Computational reconstruction of the Transcriptional Regulatory Modules (TRMs) in yeast

Wei-Sheng Wu

Yeast Bioinformatician

Associate Professor

Department of Electrical Engineering (EE)

National Cheng Kung University

Taiwan

Visiting Scholar

Department of Ecology and Evolution (EE)

University of Chicago

USA

Where is Taiwan?

http://www.taroko.gov.tw/English/?mm=2&sm=2&page=1

https://travel.state.gov/content/passports/en/country/taiwan.html

Outlines

- Background of Transcriptional Regulatory Modules (TRMs)
- My algorithm for reconstructing TRMs (published in BMC Genomics)
  -using yeast heat shock response as an example
- My other expertise: databases and web tools development

Outlines

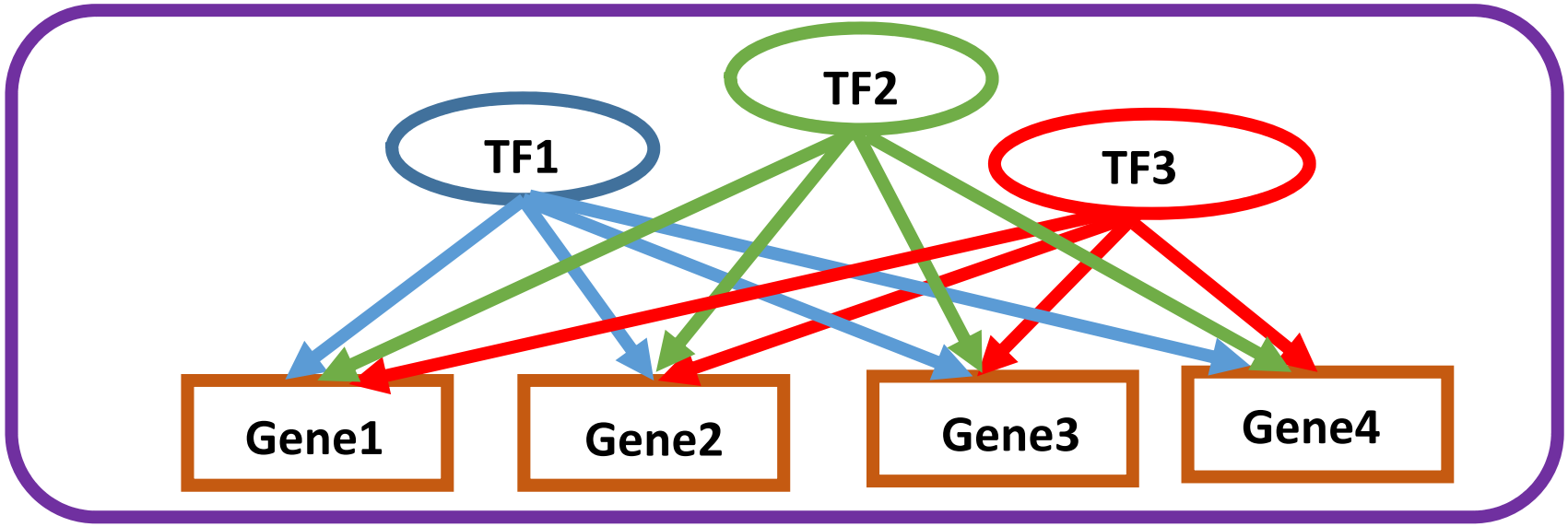
- Background of Transcriptional Regulatory Modules (TRMs)

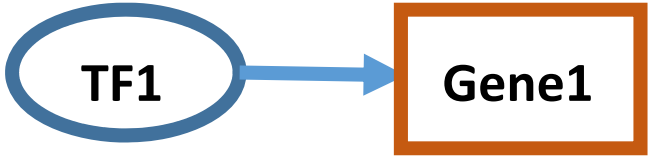
- My algorithm for reconstructing TRMs
  -using yeast heat shock response as an example

- My other expertise: databases and web tools development

What is a transcriptional regulatory module (TRM)?

- **TRM:** a set of genes that is co-regulated by the same set of TFs.
- **TRM** = {[TF1, TF2, TF3]→[Gene1, Gene2, Gene3, Gene4]}





Direct regulation: TF1 **binds & regulates** Gene1

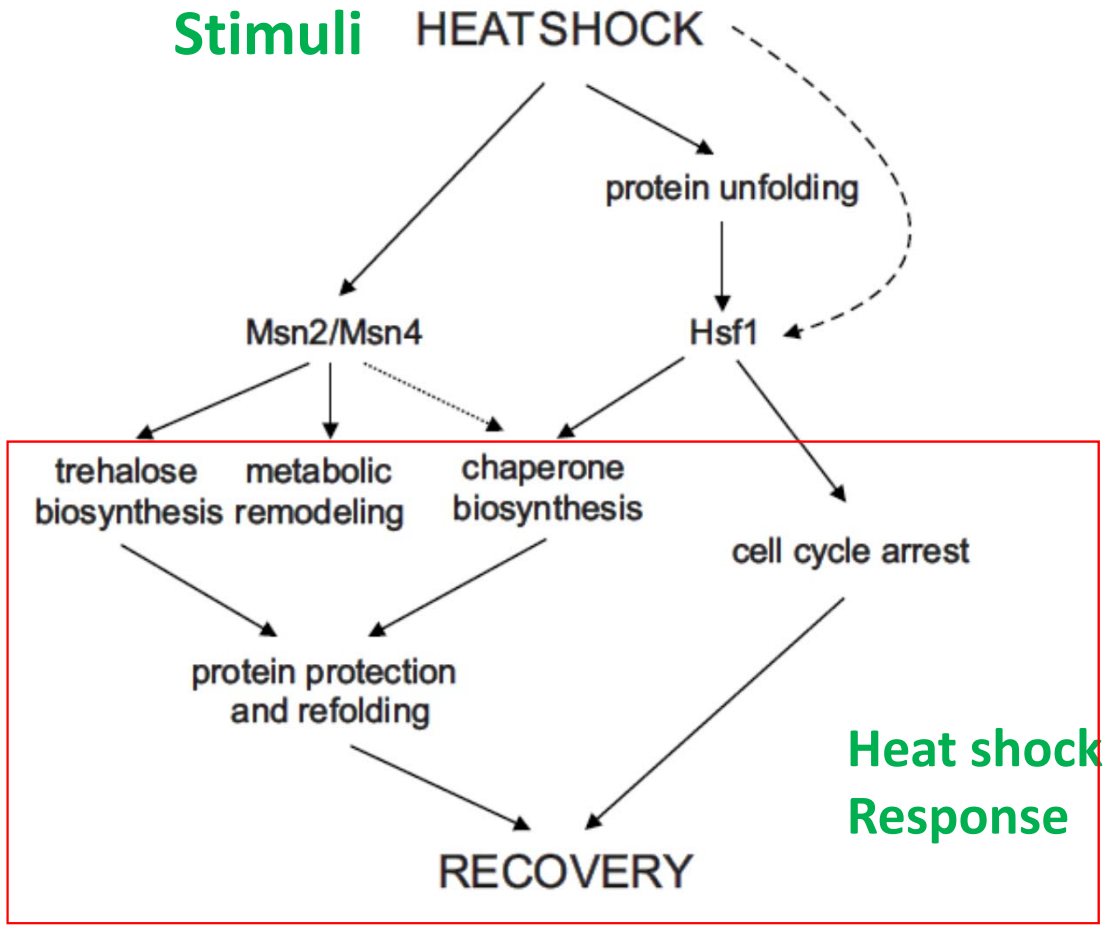
Why study TRM?
(using yeast heat shock response as an example)

- Single-cell organisms such as yeasts constantly face changing or even harsh environments such as high temperature that threaten their survival.

- By organizing the genome into transcriptional regulatory modules (TRMs), a yeast cell can coordinate the activities of many genes and carry out complex functions in response to high temperature.

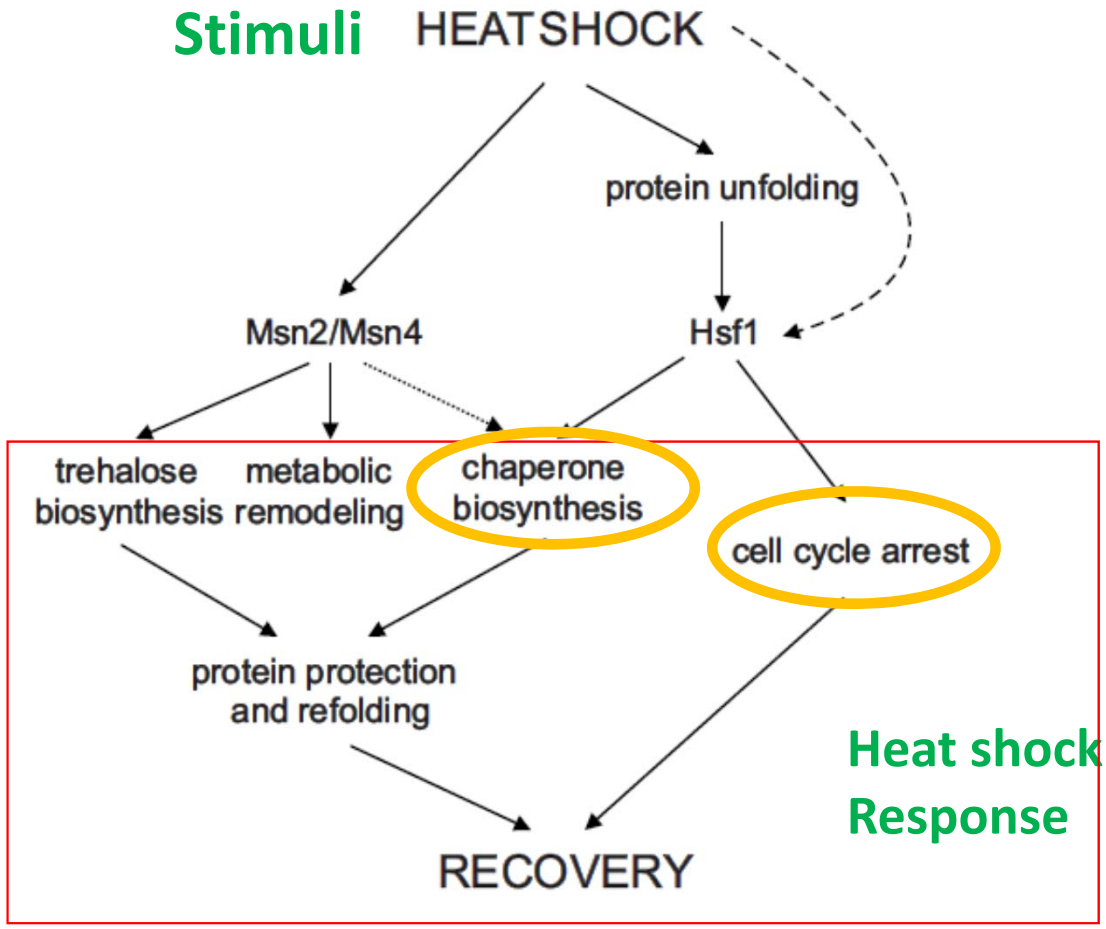
- Therefore, identifying TRMs of heat response is instrumental for understanding cellular responses to heat shock.

