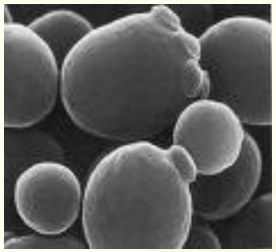
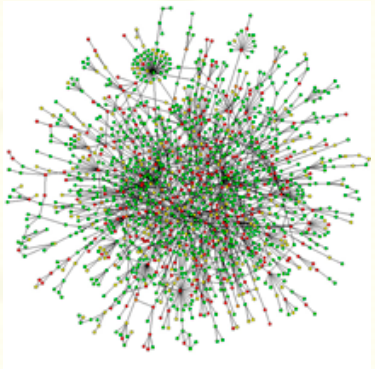
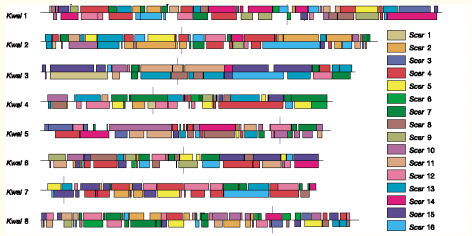
Lecture 3 - 5: Yeast as a model organism for functional and evolutionary genomics and systems biology







Lecture outline

- **Objective:** using yeast as an example to introduce basic concepts, experimental and computational techniques

- Basics on budding yeast
  - Haploid （单被）, diploid （双被体）, essential and non-essential genes

- Study gene function using genomics and proteomics
  - Microarray, transcriptional regulation, ChIP-chip, regulatory evolution
  - Protein-protein interactions, protein complexes, biological network, genetic interactions, high content cell biology

- Genome duplication and evolution

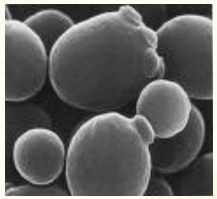
Budding Yeast: *Saccharomyces cerevisiae*

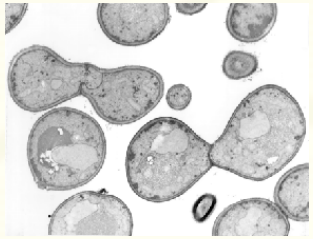
- "domesticated" by ancient human, used in baking and brewing
  - *"Saccharo-"* = sugar, "myces' = mushroom, fungas,
  - "cerevisiae" = "of beer"

Budding Yeast: *Saccharomyces cerevisiae*

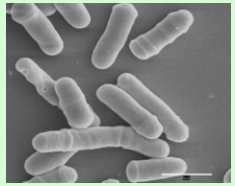
- Can grow under both aerobic（有氧）and anaerobic（无氧）(fermentation) conditions

- Can exist as both haploid and diploid form.
  - In haploid (单被） form, yeast undergo mitosis (有丝分裂） (asexual budding)
  - In diploid （双倍） form, yeast can undergo sporulation, enter sexual reproduction (meiosis 减数分裂) and produce haploid spores (**a** and ***α***). The spores can mate (conjugate) and form diploid again.

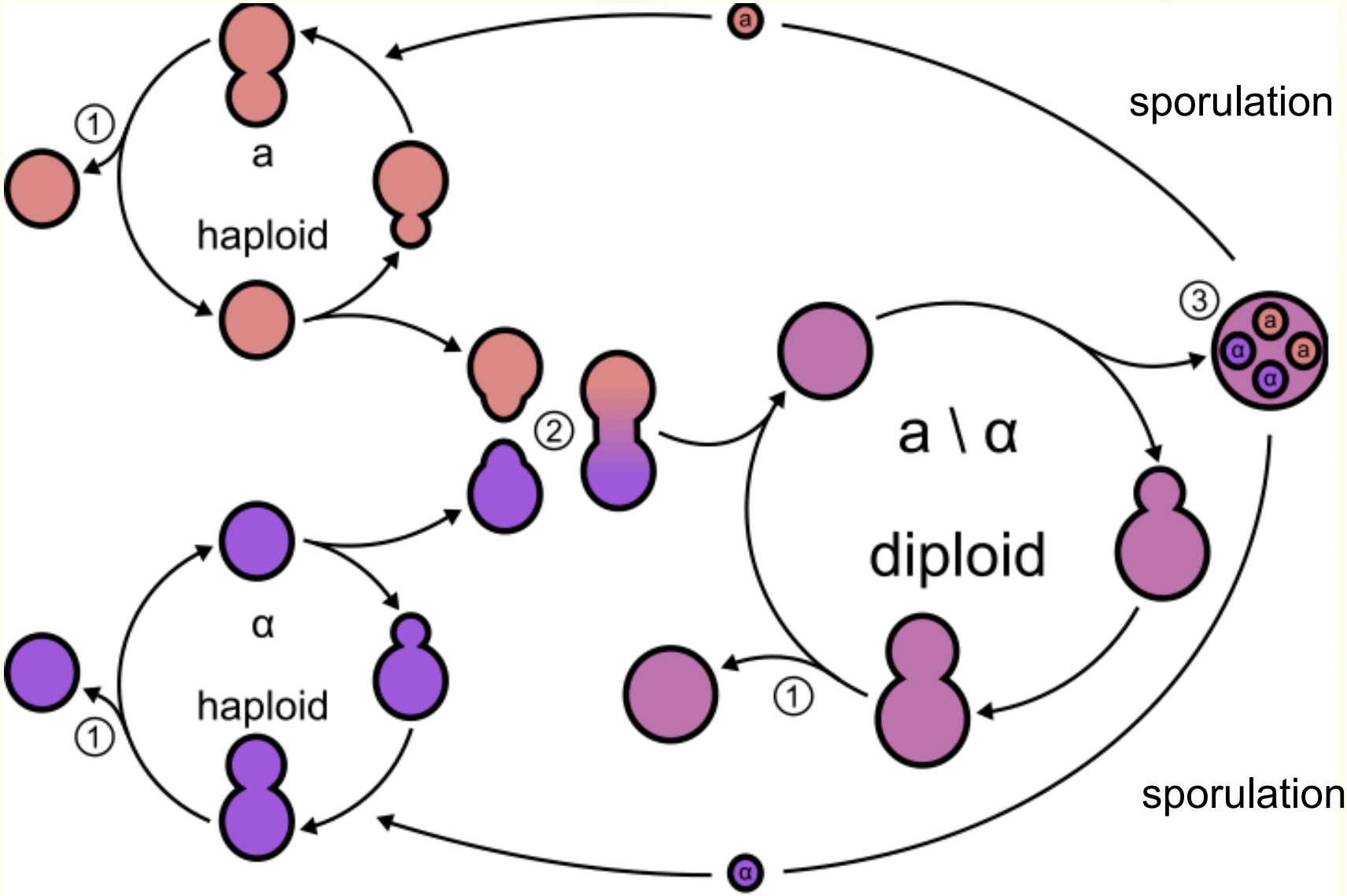




fission yeast is another less commonly used yeast model organism:
Schizosaccharomyces pombe

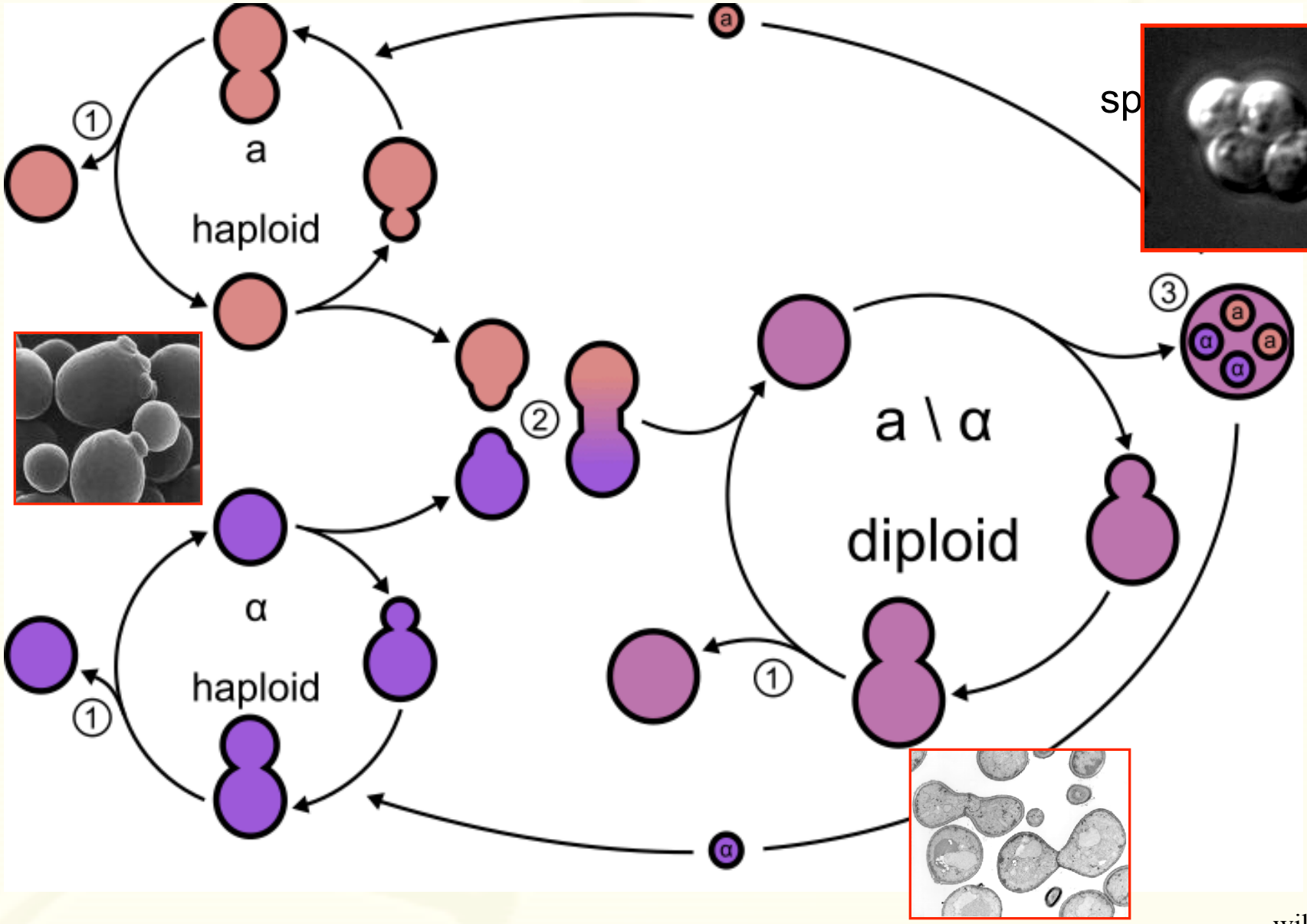


Life cycle of budding yeast



wikipedia

Life cycle of budding yeast



S. cerevisiae genome

- 16 chromosomes (haploid), 12 million bp, ~6,000 genes,
  - Human has 22,000 genes, 3,000 million bp
- 46% of yeast genes have orthologs in humans
- 290 yeast genes are orthologs of human disease genes
- Only 220 genes have introns (4%),
  - In fission yeast S. pombe 40% genes have introns

- Yeast is the most important eukaryotic model organism, many important biological discoveries were made in yeast, e.g. cell cycle.

- Fully automated experimental protocol allows genome-wide study, very collaborative research community

many yeast genes are “non-essential”

- ***Saccharomyces Genome* Deletion Project**: systematically delete every gene and assay for survival and fitness effects.

