence	S288C	Venters BJ, et al. (2011) PMID:21321985

## Regulation Network

REGULATOR TARGET FOCUS

SGD 2018-04-02



Maximum Number of Nodes  
50 150

Download (.png)

gene lists in each table can be ported  
to other analysis tools within SGD:

GO Term Finder, SPELL, YeastMine

# Future Directions

- regulatory complexes
  - *in addition to single protein regulators*
- other types of regulators
  - *RNA-binding proteins, RNA modifiers, other protein-binding entities*
- other types of regulation
  - *RNA stability, protein stability*



## Database update

# Updated regulation curation model at the *Saccharomyces* Genome Database

Stacia R. Engel, Marek S. Skrzypek, Sage T. Hellerstedt, Edith D. Wong, Robert S. Nash, Shuai Weng, Gail Binkley, Travis K. Sheppard, Kalpana Karra and J. Michael Cherry\*

Department of Genetics, Stanford University, Stanford, CA 94305, USA

\*Corresponding author; Email: [cherry@stanford.edu](mailto:cherry@stanford.edu)

Citation details: Engel,S.R., Skrzypek,M.S., Hellerstedt,S.T. et al. Updated regulation curation model at the *Saccharomyces* Genome Database. *Database* (2018) Vol. 2018: article ID bay007; doi:10.1093/database/bay007



Marek Skrzypek



Sage Hellerstedt



Edith Wong



Rob Nash



Shuai Weng



Gail Binkley



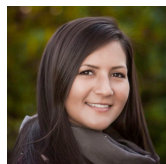
Travis Sheppard



Kalpana Karra



J. Michael Cherry



Suzi Aleksander



Barbara Dunn



Felix Gondwe



Kevin MacPherson



Stuart Miyasato



Patrick Ng



Ajay Prathivadhiba



Matt Simison

# *Thank you!*

- Funding: National Human Genome Research Institute at the United States National Institutes of Health [U41 HG001315].
- Please contact us at [sgd-helpdesk@lists.stanford.edu](mailto:sgd-helpdesk@lists.stanford.edu).