Current knowledge of yeast heat shock response (from numerous experimental studies for many years)



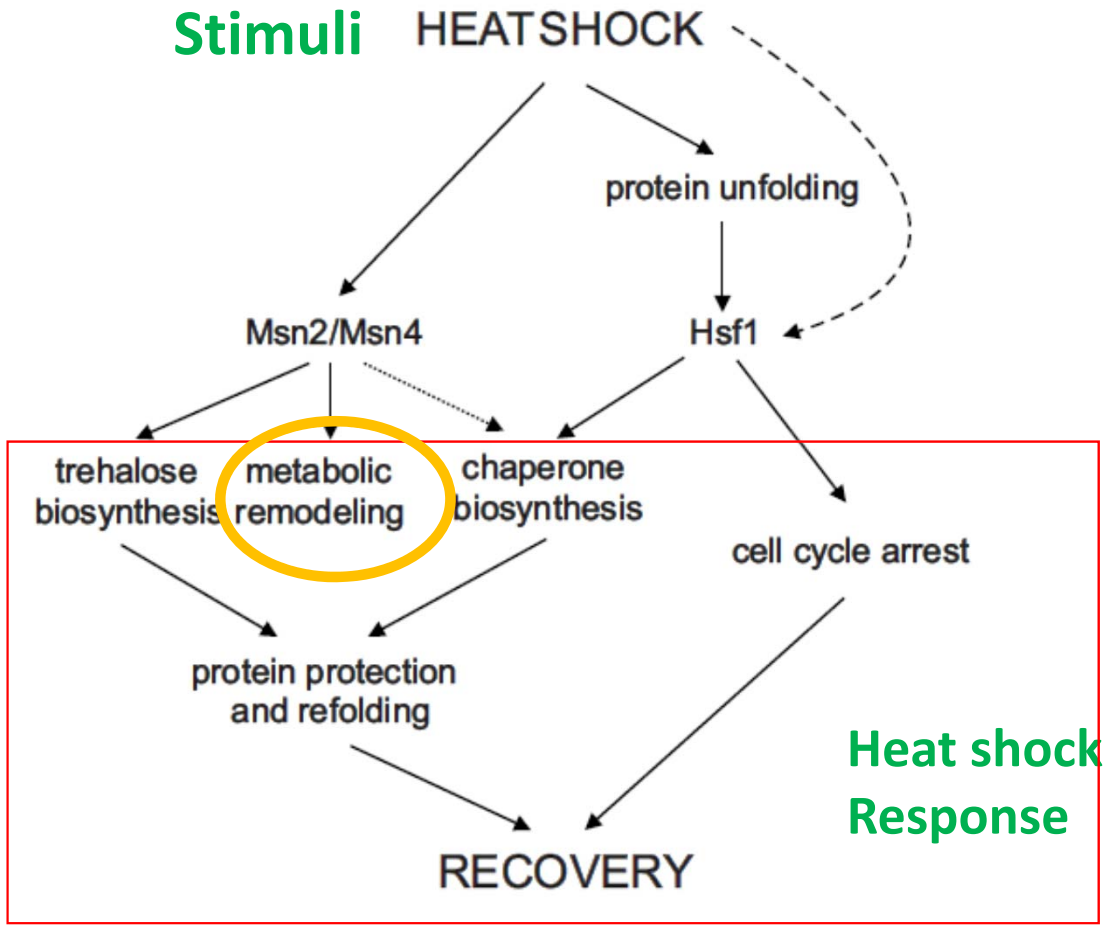
- Disruption of a large number of cellular assemblies and processes, an increased protein unfolding and aggregation, and membrane structure alterations are paramount in cells exposed to high temperature.
- Heat shock response serves to counteract these deleterious effects.
- Through it cells increase their thermotolerance or ability to withstand heat stress.

Many events occur in yeast cells during heat shock response (from numerous experimental studies for many years)



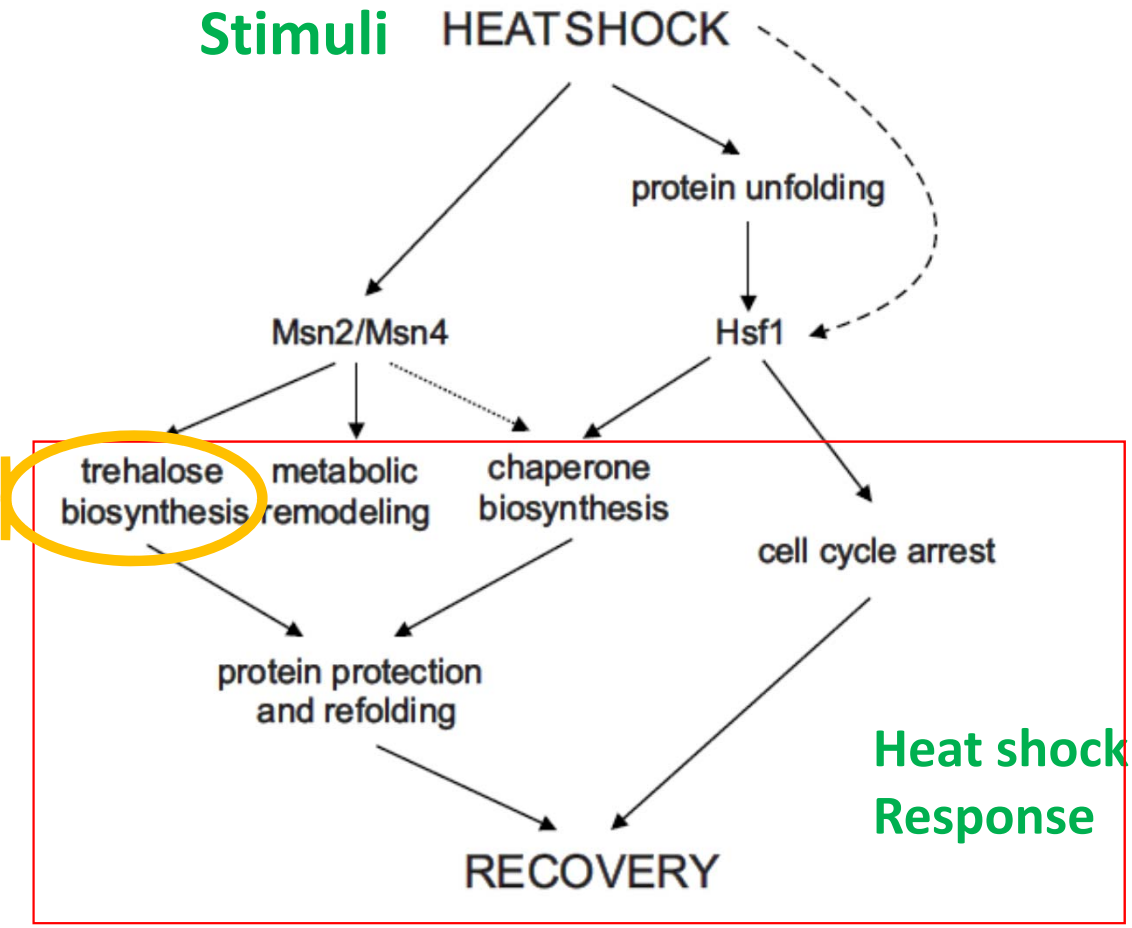
- Cell cycle transiently arrests during a heat shock stress.
- Heat shock proteins (HSPs) are rapidly synthesized.
- Many HSPs function as protein chaperones, so named because of their ability to bind to partially unfolded proteins to protect them from degradation or aggregation.

Many events occur in yeast cells during heat shock response (from numerous experimental studies for many years)



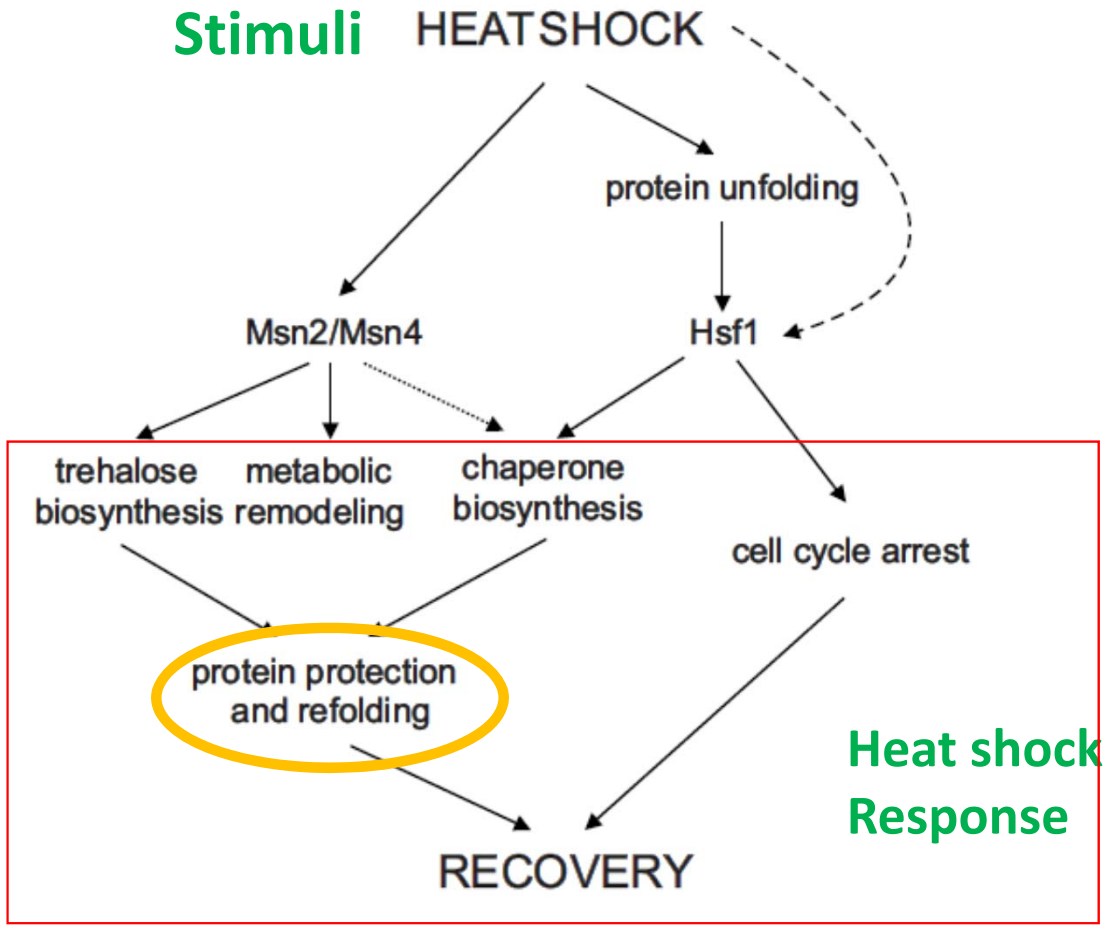
- Heat shock cells induce a variety of genes related to carbohydrate metabolism, fatty acid metabolism, respiration and others.