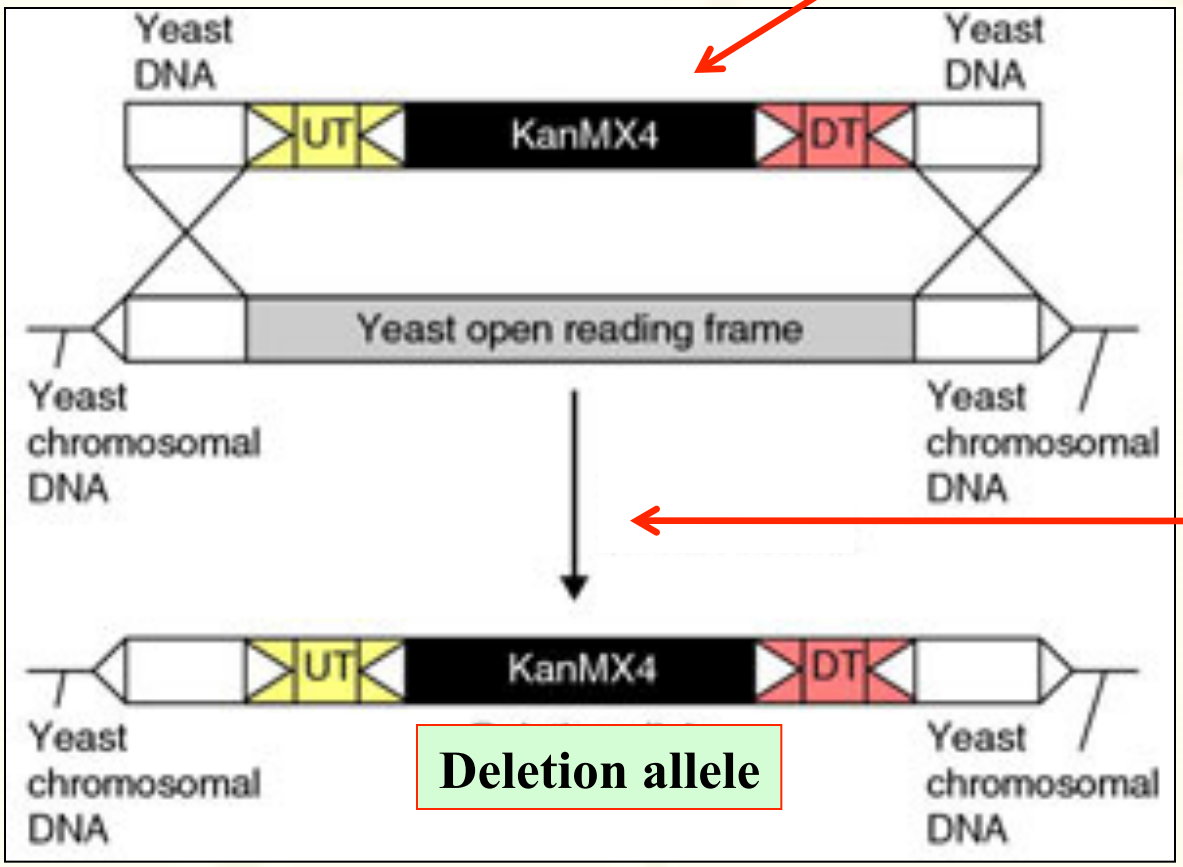
- Only 19% of yeast genes (1100 of ~5900) are essential for growth on rich glucose media, the majority of yeast deletion strains are viable.

- Why are some genes “nonessential” ?
  - Functionally not vital
  - Only essential in specific condition
  - Redundancy by duplicated gene
  - Subtle fitness effect that only manifest after long evolutionary time

Winzeler et al Science 1999, *Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis*

Create yeast deletion strain

drug resistance marker used to select successful deletion strain



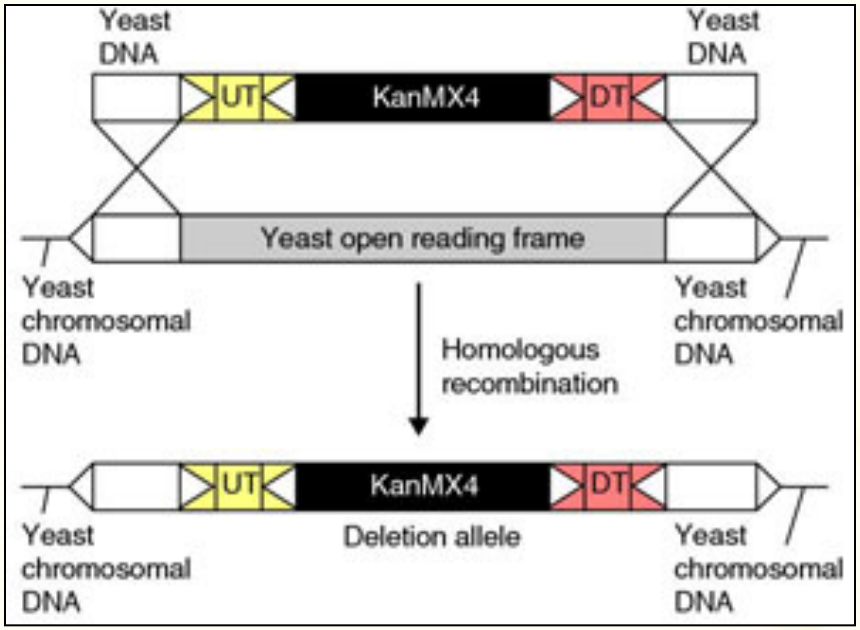
Unique 20 nt tags

Homologous recombination

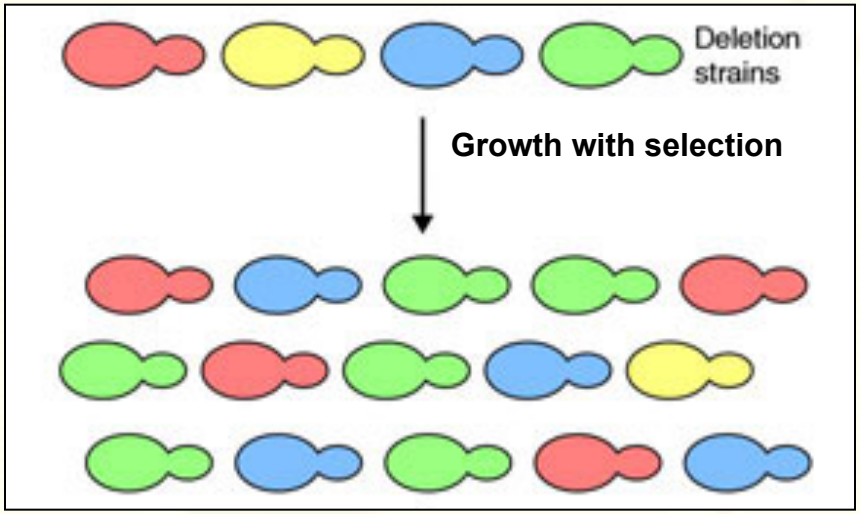
Scherens et al Genome Biology 2004

Scherens Genome Biology 2004

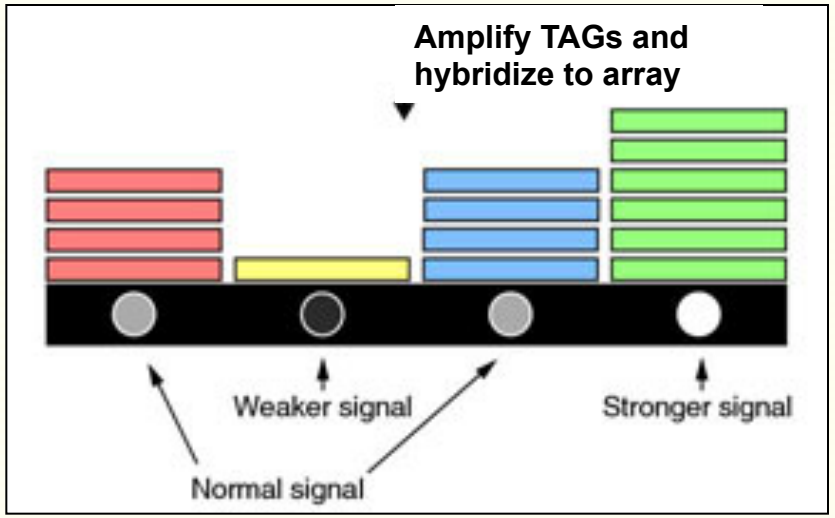
Yeast deletion strain collection (and a preview of chemical genomics)



Delete genes and add barcodes

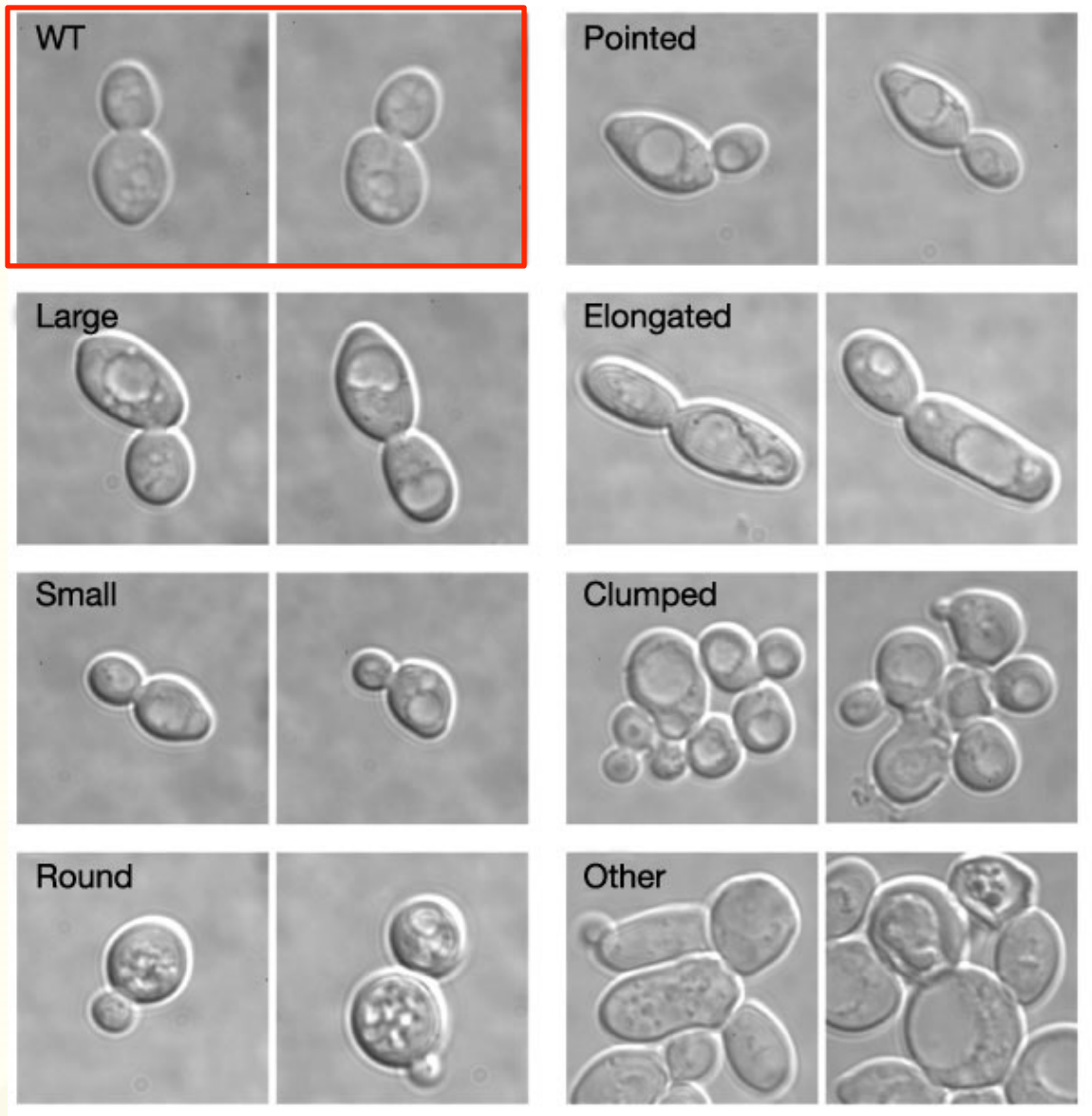


Grow pooled strains, (chemical) selection



Read-out by microarray or sequencing

Deletion of nonessential genes have morphological phenotypes



Giaever et al *Functional profiling of the Saccharomyces cerevisiae genome* Nature 2002