Many events occur in yeast cells during heat shock response (from numerous experimental studies for many years)



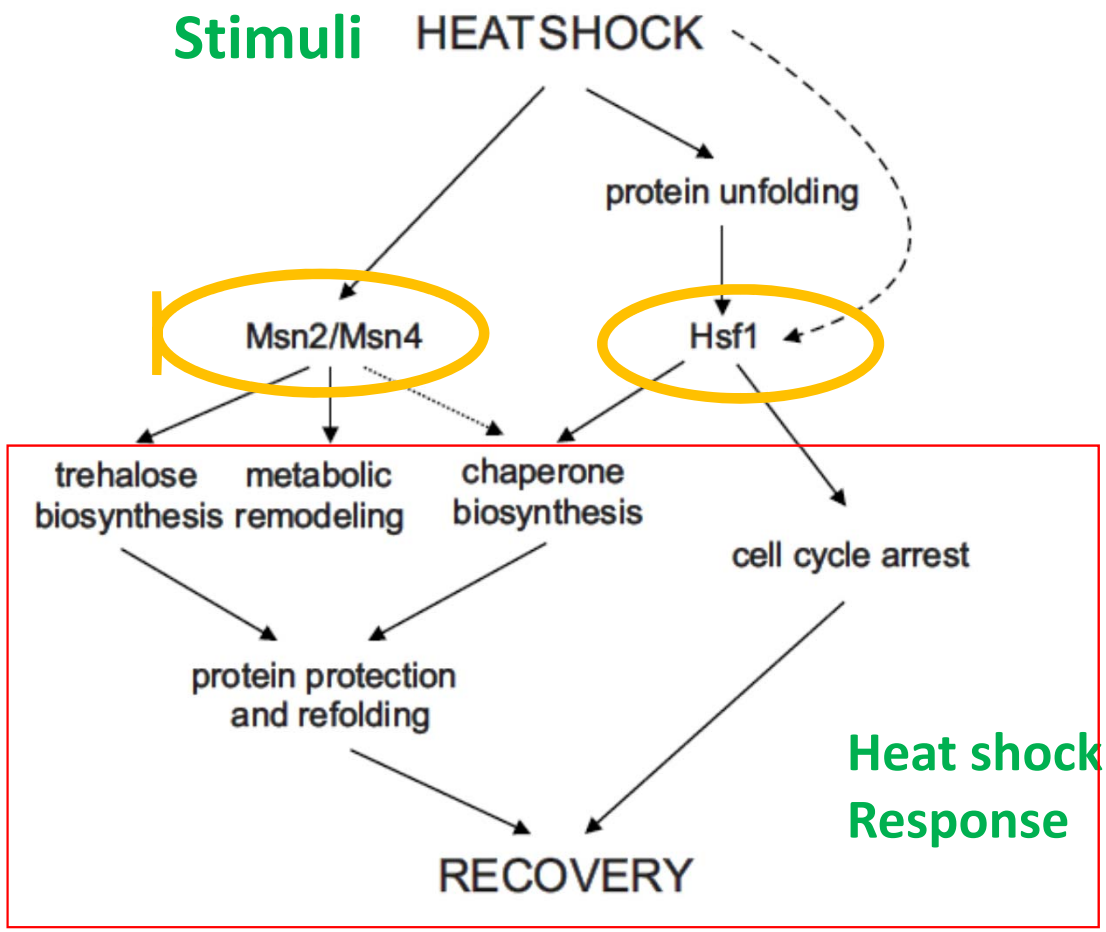
- Heat shock causes the extremely rapid accumulation of a large cytoplasmic pool of trehalose.
- Trehalose is one of the most effective substances known for preservation of membranous structures and enzyme activities during heating.

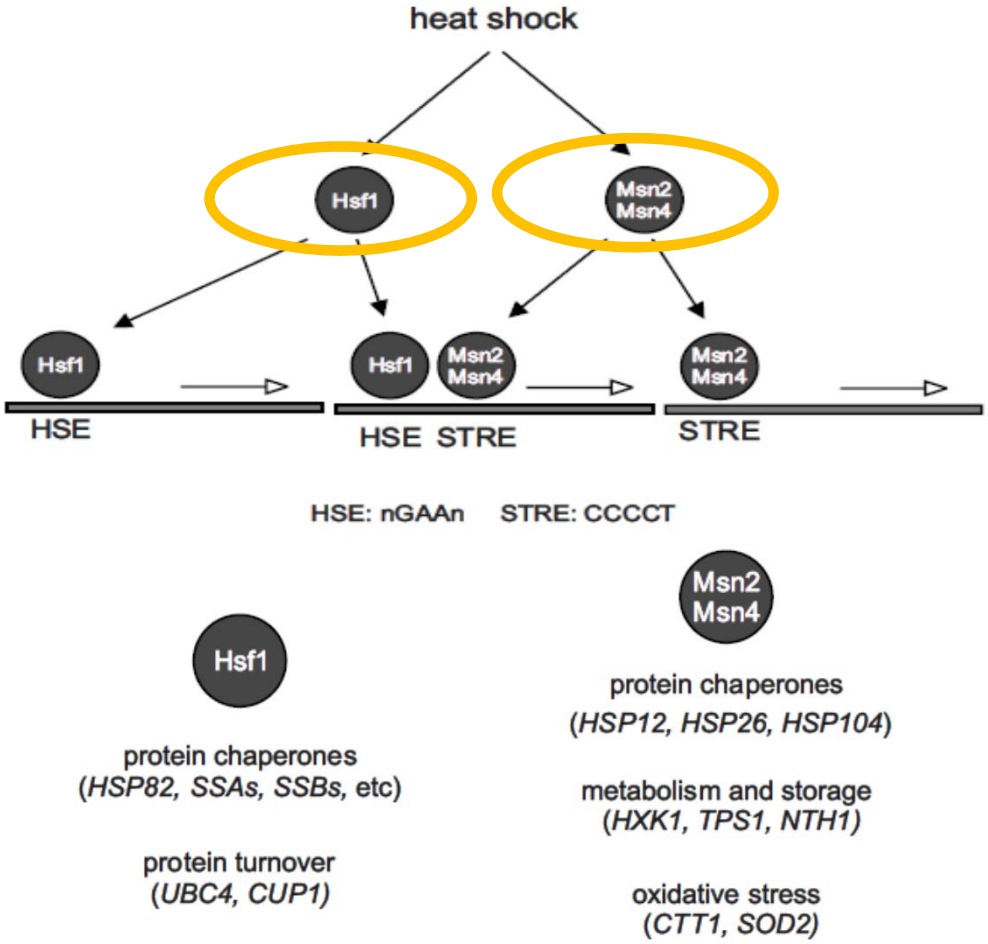
Many events occur in yeast cells during heat shock response (from numerous experimental studies for many years)



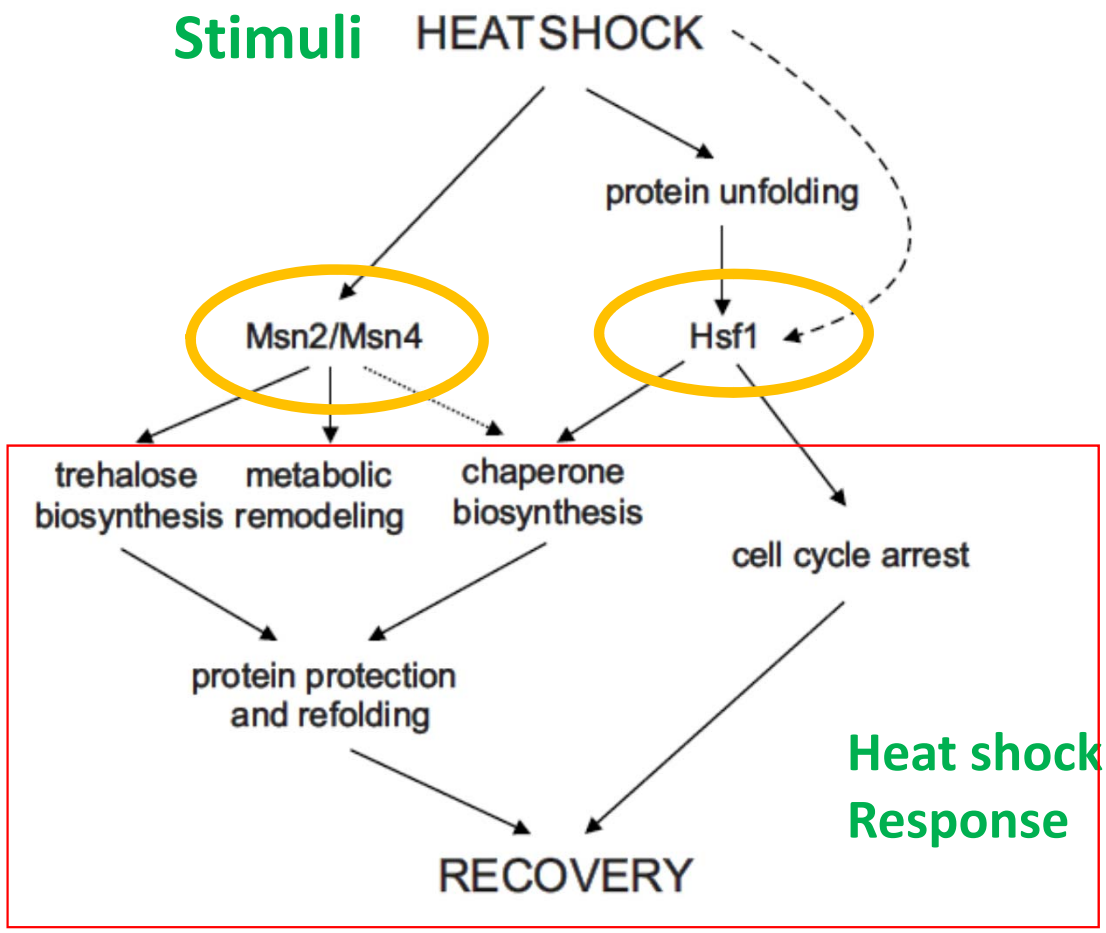
- Heat shock causes an increased protein unfolding and aggregation in a cell.
- Thus, many genes that are involved in protein (re)folding are induced to bind to partially unfolded proteins to protect them from degradation or aggregation.

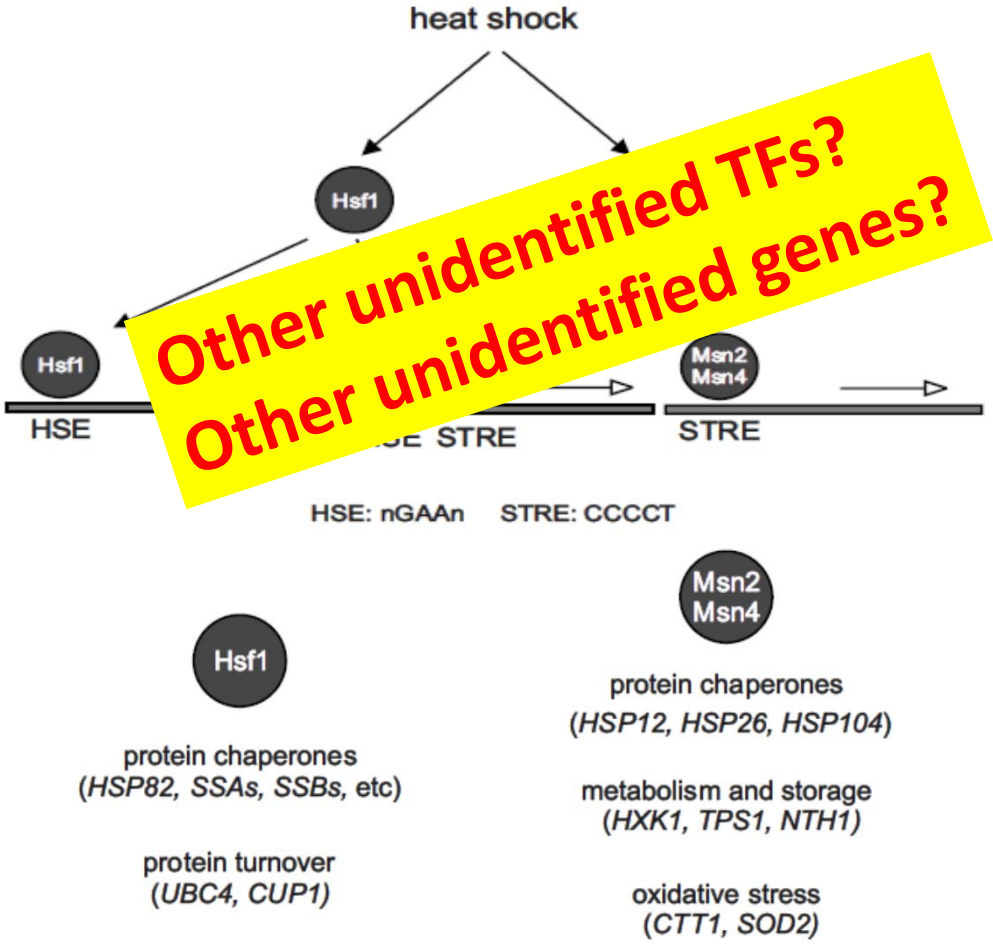
TFs and genes known to be involved in heat shock response (from numerous experimental studies for many years)



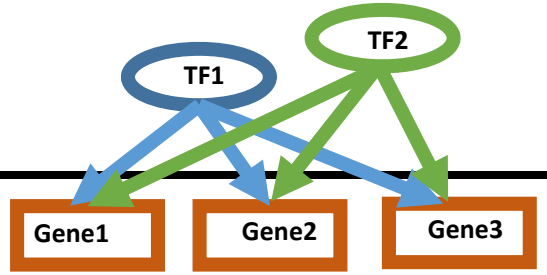


TFs and genes known to be involved in heat shock response (from numerous experimental studies for many years)

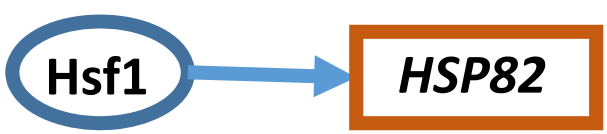




Computational reconstruction of TRMs



- Identifying direct regulations one by one using traditional experimental approaches is very time-consuming and labor-intensive to reconstruct TRMs.