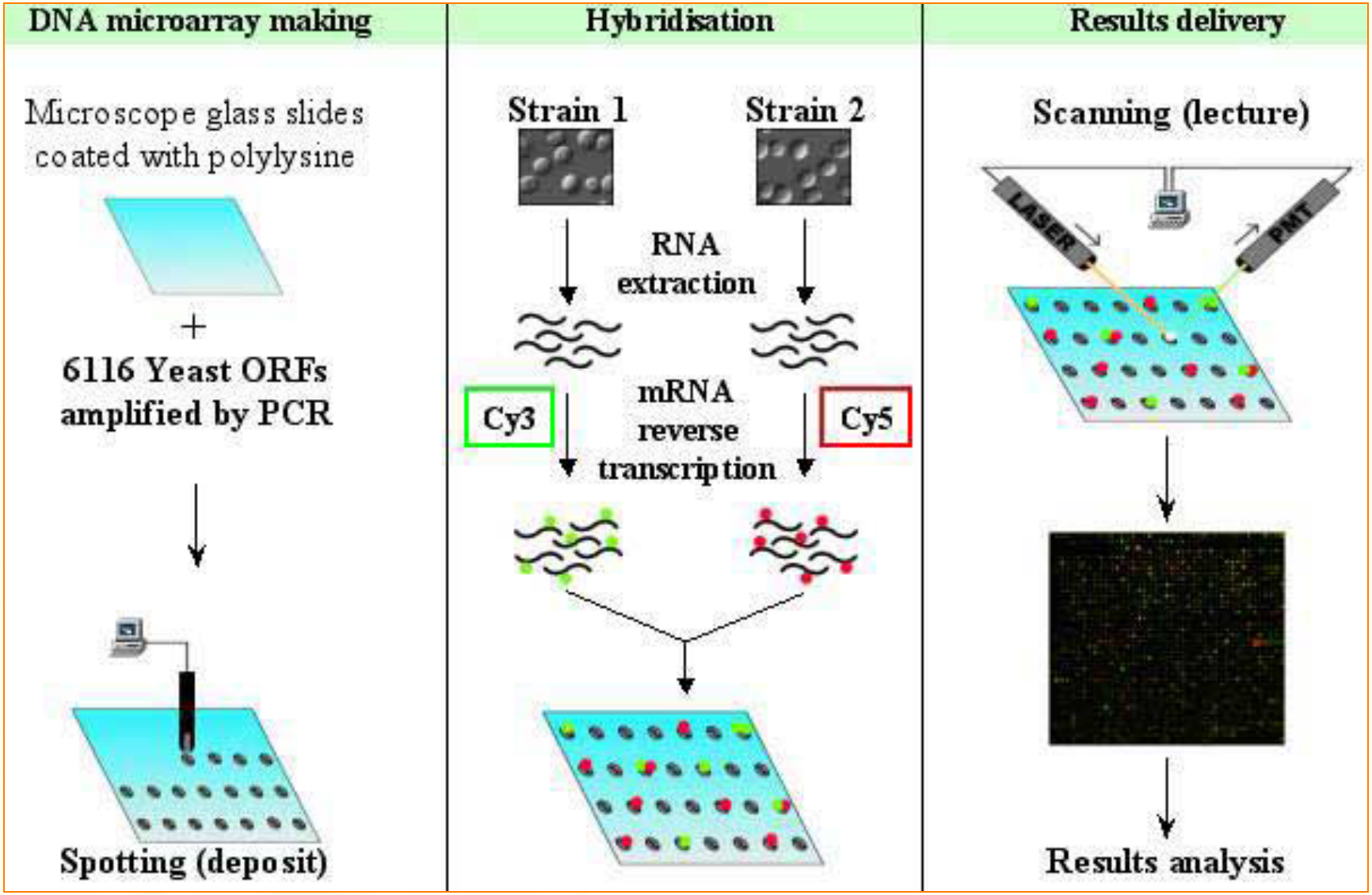
Yeast has many high-throughput genomics data

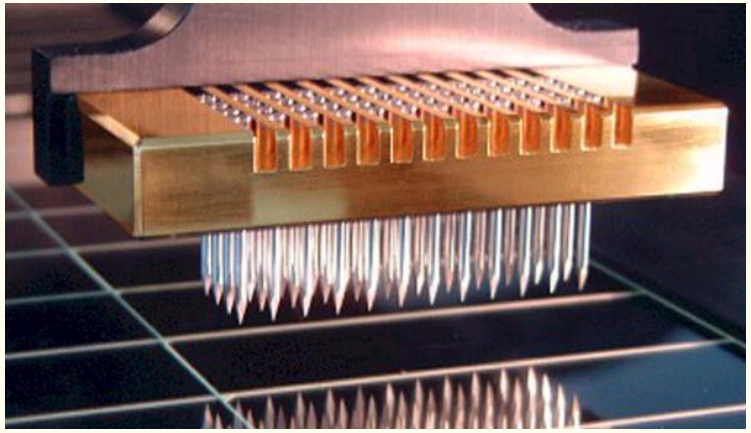
- **Gene expression (by microarray or RNA-seq)**
  - Cell cycle, deletion strain, chemical perturbations
- **Transcription regulation (binding by transcription factors)**
  - ChIP-chip, ChIP-seq
- **Protein-protein interactions and complex data**
  - Yeast Two-Hybrid (Y2H), TAP-tagging, literature curation
- **Genetic interactions and pathways**
  - Synthetic Genetic Array (SGA)
- **Chemical genomics**
  - Small molecule – gene interactions
- **High content morphological screening**

Gene expression

- Microarray or next-generation sequencing can measure gene expression level of tens of thousands of genes at once.

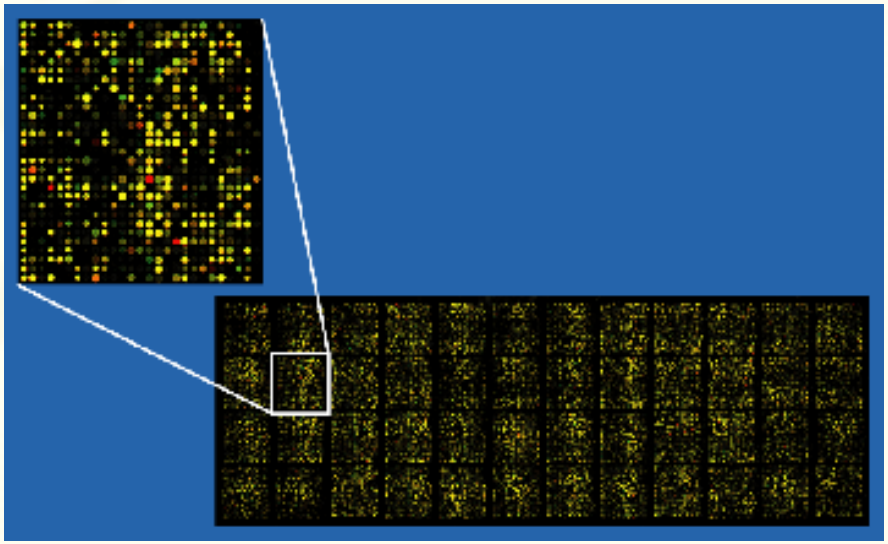
- There are two types of microarray experiments:
  - to measure the absolute mRNA abundance in the sample,
  - to measure the difference between two samples, or after perturbation e.g. DNA damage, starvation, heat shock, or small molecules (drugs)


DNA microarray making
Hybridisation
Results delivery
Microscope glass slides coated with polylysine
+
6116 Yeast ORFs amplified by PCR
Spotting (deposit)
Strain 1
Strain 2
RNA extraction
Cy3
Cy5
mRNA reverse transcription
Scanning (lecture)
LASER
PMT
Results analysis



"printing" cDNAs onto slides

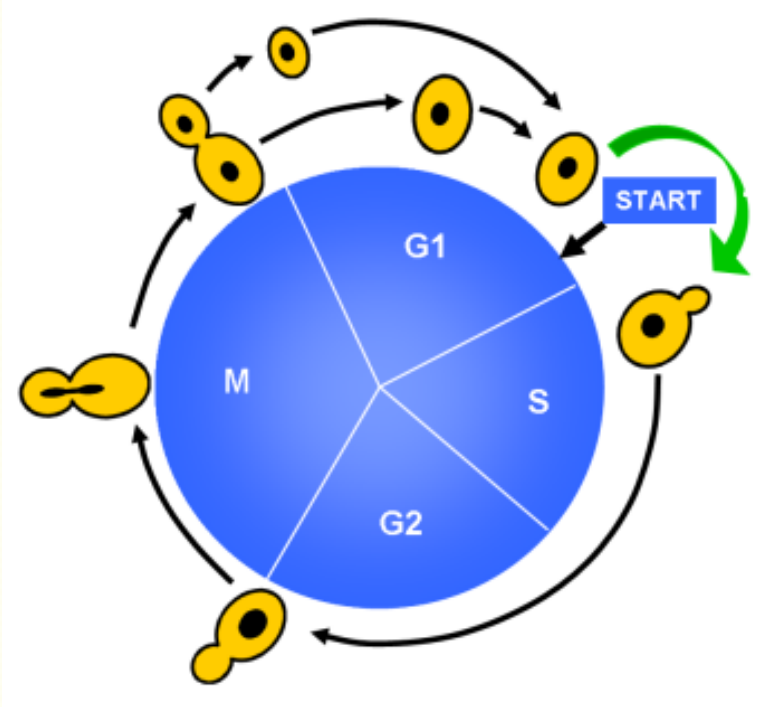
Array scanner



Gene expression in yeast cell cycle

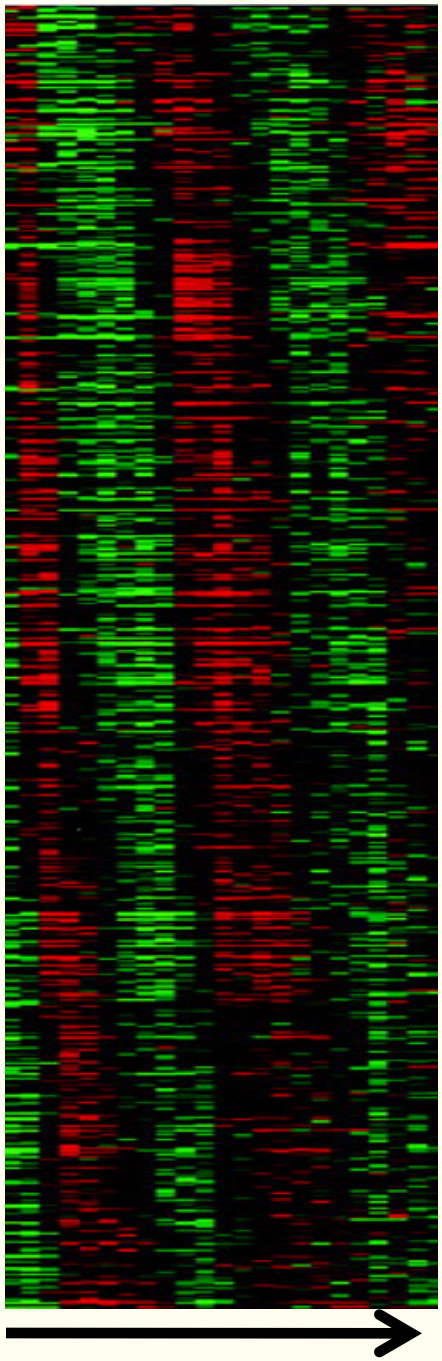
- Spellman et al identified ~800 yeast genes whose mRNA expression level are tightly regulated through cell-clcye.
- "*Comprehensive identification of cell cycle-regulated genes of the yeast"* MCB 1998

.