Direct regulation: Hsf1 **binds & regulates** *HSP82*

-**Binding evidence** (DNA footprinting, Giardina et al. 1995)

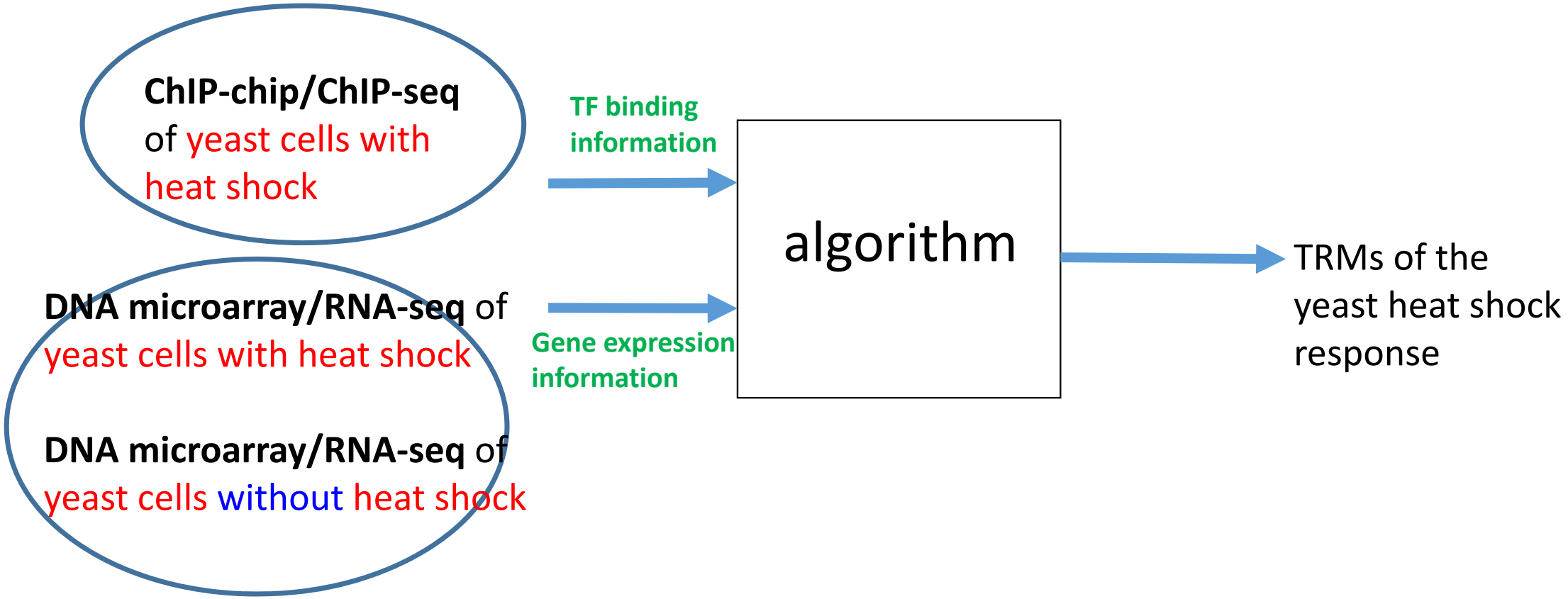
-**Regulation evidence** (Northern blotting WT vs. hsf1Δ; Eastmond and Nelson 2006)

- Since various kinds of high-throughput experimental technologies (e.g. **DNA microarray/RNA-seq**, **ChIP-chip/ChIP-seq**, …) are available now, it is possible to reconstruct TRMs using computational approaches.

Outlines

- Background of Transcriptional Regulatory Modules (TRMs)
- My algorithm for reconstructing TRMs
  -using yeast heat shock response as an example
- My other expertise: databases and web tools development

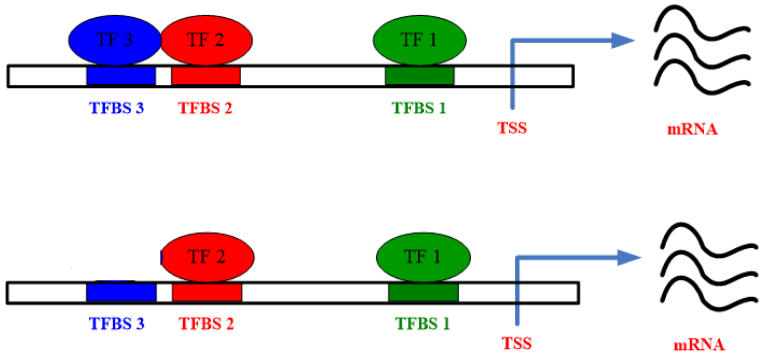
My algorithm for reconstructing TRMs of the yeast heat shock response



My algorithm for reconstructing TRMs of the yeast heat shock response

Step 1: For each gene in the yeast genome (~6500 genes), **identify its promoter-binding TFs** under heat shock

Step 2: For each gene in the yeast genome, **extract its regulatory TFs** from its promoter-binding TFs under heat shock



Step 3: **Identify heat-responsive genes** from the yeast genome

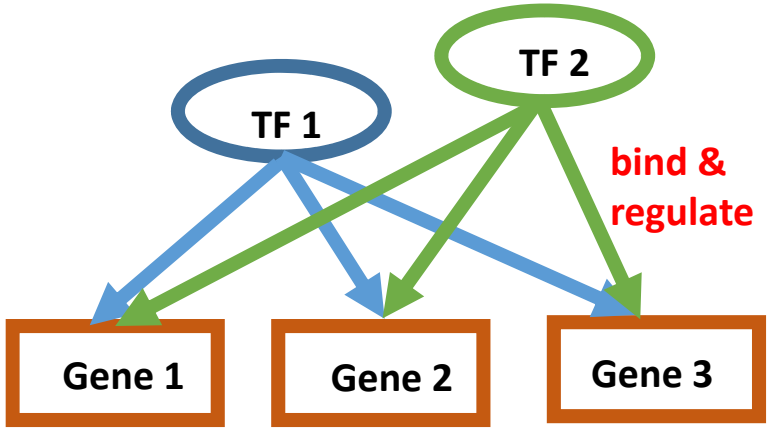
e.g. Gene1, Gene2, Gene3, Gene4, Gene5, Gene6,…

Step 4: **Identify heat-responsive TF sets**

e.g. **[TF1,TF2]**, [TF3,TF4,TF5],…

Step 5: **Reconstruct heat-responsive TRMs**

e.g. {**[TF1,TF2]**➔[Gene1, Gene2, Gene3]}

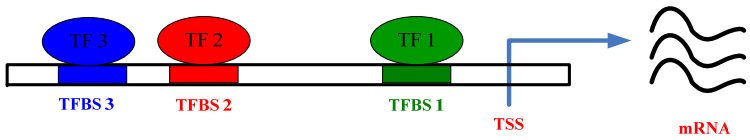


Step 1: For each gene in the yeast genome (~6500 genes), identify its promoter-binding TFs under heat shock

- Using **ChIP-chip/ChIP-seq** technology, researchers can know the **genome-wide binding target genes of a specific TF**.

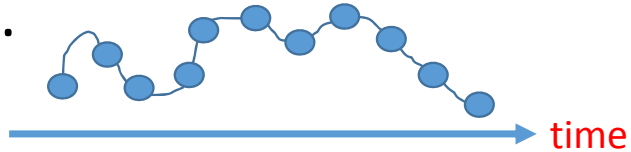
- For example, Harbison et al. (Nature 2004) used the ChIP-chip technology to determine the **genome-wide binding target genes of 200 yeast TFs in rich media conditions** and 7 TFs (Adr1, Gat1, Hsf1, Msn2, Skn7, Xbp1, Yap1) under heat shock.

- So, for each yeast gene, we know its promoter-binding TFs under heat shock from Harbison's ChIP-chip data.

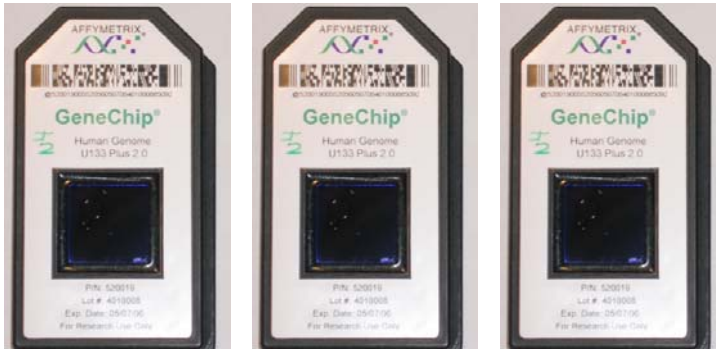


Step 2: For each gene in the yeast genome, extract its regulatory TFs from its promoter-binding TFs under heat shock

- Using **DNA microarray/RNA-seq** technology, researchers can have the **mRNA time profile of each gene** in the genome.

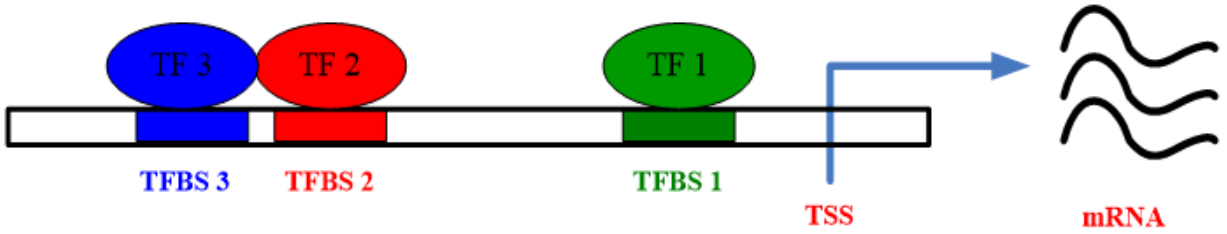


- For example, Causton et al. (MBC 2001) used DNA microarray technology to measure the **mRNA time profile of each gene in the yeast genome at 0 (before heat shock; 25˚C), 15, 30, 45, 60, 120 min (after heat shock; 37˚C)**.



0 min 15 min 30 min

Step 2: For each gene in the yeast genome, extract its regulatory TFs from its promoter-binding TFs under heat shock



$x_3[t]$: mRNA time profiles of TF3

$x_2[t]$: mRNA time profiles of TF2

$x_1[t]$: mRNA time profiles of TF1

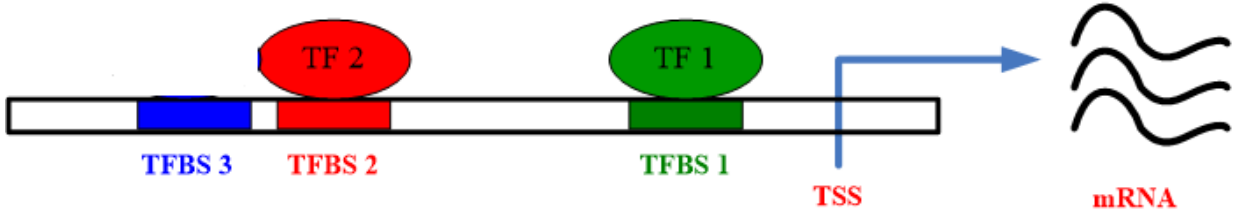
$y[t]$: mRNA time profiles of gene y

{TF1, TF2, TF3} are promoter-binding TFs of gene y

$$y[t+1] = \left(b_1 \cdot x_1[t] + b_2 \cdot x_2[t] + b_3 \cdot x_3[t] + k\right) - a \cdot y[t] + \varepsilon[t]$$

Discrete-time dynamic system model

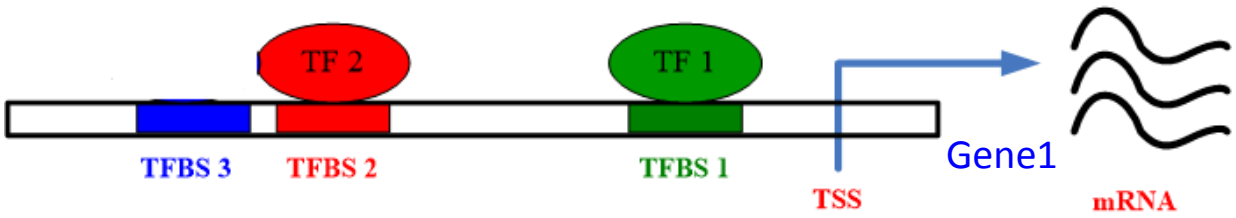
If maximum likelihood (ML) parameter estimation results show that $|b_1|$ & $|b_2|$ are >> 0, but not $|b_3|$



{TF1, TF2} are regulatory TFs of gene y

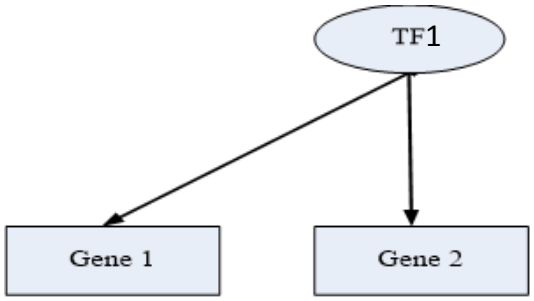
After running Step 2, we will know

- **The regulatory TFs of each gene** in the yeast genome under heat shock



	TF1	TF2	…	TFM
Gene1	v	v		
Gene2	v			v
…				
GeneK		v		

- **The regulatory target genes of each TF** under heat shock

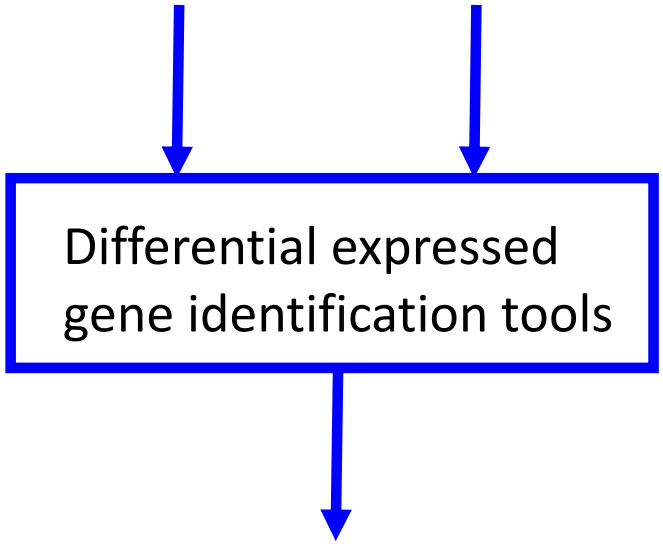


	TF1	TF2	…	TFM
Gene1	v	v		
Gene2	v			v
…				
GeneK		v		

Step 3: Identify heat-responsive genes from the yeast genome

DNA microarray/RNA-seq data of yeast cells without heat shock

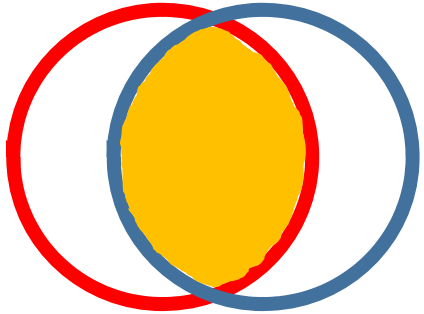
DNA microarray/RNA-seq data of yeast cells with heat shock



Step 4: Identify heat-responsive TF sets

- **A TF set (e.g. [TF1,TF2]) is said to be heat-responsive only if a significant portion of the regulatory target genes of the TF set is heat-responsive.**
- The number of TFs in a TF set could be one, two or more.

The regulatory targets of a TF set [TF1,TF2]



Heat-responsive genes identified in Step 3

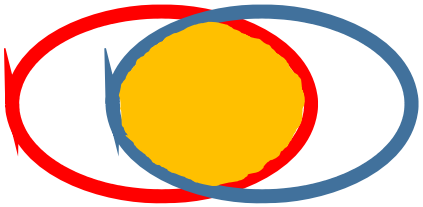
	TF1	TF2	…	TFM
Gene1	v	v		
Gene2	v	v		v
…				
GeneK		v		

- The hypergeometric distribution is used to test the statistical significance.

Step 5: Reconstruct heat-responsive TRMs

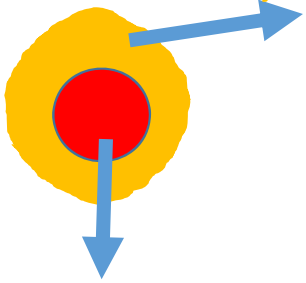
- For each heat-responsive TF set (e.g. [TF1, TF2]), we collect **all their regulatory targets that are heat-inducible** & **highly co-expressed** to form a TRM.

The regulatory targets of a TF set [TF1,TF2]



Heat-responsive genes identified in Step 3

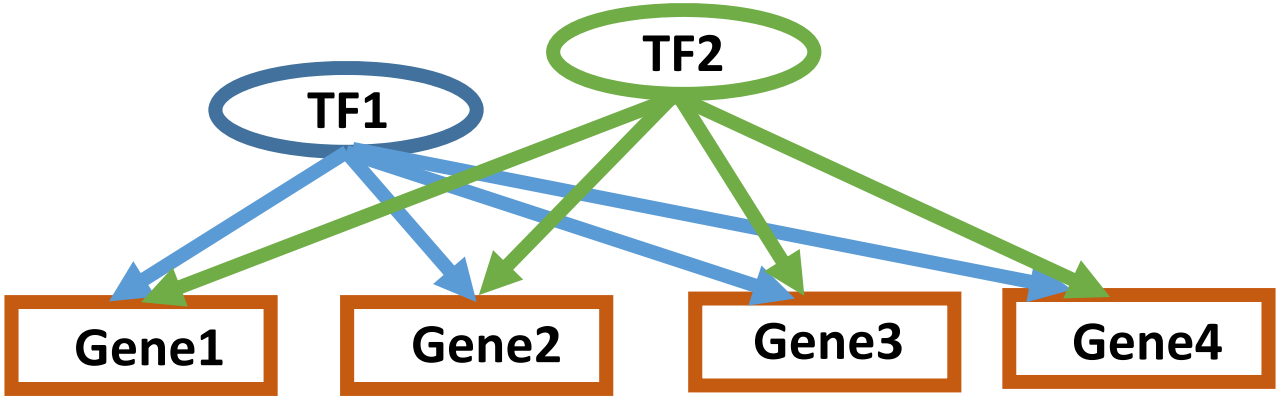
[Gene1, Gene2, Gene3, Gene4, Gene5, Gene6]



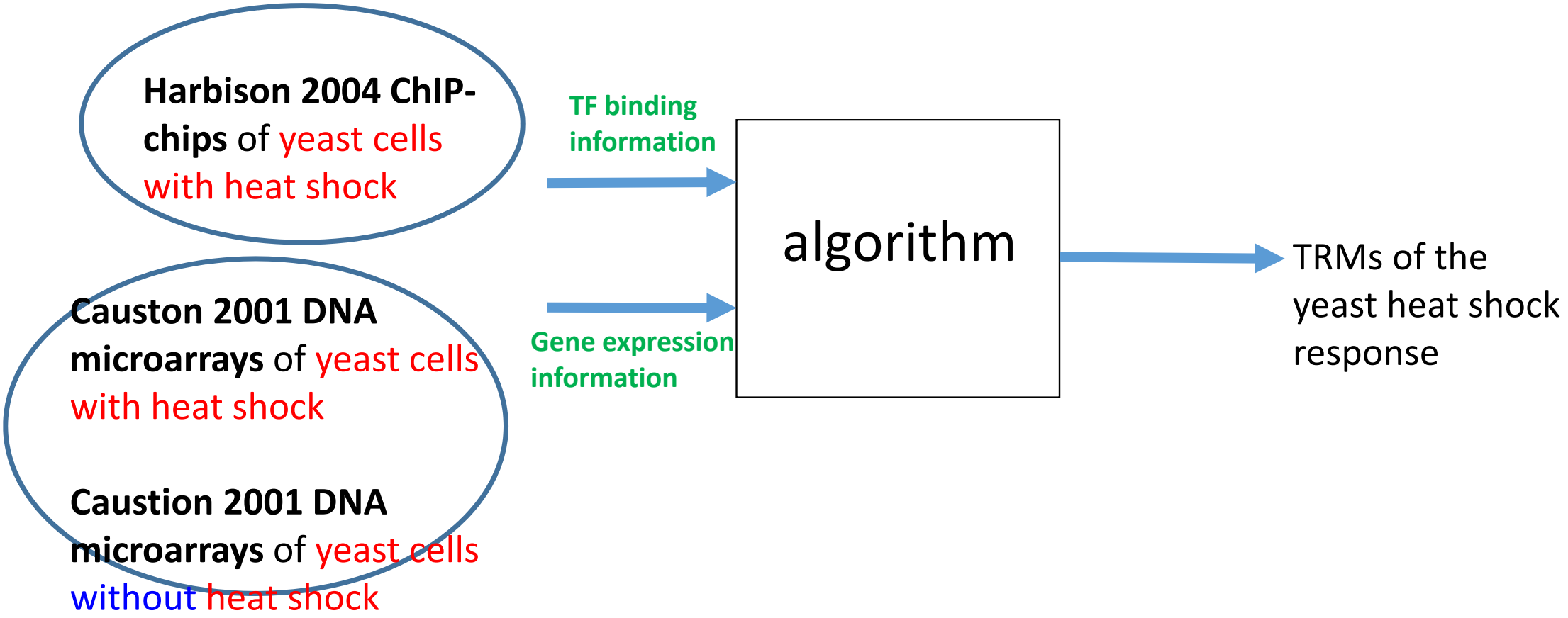
[Gene1, Gene2, Gene3, Gene4]

TRM:

{[TF1, TF2]→[Gene1, Gene2, Gene3, Gene4]}



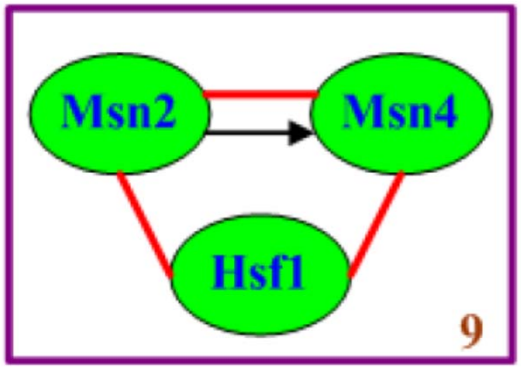
My algorithm for reconstructing TRMs of the yeast heat shock response