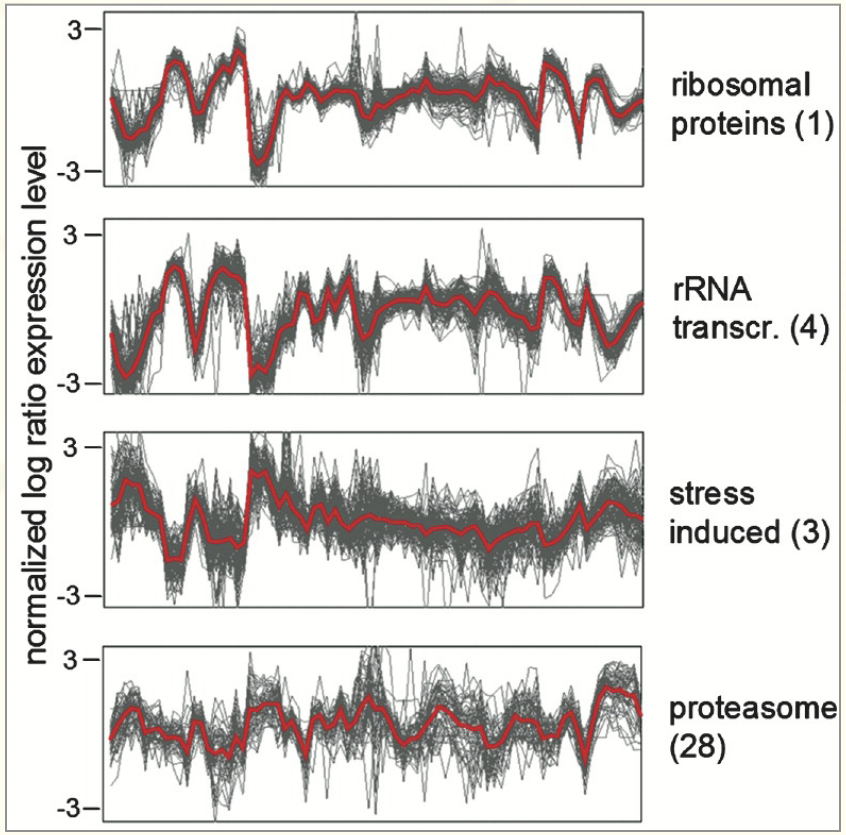
Cell cycle

Genes



Genes of similar function often have similar expression profile

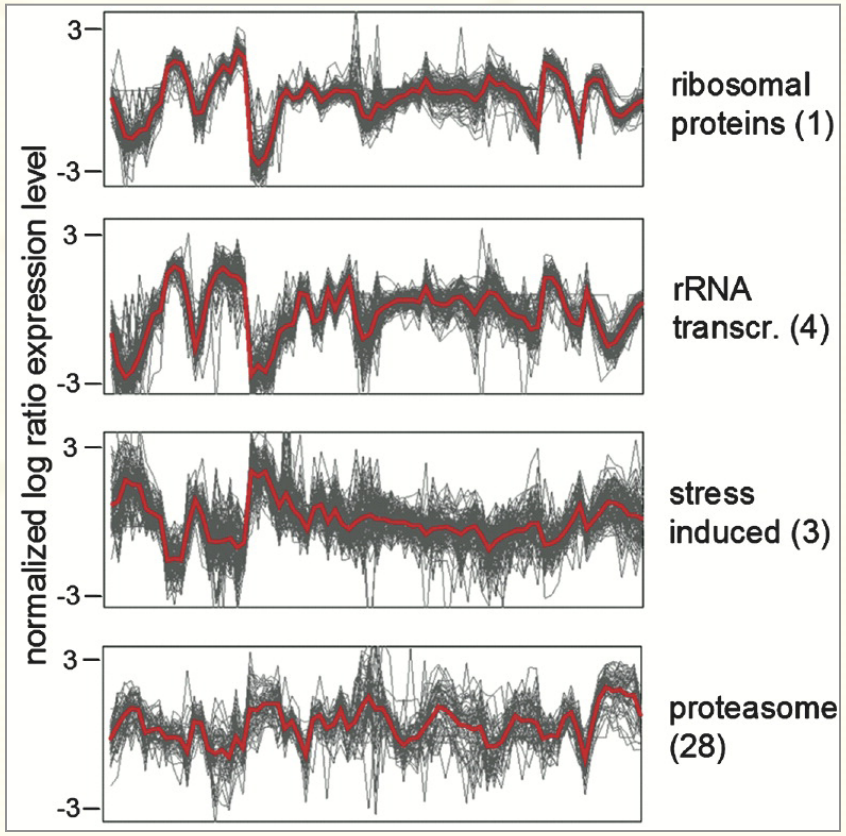
- Genes that have similar expression profile (i.e. co-expressed) often share similar function or on the same pathway: “guilt by association”.



Beer & Tavazoie Cell 2004

Genes of similar function often have similar expression profile

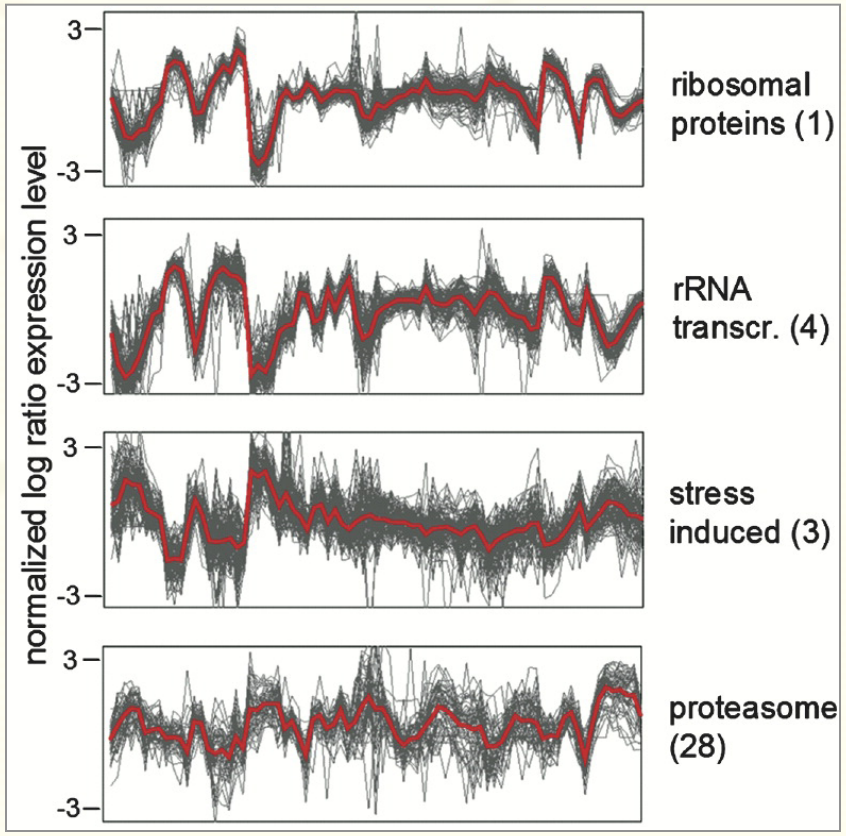
- Gene expression experiments can identify targets of regulators such as transcription factors (TFs), microRNAs, or drugs.



Beer & Tavazoie Cell 2004

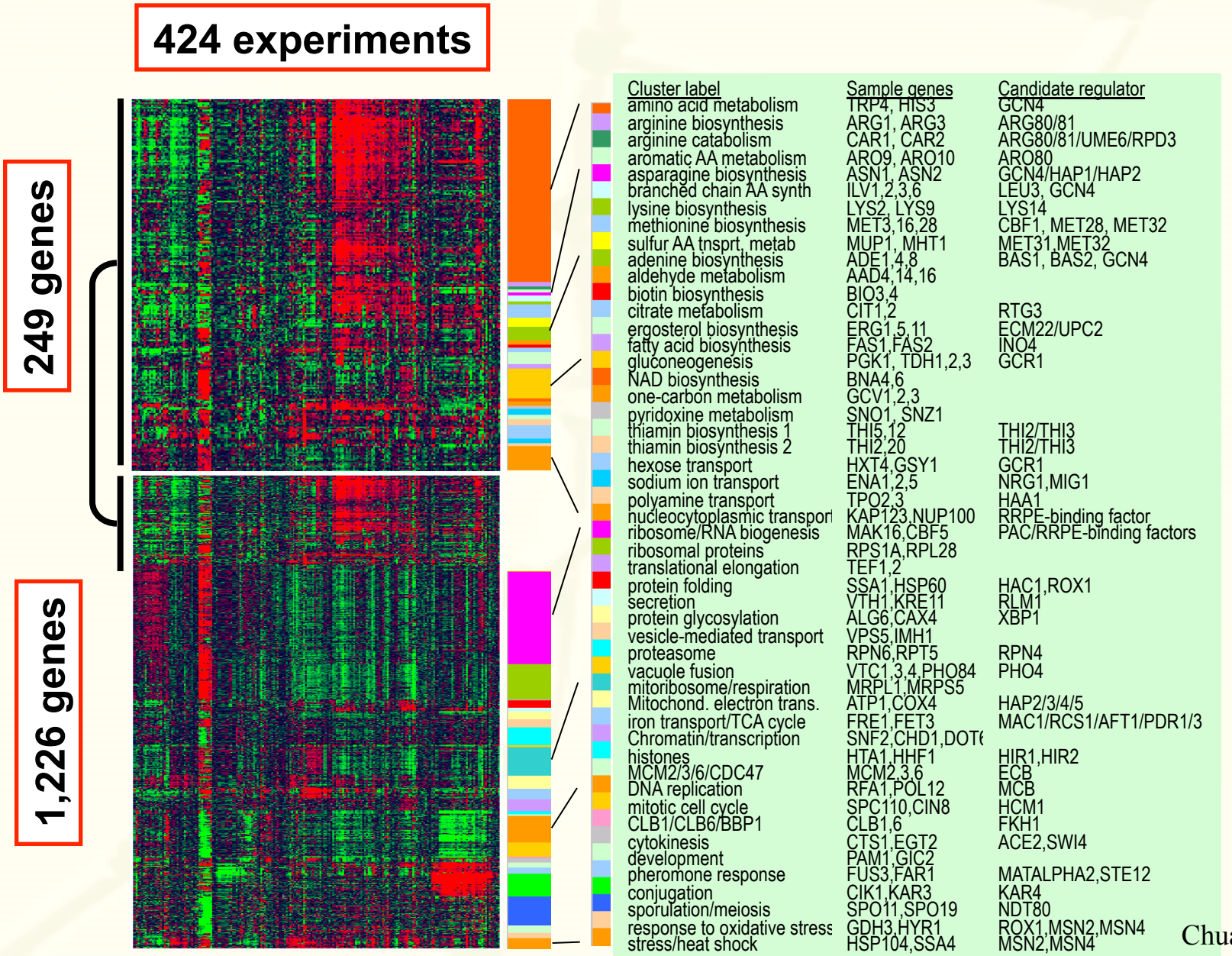
Genes of similar function often have similar expression profile

- Genes that are co-expressed often share common regulators (transcription factors).