A TRM example

- **Msn2** colored blue: known to be involved in heat shock response
- The periphery of a rectangle colored **purple**: the module has at least one enriched MIPS functional category
- An oval colored **green**: the TF's function is consistent with at least one of the module's enriched MIPS functional categories

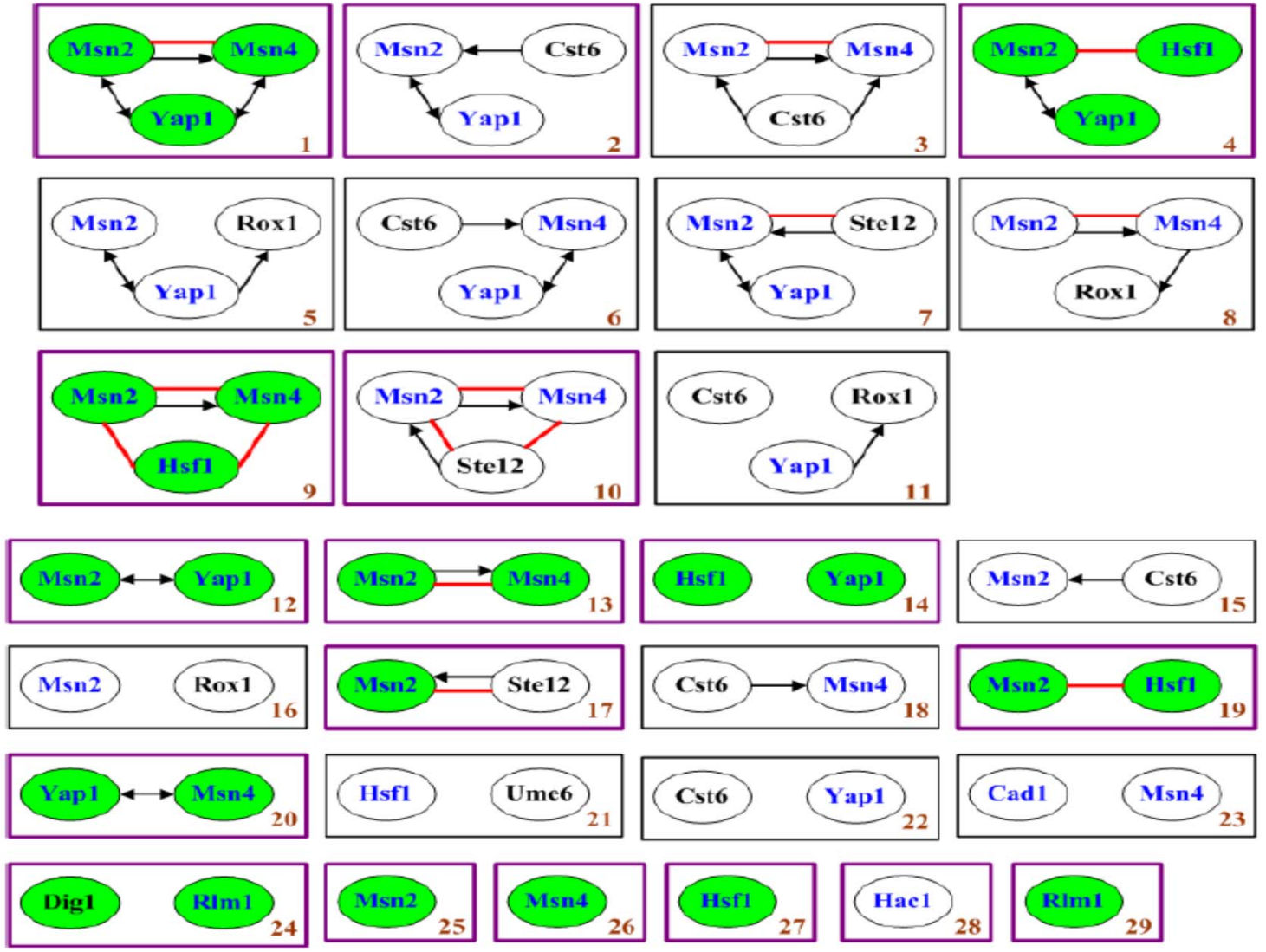
- Physical interaction
- Transcriptional regulation



(3,8) Module 9 regs:	Hsf1 Msn2 Msn4	
YBR117C	YBR117C	TKL2 Transketolase, similar t
YDR343C	YDR343C	HXT6 High-affinity glucose t
YDR533C	YDR533C	HSP31 Possible chaperone an
YEL039C	YEL039C	CYC7 Cytochrome c isoform
YER103W	YER103W	SSA4 Heat shock protein tha
YER150W	YER150W	SPI1 GPI-anchored cell wal
YKR076W	YKR076W	ECM4 Omega class glutathion
YLL026W	YLL026W	HSP104 Heat shock protein tha
P-VALUE	ENRICHED FUNCTIONAL CATEGORY	
1.58E-03	32.01 stress response	
3.35E-03	32.01.07 unfolded protein response (e.g. E	
3.44E-03	32 CELL RESCUE, DEFENSE AND VIRU	
6.01E-03	14.01 protein folding and stabilization	

TRMs of yeast heat shock response

- We identify 29 GRMs, containing 182 heat-responsive genes regulated by 12 heat-responsive TFs.
- 108/182 genes and 7/12 TFs are known to be involved in heat shock response.

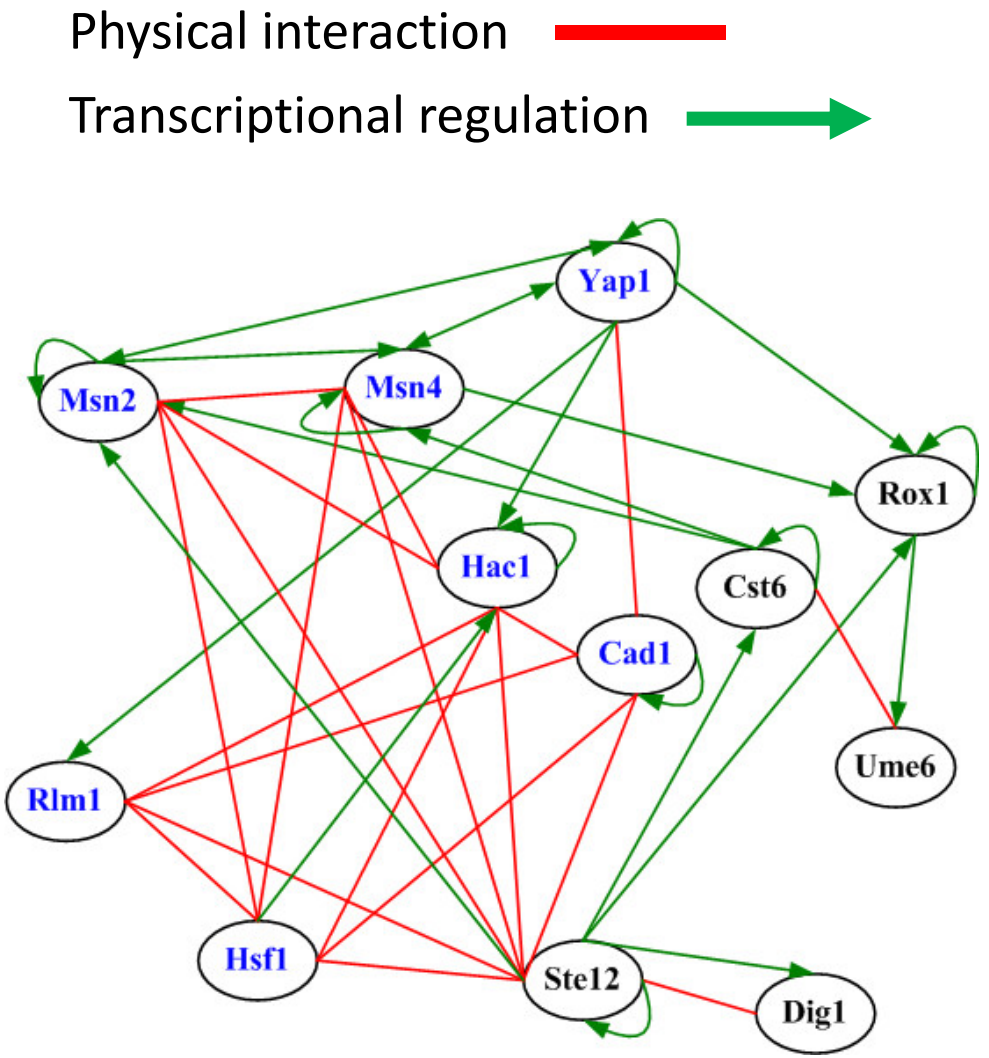


Identification of 12 heat-responsive TFs

- 7 TFs (Msn2, Msn4, Hsf1, Yap1, Hac1, Rlm1 and Cad1) known to be involved in heat shock response.
- Five novel heat-responsive TFs (Rox1, Cst6, Ume6, Ste12 and Dig1) have also been identified.

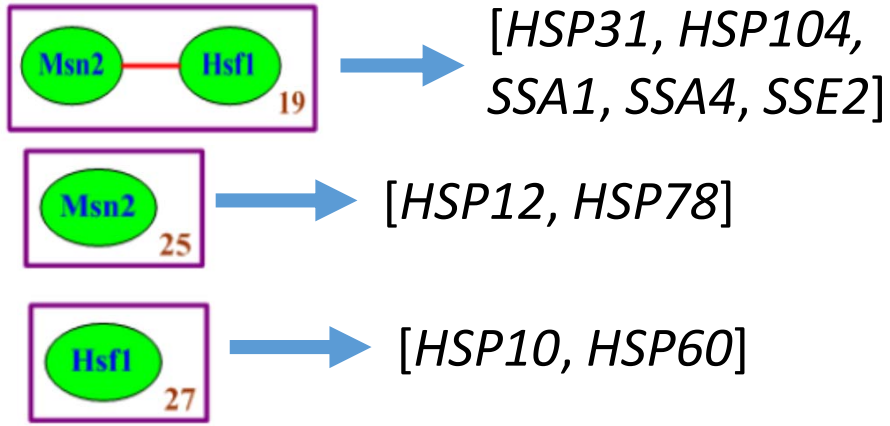
- These five novel heat-responsive TFs and the other seven known heat-responsive TFs form a highly connected network of interactions, suggest

1. these five novel heat-responsive TFs may play a role in heat shock response.
2. different combinations of a fairly small number of heat-responsive TFs may be sufficient to regulate a large number of genes involved in heat shock response.

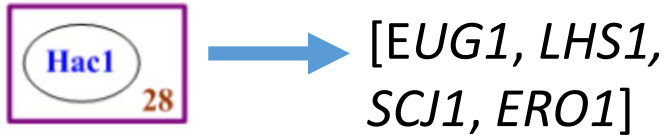


Identification of known heat-responsive genes

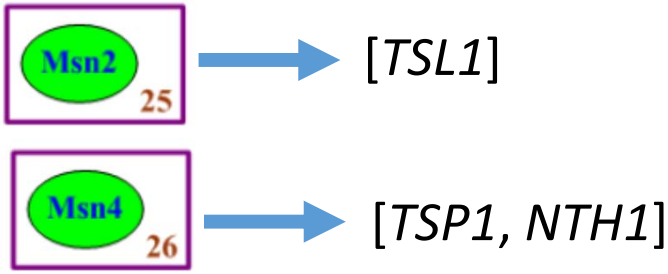
- Genes code for heat shock proteins (HSPs)



- Genes involved in the protein folding or refolding

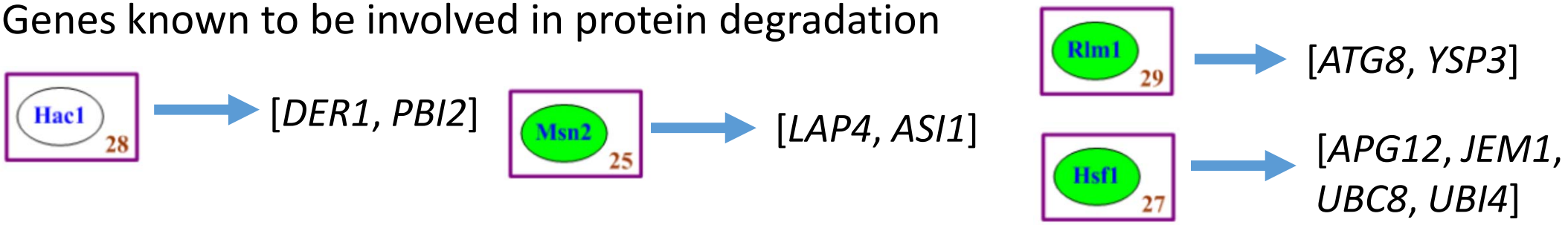


- Genes code for trehalose synthease subunits

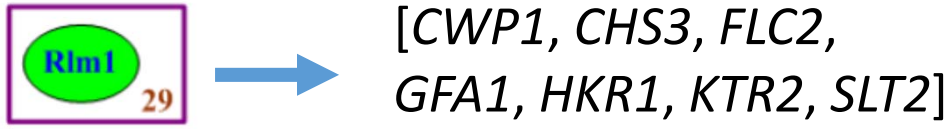


Identification of known heat-responsive genes

- Genes known to be involved in protein degradation

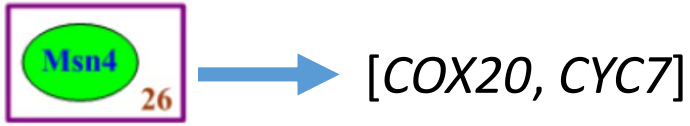


- Genes known to be involved in the cell wall biogenesis and maintenance



- Genes involved in glucose metabolism

- Genes involved in fatty acid metabolism



- Genes involved in respiration

Annotating 68 uncharacterized genes

- Among the 182 identified heat-responsive genes, 68 genes have unknown function according to the Saccharomyces Genome Database.
- We suggest that these genes are involved in heat shock response.

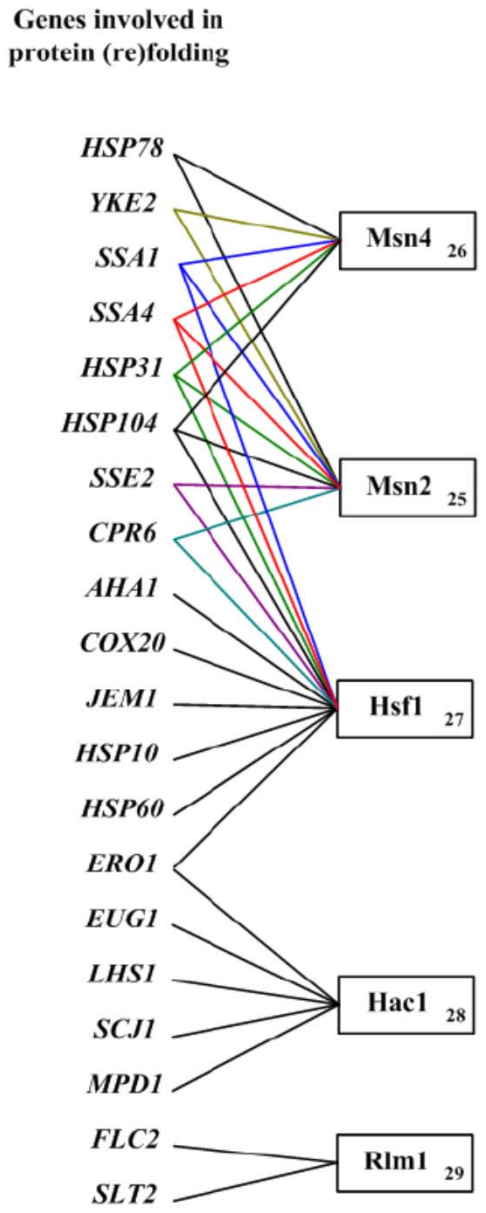
- Our predictions are supported by the fact that

1. all these 68 genes are induced by more than three folds at least at two time points of their expression profiles under heat shock
2. all these 68 genes are regulated by known heat-responsive TFs: 14 of these genes are regulated by Hac1, 25 by Hsf1, 24 by Msn2, 20 by Msn4, 17 by Yap1, and so on

- However, further experimental validations are needed to confidently annotate these uncharacterized genes as heat-responsive genes.

Refining the clusters of the genes involved in the protein (re)folding

- Heat shock causes an increased protein unfolding and aggregation in a cell.
- Thus, many genes that are involved in protein (re)folding are induced to bind to partially unfolded proteins to protect them from degradation or aggregation.
- Among the 182 identified heat-responsive genes, 20 genes are known to be involved in protein (re)folding.
- Although these genes are functionally similar, they may be under different transcriptional controls. Indeed, my algorithm assigns these 20 genes into five modules.
- My algorithm can refine the cluster of genes involved in protein (re)folding and degradation and can provide a better understanding of how the cell regulates the complex expression program of these genes.



Summary

- I developed an algorithm which combine ChIP-chip/ChIP-seq and DNA microarray/RNA-seq data to reconstruct TRMs of the yeast heat shock response.

- My algorithm identified 29 GRMs, which in total contain 182 heat-responsive genes regulated by 12 heat-responsive TFs.
- The literature indicates that 108 of the 182 genes and 7 of the 12 TFs are known to be involved in heat shock response.

- My algorithm suggested that 68 uncharacterized genes may be involved in heat shock response and it also identified their plausible heat-responsive regulators.

- My algorithm refined the cluster of genes that are involved in the protein (re)folding and provided a better understanding of how the complex expression program of heat shock is regulated.

Outlines

- Background of Transcriptional Regulatory Modules (TRMs)
- My algorithm for reconstructing TRMs
  -using yeast heat shock response as an example
- My other expertise: databases and web tools development

Yeast databases

YPA (Yeast Promoter Atlas) NAR featured article	**NAR 2011**
YTRP (Yeast Transcriptional Regulatory Pathway)	**Database 2014**
YNA (Yeast Nucleosome Atlas)	**BMC Genomics 2014**
CoopTFD (Cooperative TFs Database)	**Database 2016**
YCRD (Yeast Combinatorial Regulation Database)	**Plos One 2016**
YLBP (Yeast Lipid-Binding Proteins)	**Under construction**
YARG (Yeast Arsenic-Related Genes)	**Under construction**
YPRG (Yeast Prion-Related Genes)	**Planning to do**

Yeast web tools

YGA (Yeast Genes Analyzer)	**Gene 2012**
YNA (Yeast Nucleosome Atlas)	**BMC Genomics 2014**
YAGM (Yeast Associated Gene Miner)	**BMC Systems Biology 2015**
PCTFPeval (Predicted Cooperative TF Pairs evaluator)	**BMC Bioinformatics 2015**
YGSE (Yeast Gene Set Enrichment)	**Under construction**

Other databases and tools

- **Drosophila database**

cis-MEP (cis-regulatory Module Epigenetic Profile Database for Drosophila Melanogaster)	BMC Systems Biology 2014

- **Human database**

CSmirTar (Condition-Specific miRNA Targets)	Ready for submission
IRES (Internal Ribosome Entry Zone)	Ready for submission
RPDV (Ribosome Profiling Data Viewer)	Under construction
p53BLV (p53 Binding Location Viewer)	Under construction

- **Phosphoproteomic tool**

iPhos (a toolkit to streamline the alkaline phosphatase-assisted comprehensive LC-MS phosphoproteome investigation)	BMC Bioinformatics 2014

- **Microarray missing value imputation**

MissVIA (Missing Value Imputation Atlas)	BMC Systems Biology 2013
MVIAeval (Missing Value Imputation Algorithm evaluator)	Submitted

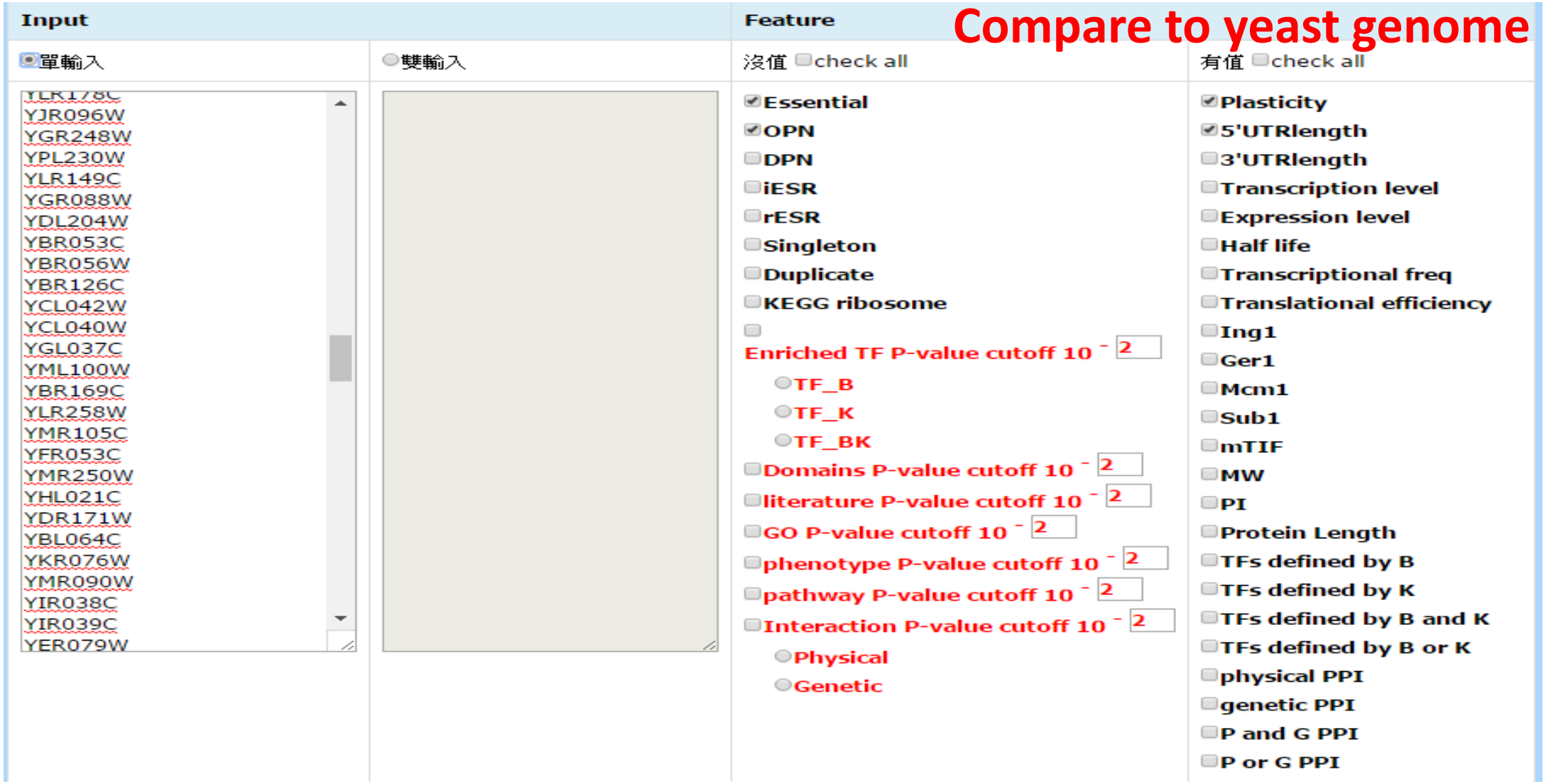
A tool for yeast gene sets enrichment & comparison