Beer & Tavazoie Cell 2004

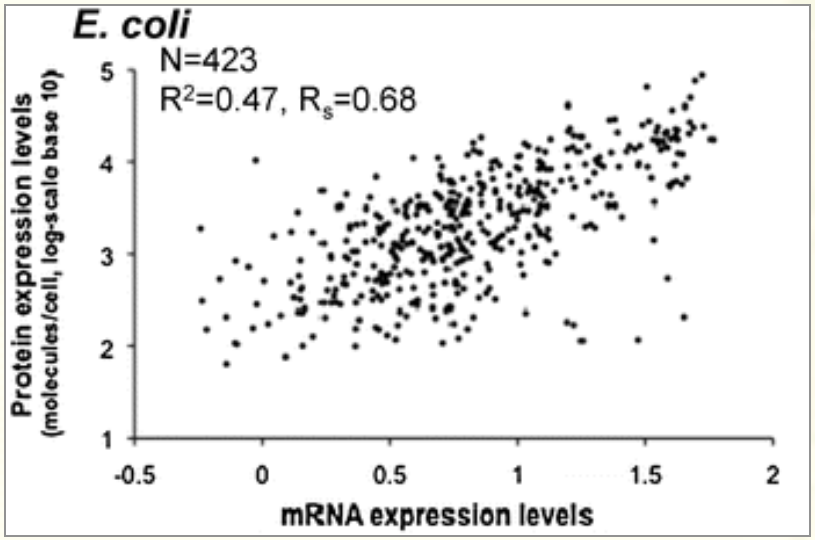
Gene expression clustering reveal functional similarity

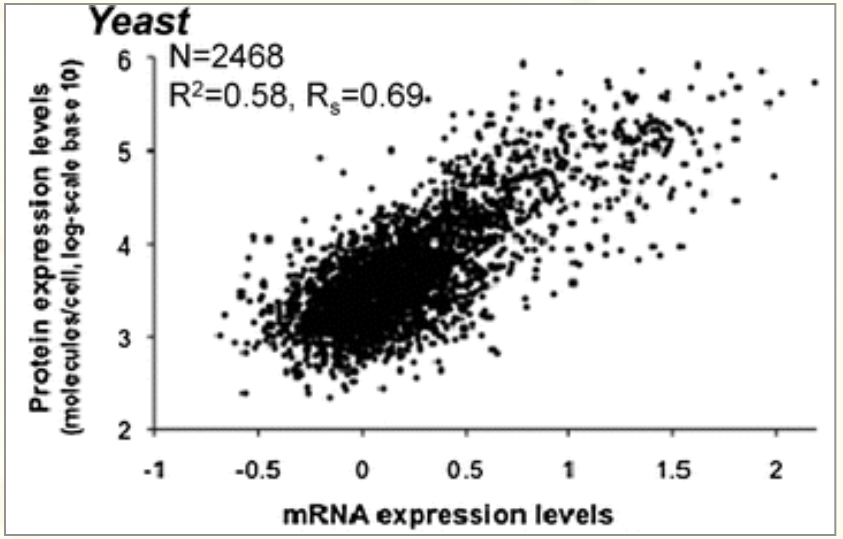


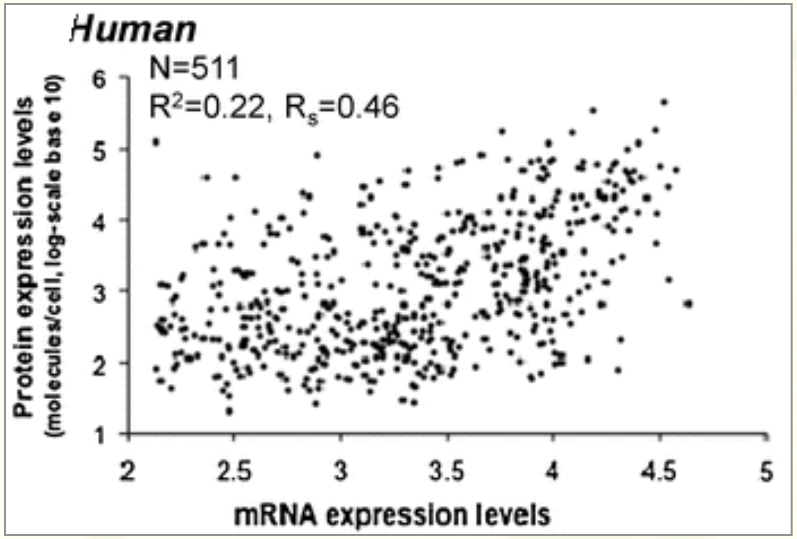
Cluster label	Sample genes	Candidate regulator
amino acid metabolism	TRP4, HIS3	GCN4
arginine biosynthesis	ARG1, ARG3	ARG80/81
arginine catabolism	CAR1, CAR2	ARG80/81/UME6/RPD3
aromatic AA metabolism	ARO9, ARO10	ARO80
asparagine biosynthesis	ASN1, ASN2	GCN4/HAP1/HAP2
branched chain AA synth	ILV1,2,3,6	LEU3, GCN4
lysine biosynthesis	LYS2, LYS9	LYS14
methionine biosynthesis	MET3,16,28	CBF1, MET28, MET32
sulfur AA tnsprt, metab	MUP1, MHT1	MET31,MET32
adenine biosynthesis	ADE1,4,8	BAS1, BAS2, GCN4
aldehyde metabolism	AAD4,14,16	
biotin biosynthesis	BIO3,4	
citrate metabolism	CIT1,2	RTG3
ergosterol biosynthesis	ERG1,5,11	ECM22/UPC2
fatty acid biosynthesis	FAS1,FAS2	INO4
gluconeogenesis	PGK1, TDH1,2,3	GCR1
NAD biosynthesis	BNA4,6	
one-carbon metabolism	GCV1,2,3	
pyridoxine metabolism	SNO1, SNZ1	
thiamin biosynthesis 1	THI5,12	THI2/THI3
thiamin biosynthesis 2	THI2,20	THI2/THI3
hexose transport	HXT4,GSY1	GCR1
sodium ion transport	ENA1,2,5	NRG1,MIG1
polyamine transport	TPO2,3	HAA1
nucleocytoplasmic transport	KAP123,NUP100	RRPE-binding factor
ribosome/RNA biogenesis	MAK16,CBF5	PAC/RRPE-binding factors
ribosomal proteins	RPS1A,RPL28	
translational elongation	TEF1,2	
protein folding	SSA1,HSP60	HAC1,ROX1
secretion	VTH1,KRE11	RLM1
protein glycosylation	ALG6,CAX4	XBP1
vesicle-mediated transport	VPS5,IMH1	
proteasome	RPN6,RPT5	RPN4
vacuole fusion	VTC1,3,4,PHO84	PHO4
mitoribosome/respiration	MRPL1,MRPS5	
Mitochond. electron trans.	ATP1,COX4	HAP2/3/4/5
iron transport/TCA cycle	FRE1,FET3	MAC1/RCS1/AFT1/PDR1/3
Chromatin/transcription	SNF2,CHD1,DOT(	
histones	HTA1,HHF1	HIR1,HIR2
MCM2/3/6/CDC47	MCM2,3,6	ECB
DNA replication	RFA1,POL12	MCB
mitotic cell cycle	SPC110,CIN8	HCM1
CLB1/CLB6/BBP1	CLB1,6	FKH1
cytokinesis	CTS1,EGT2	ACE2,SWI4
development	PAM1,GIC2	
pheromone response	FUS3,FAR1	MATALPHA2,STE12
conjugation	CIK1,KAR3	KAR4
sporulation/meiosis	SPO11,SPO19	NDT80
response to oxidative stress	GDH3,HYR1	ROX1,MSN2,MSN4
stress/heat shock	HSP104,SSA4	MSN2,MSN4

Chua et al., 2004

How well can mRNA level reflect protein abundance ?



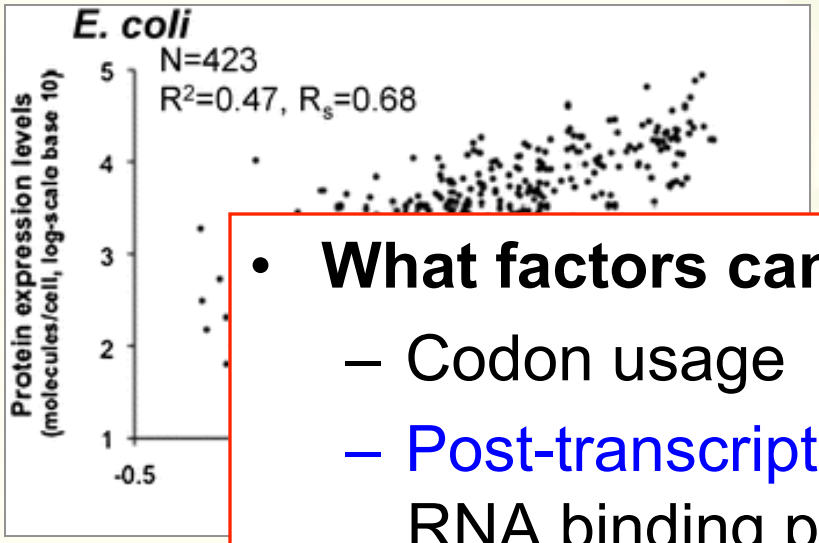


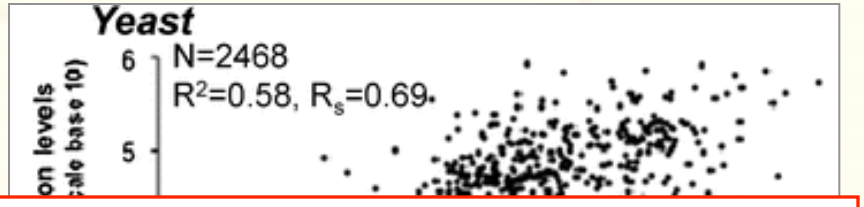


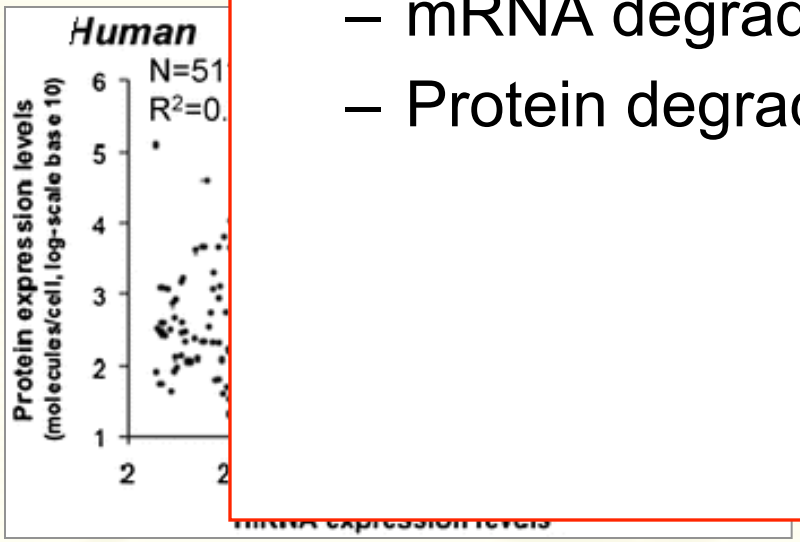
- **What factors can cause the discrepancy ?**
  - Codon usage
  - Post-transcriptional regulation: microRNAs, RNA binding proteins
  - mRNA degradation and half-life
  - Protein degradation and half-life

Vogel C. Mol Biosystems 2009

How well can mRNA level reflect protein abundance ?



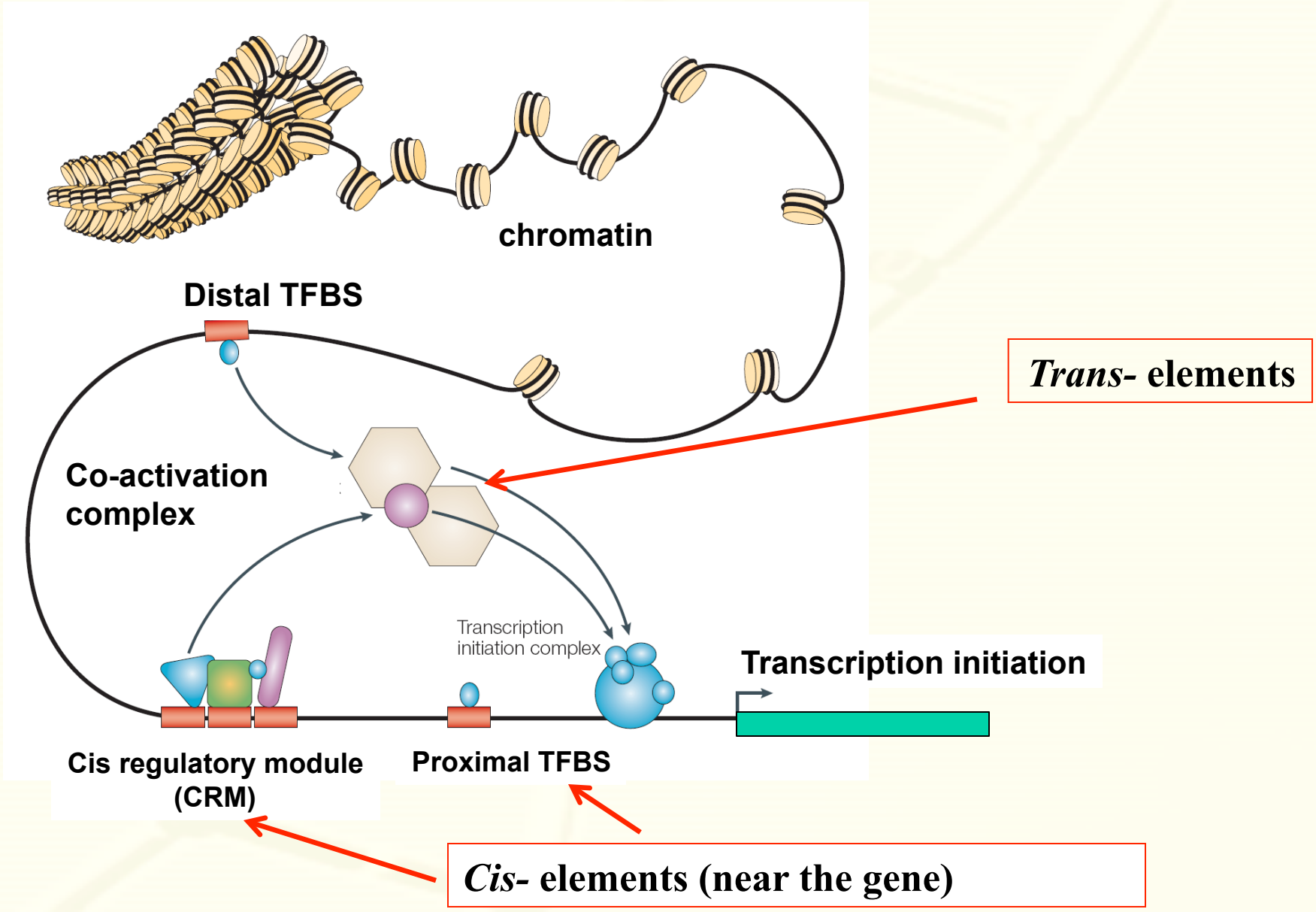




- **What factors can cause the discrepancy ?**
  - Codon usage
  - Post-transcriptional regulation: microRNAs, RNA binding proteins
  - mRNA degradation and half-life
  - Protein degradation and half-life

Vogel C. Mol Biosystems 2009

How are gene expression regulated ?

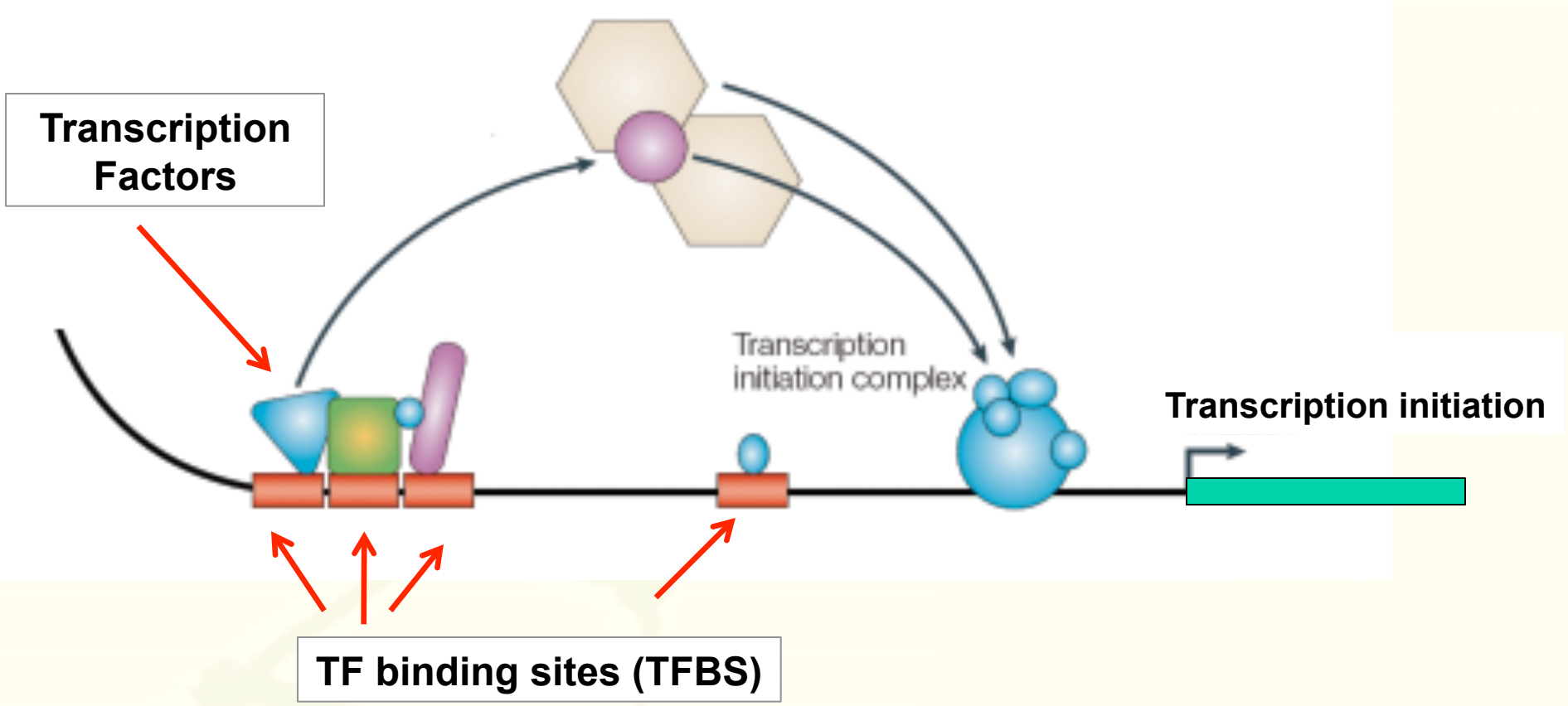


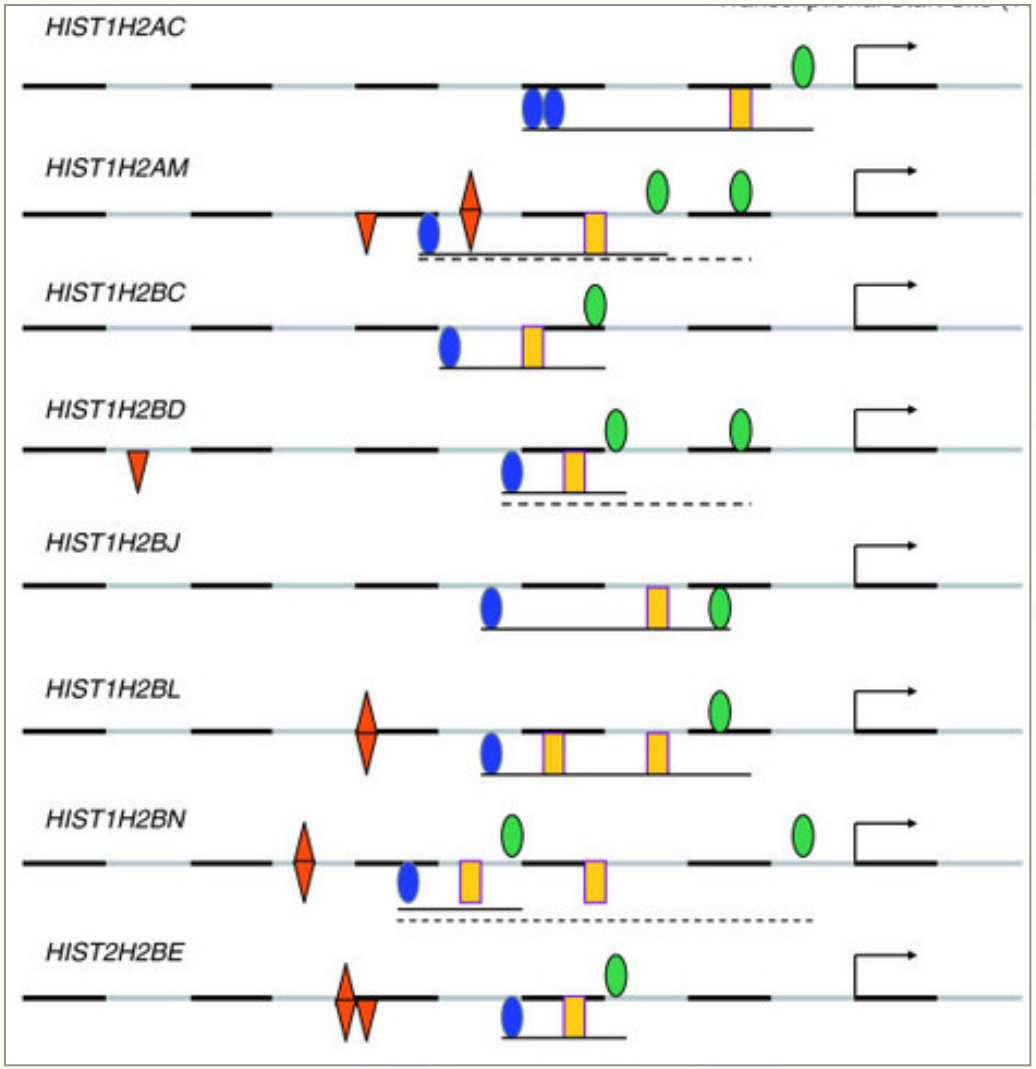
Trans- elements

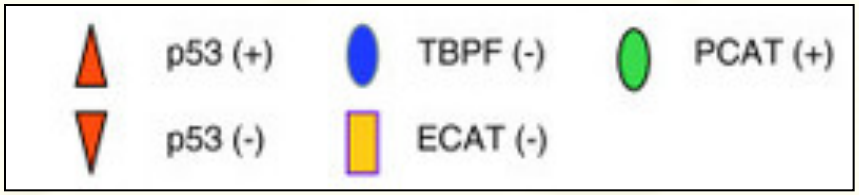
Cis- elements (near the gene)

Regulation by transcription factors (TF)

- Transcription factor – DNA interactions is the most important regulatory mechanism.







Regulation of histone variants by different combination of transcription factors

Different combination, spacing, position of TFs result in different timing, tissue of expression of these histone homologs

Yan et al Genome Biology 2007

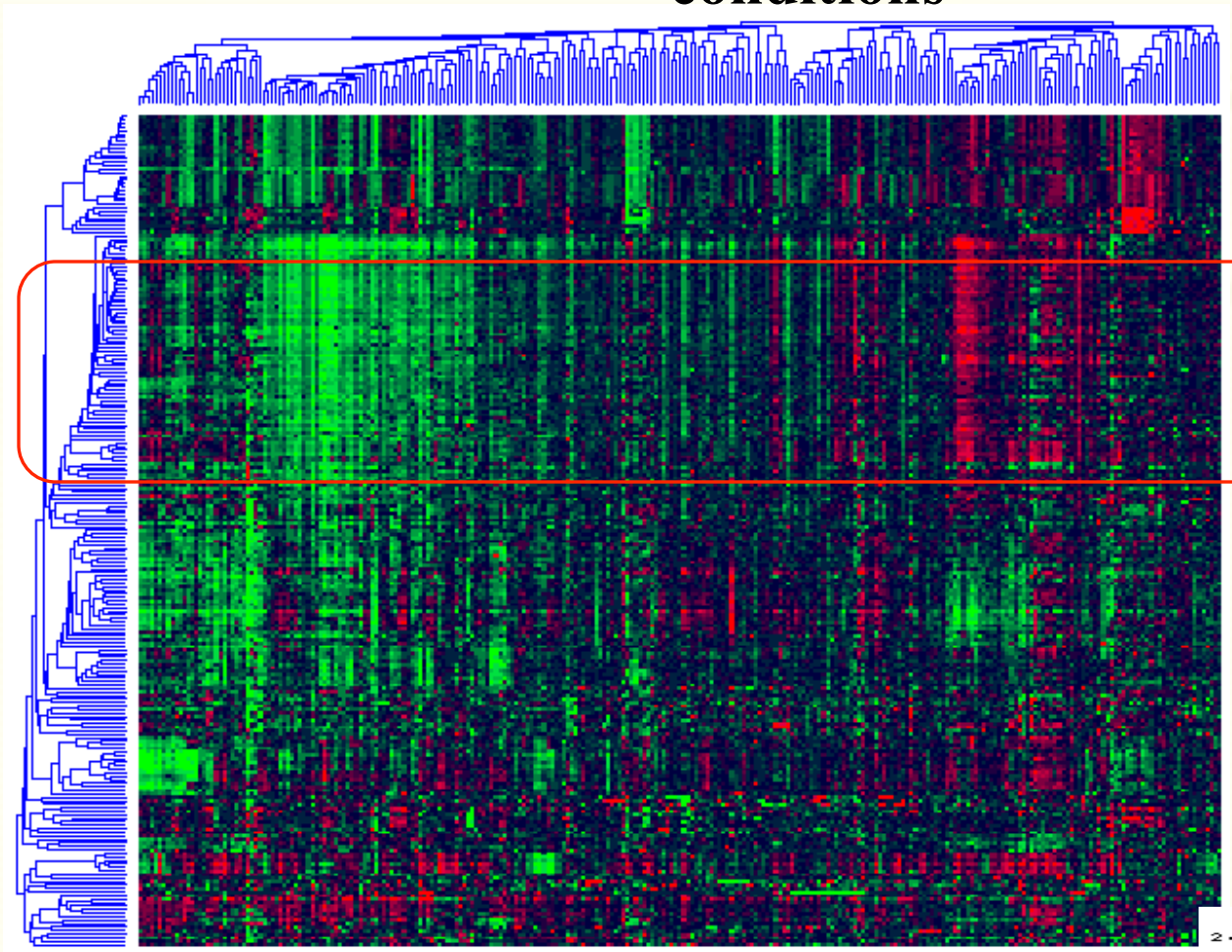
Computational prediction of TF binding sites