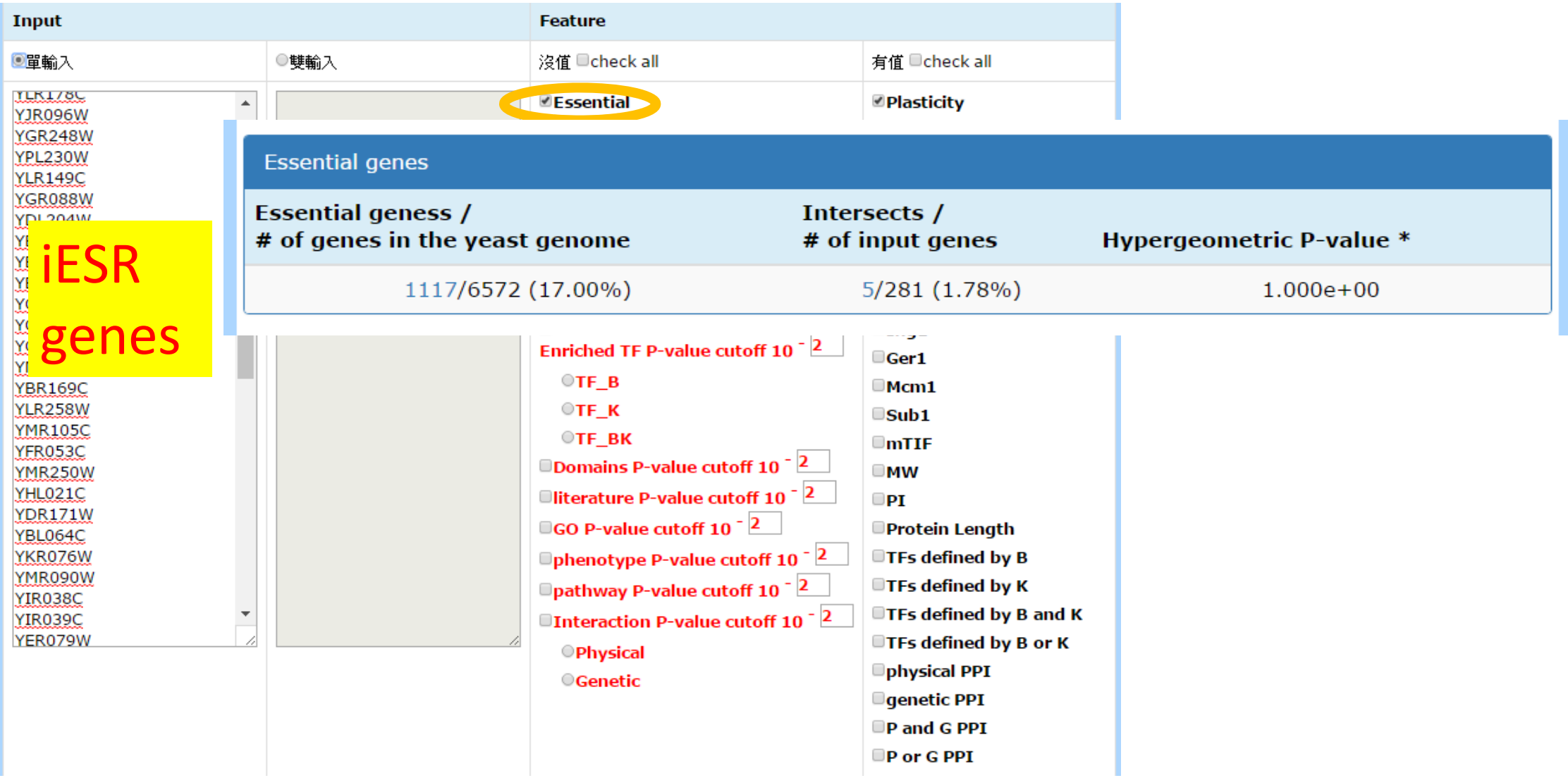
- Nowadays, yeast biologists can have a set of genes (e.g. differentially expressed genes under heat shock) or two sets of genes (induced genes under heat shock vs. repressed genes under heat shock) easily by using **RNA-seq technology**.

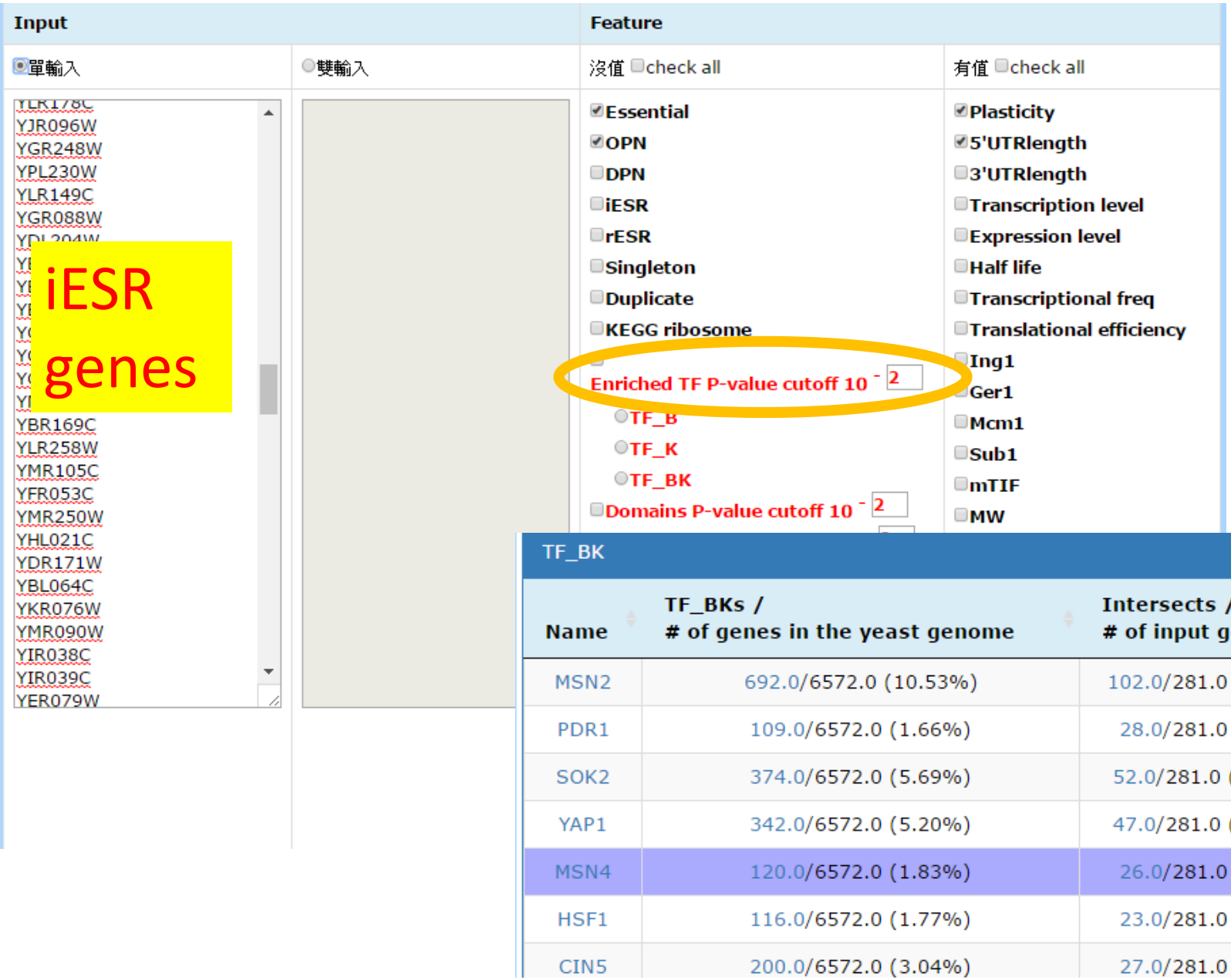
- How to make biological sense of these genes is challenging.

- My lab is developing a web tool for yeast biologists.

Web interface of a single list of genes




Input
Feature
單輸入
雙輸入
沒值 check all
有值 check all
YLR178C
YJR096W
YGR248W
YPL230W
YLR149C
YGR088W
Essential
Plasticity
Essential genes
Essential geness / # of genes in the yeast genome
Intersects / # of input genes
Hypergeometric P-value *
1117/6572 (17.00%)
5/281 (1.78%)
1.000e+00
iESR genes
YBR169C
YLR258W
YMR105C
YFR053C
YMR250W
YHL021C
YDR171W
YBL064C
YKR076W
YMR090W
YIR038C
YIR039C
YER079W
Enriched TF P-value cutoff 10 - 2
TF_B
TF_K
TF_BK
Domains P-value cutoff 10 - 2
literature P-value cutoff 10 - 2
GO P-value cutoff 10 - 2
phenotype P-value cutoff 10 - 2
pathway P-value cutoff 10 - 2
Interaction P-value cutoff 10 - 2
Physical
Genetic
Ger1
Mcm1
Sub1
mTIF
MW
PI
Protein Length
TFs defined by B
TFs defined by K
TFs defined by B and K
TFs defined by B or K
physical PPI
genetic PPI
P and G PPI
P or G PPI



TF_BK

Name	TF_BKs / # of genes in the yeast genome	Intersects / # of input genes	Hypergeometric P-value *
MSN2	692.0/6572.0 (10.53%)	102.0/281.0 (36.30%)	4.061e-30
PDR1	109.0/6572.0 (1.66%)	28.0/281.0 (9.96%)	7.181e-13
SOK2	374.0/6572.0 (5.69%)	52.0/281.0 (18.51%)	1.455e-12
YAP1	342.0/6572.0 (5.20%)	47.0/281.0 (16.73%)	5.389e-11
MSN4	120.0/6572.0 (1.83%)	26.0/281.0 (9.25%)	5.299e-10
HSF1	116.0/6572.0 (1.77%)	23.0/281.0 (8.19%)	6.638e-08
CIN5	200.0/6572.0 (3.04%)	27.0/281.0 (9.61%)	1.269e-05

Input		Feature	
單輸入	雙輸入	沒值 check all	有值 check all
		Essential	Plasticity

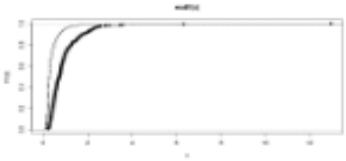
YLR178C
YJR096W
YGR248W
YPL230W
YLR149C
YGR088W
YDL204W
YBR169C
YLR258W
YMR105C
YFR053C
YMR250W
YHL021C
YDR171W
YBL064C
YKR076W
YMR090W
YIR038C
YIR039C
YER079W

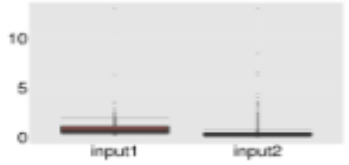
iESR genes

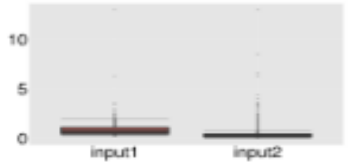
Plasticity

input1

# of input genes	Median	Mean	Yeast Genome # of genes	Median	Mean
279/281 (99.288 %)	0.72	0.938	6105/6572 (92.894 %)	0.252	0.368

K-test P-value		
A ≠ B	0	
A > B	7.012837e-77	
A < B	1	

U-test P-value		
A ≠ B	2.528223e-89	
A > B	1.264112e-89	
A < B	1	

T-test P-value		
A ≠ B	6.159119e-19	
A > B	3.07956e-19	
A < B	1	

Web interface of comparing two lists of genes

Analyses results of rESR vs. iESR

- See webpage

Be my collaborators

- I am familiar with many kinds of yeast genome-wide data (sequence, gene expression, ChIP-chip, TF knockout, nucleosome occupancy, histone modification, PPI, genetic interaction, ribosome profiling, protein phosphorylation, mutant phenotype, GO, pathway, …)
- My lab has eight master students who are good at coding.
- We are happy to develop databases or tools for you.
- Yeast biologists, tell me your needs!

Any Questions or Comments?