- **Using gene co-expression data:**
  - **Rationale:** genes sharing similar expression profile are likely to be regulated by the same transcription factors
  - Method: cluster these genes, align the promoter region, and look for enriched or conserved motifs

- **Using evolutionary conservation:**
  - **Rationale:** Orthologous genes from related organisms tend to be regulated by the same transcription factors
  - Methods: cluster these genes, align the promoter region, and look for enriched or conserved motifs

- These methods work okay for yeast but not for human or mouse, because of very large intergenic regions.

Enriched regulatory motifs in co-expressed genes

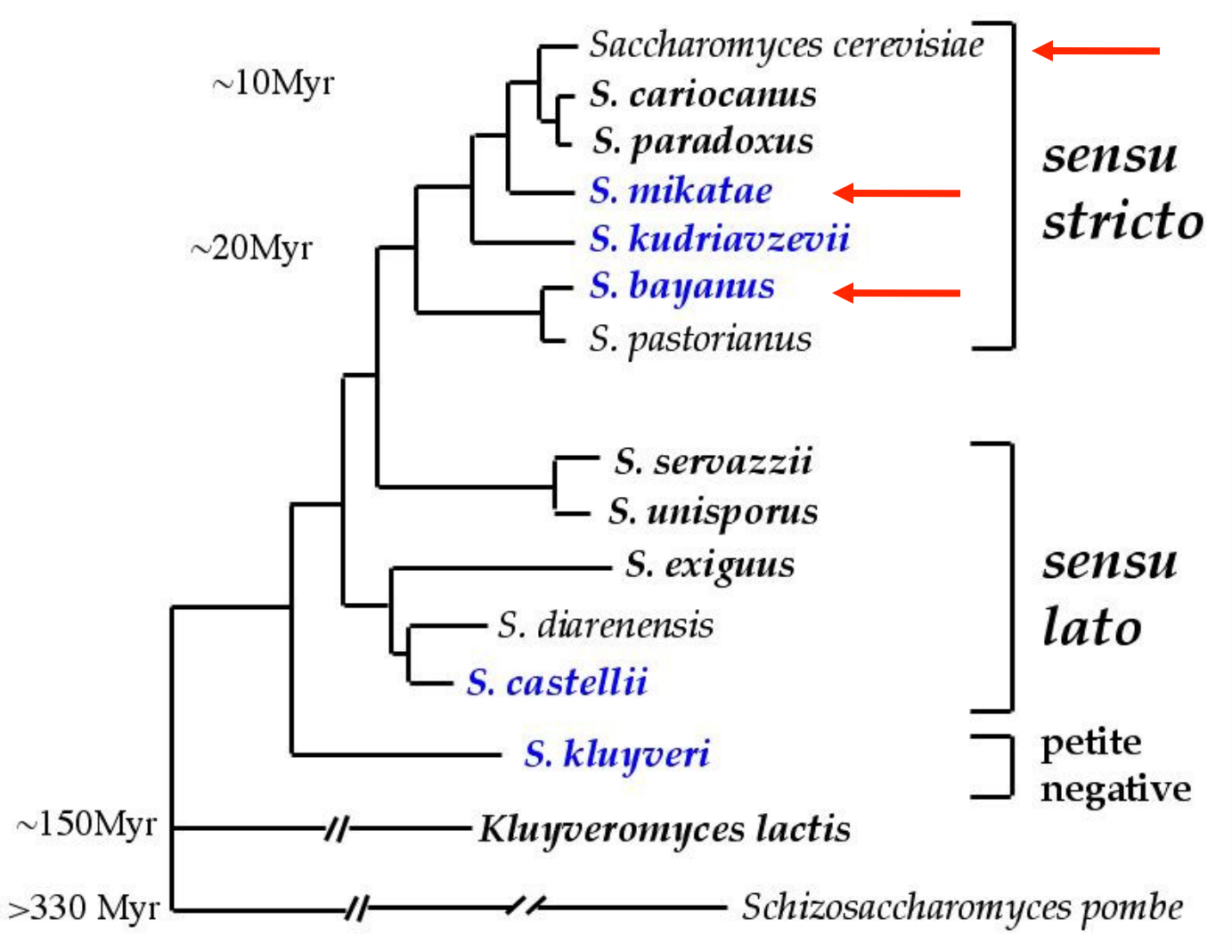


amino acid biosynthesis
amino acid metabolism

Enriched motif in promoters

From Tim Hughes

Finding regulatory elements by genome comparison



Finding regulatory elements by genome comparison

Nature **423**, 241-254 (15 May 2003) | doi:10.1038/nature01644; Received 27 February 2003; Accepted 1 April 2003

Sequencing and comparison of yeast species to identify genes and regulatory elements

Manolis Kellis[1,2], Nick Patterson[1], Matthew Endrizzi[1], Bruce Birren[1] & Eric S. Lander[1,3]

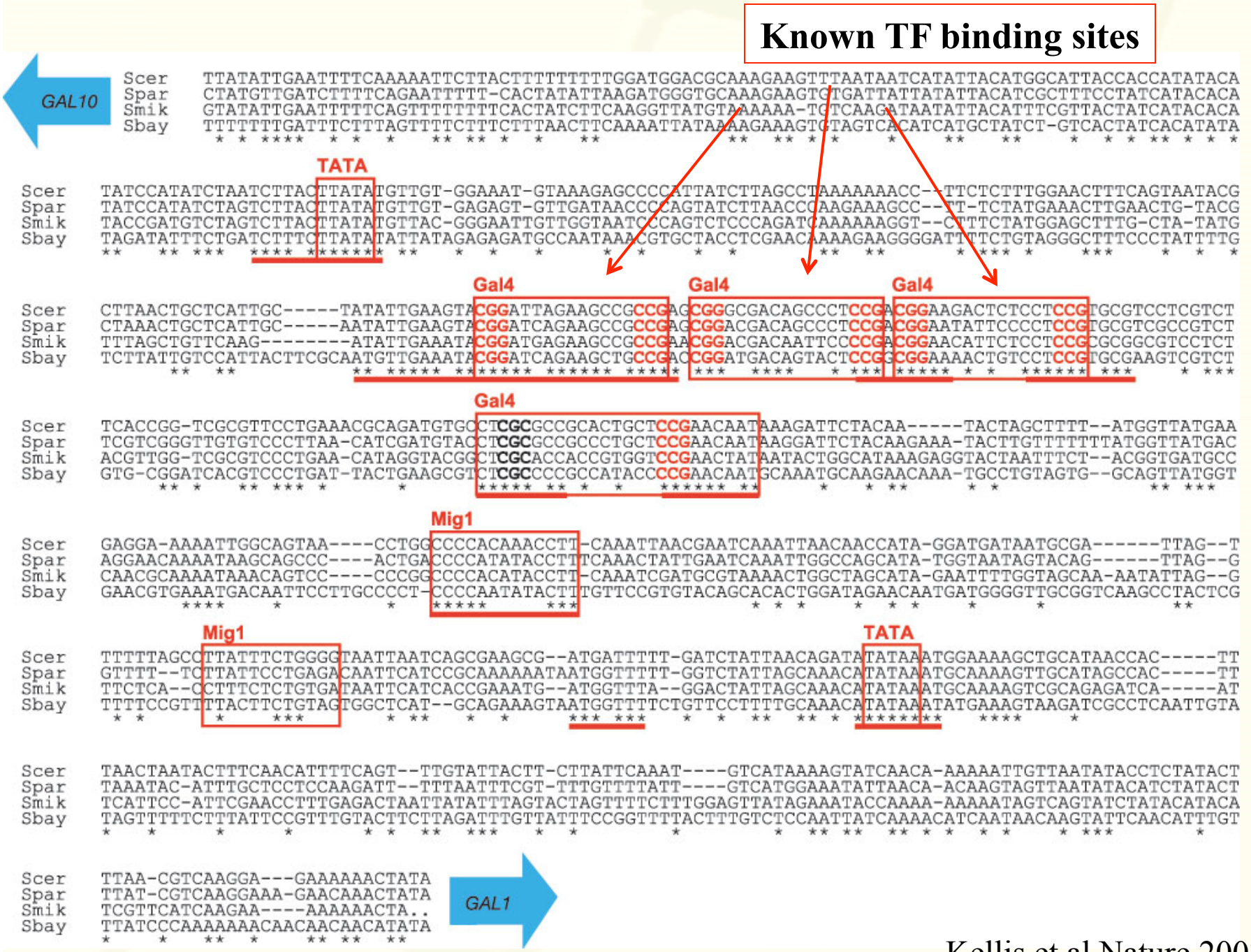
Science 4 July 2003:
Vol. 301 no. 5629 pp. 71–76
DOI: 10.1126/science.1084337

RESEARCH ARTICLE

Finding Functional Features in *Saccharomyces* Genomes by Phylogenetic Footprinting

Paul Cliften[1], Priya Sudarsanam[1], Ashwin Desikan[1], Lucinda Fulton[2], Bob Fulton[2], John Majors[3], Robert Waterston[1,2], Barak A. Cohen[1] and Mark Johnston[1,*]

- Kellis et al: … *We developed methods for direct identification of genes and regulatory motifs. …The gene analysis yielded a major revision to the yeast gene catalogue, affecting approximately 15% of all genes and reducing the total count by about 500 genes"*

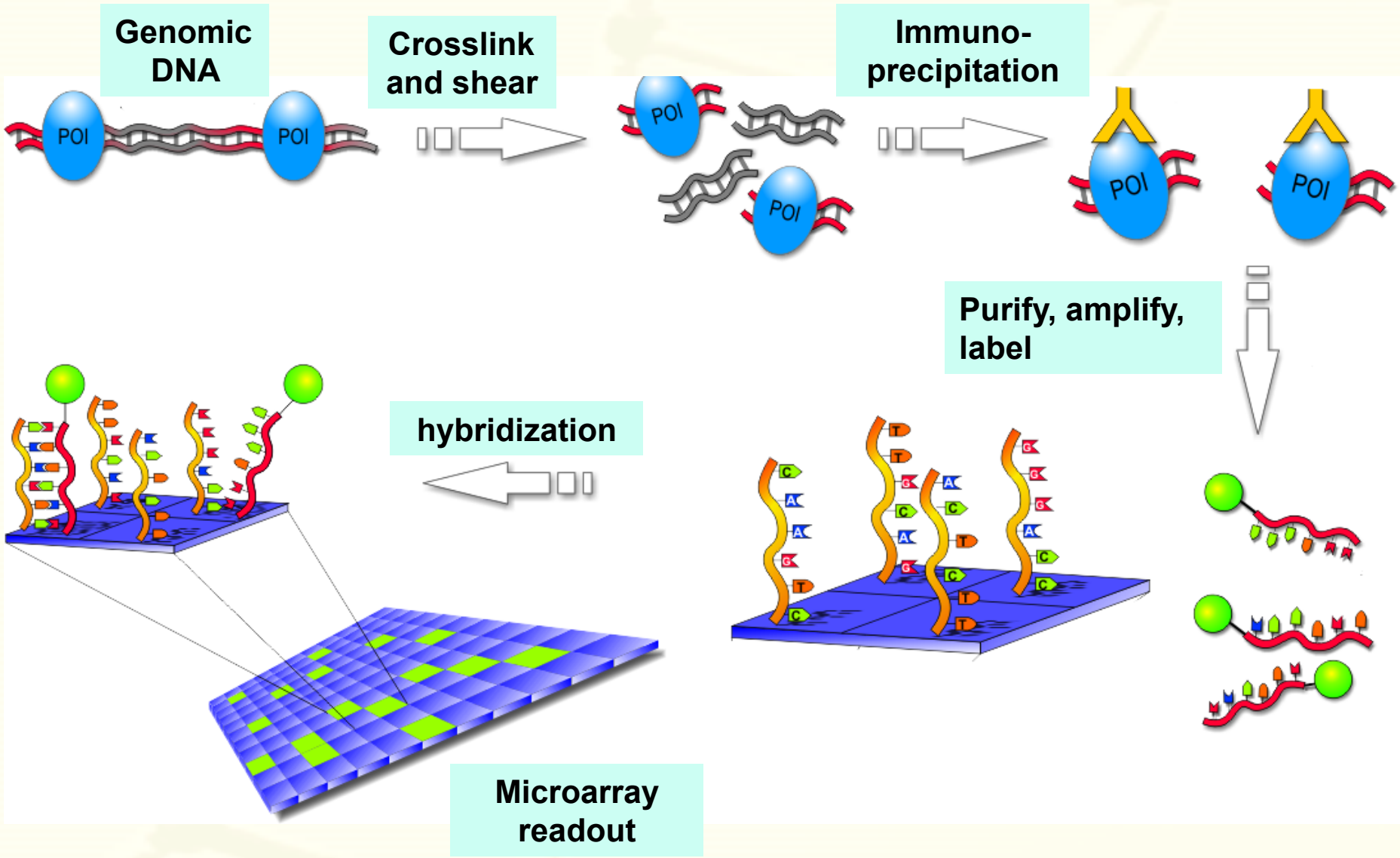


Kellis et al Nature 2003

Experimentally determine TF binding sites

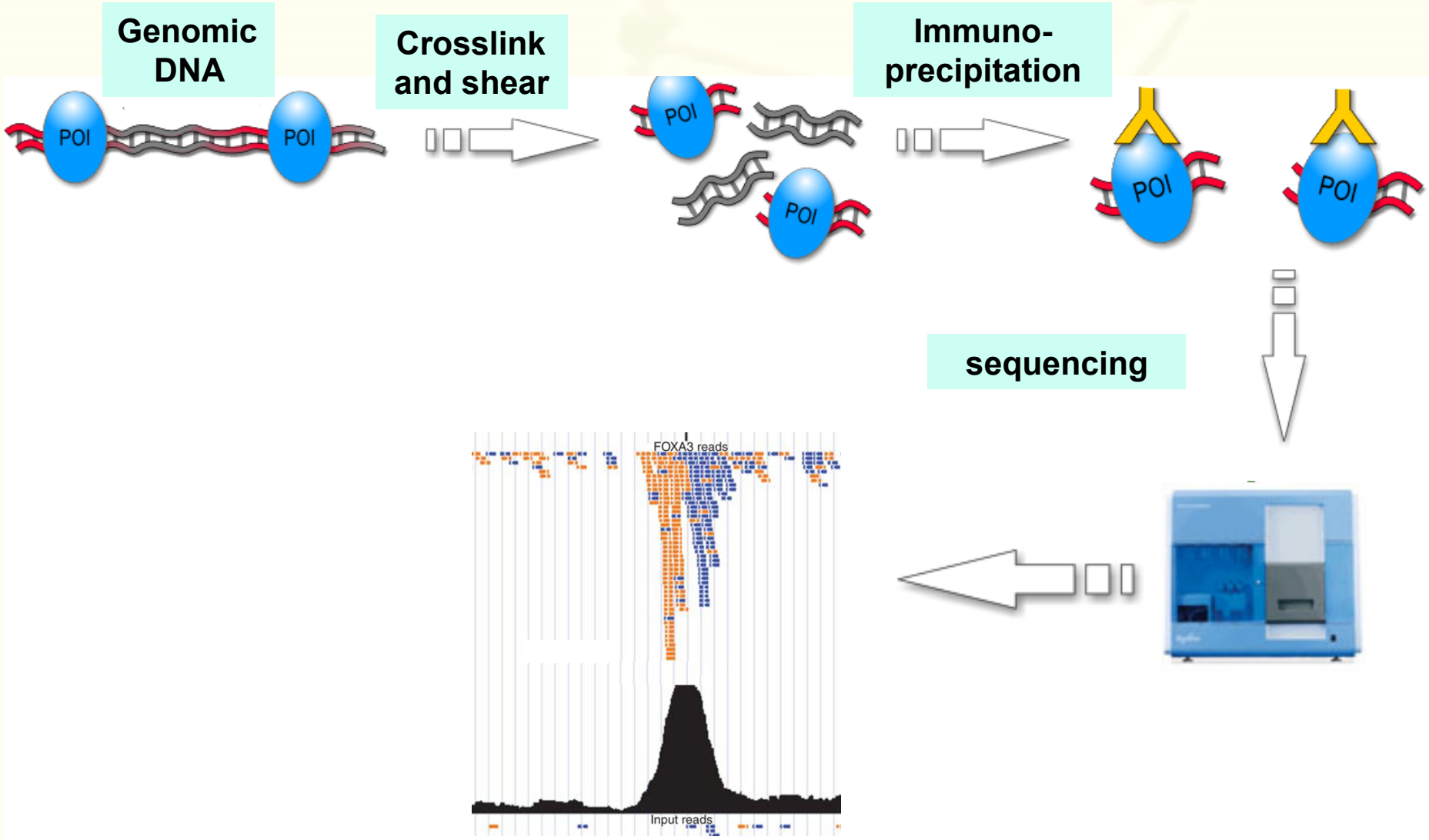
- **ChIP-chip (chromatin immunoprecipitation on a microchip)**
  - Directly detect which regions of the genome a TF binds to.
  - Use antibody to pull-down the TF protein and the bound DNA fragment, then use microarray or sequencing to map the fragments to the genome.

ChIP-chip



wikipedia

ChIP-seq

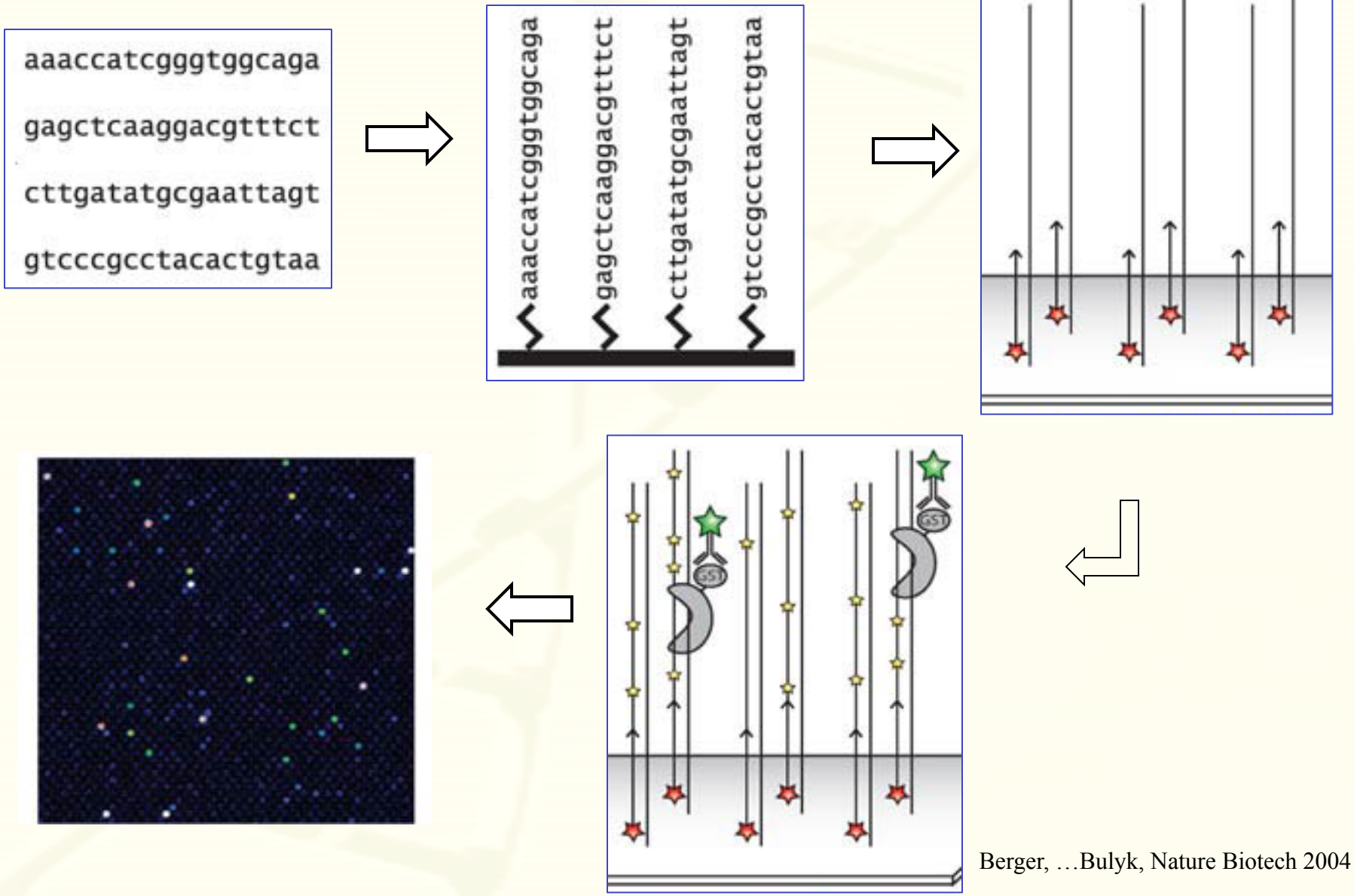


wikipedia

Experimentally determine TF binding sites

- **Protein binding array (PBM)**
  - Put all the possible 7-mer DNA fragments onto a microarray.
  - Hybridize a TF onto the array, wash out the un-bound protein, what is left is the DNA fragment bound by the TF.
  - Calculate a weight matrix from the bound fragments, and use it to scan the genome.

Protein binding microarray can determine DNA binding affinities



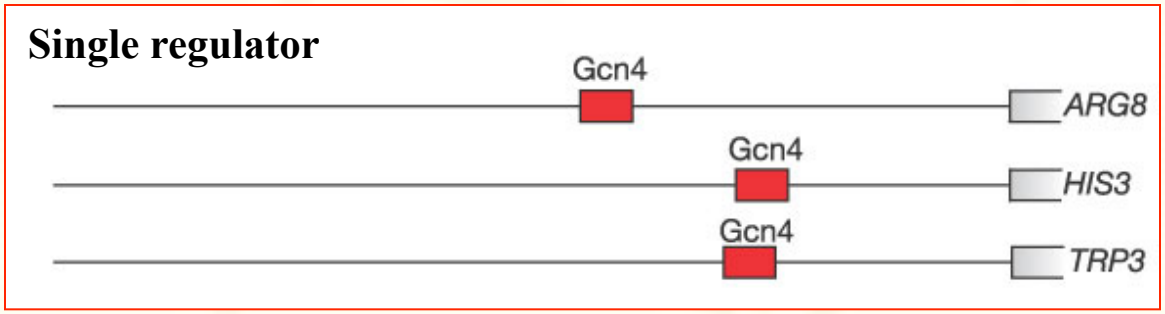
Berger, …Bulyk, Nature Biotech 2004

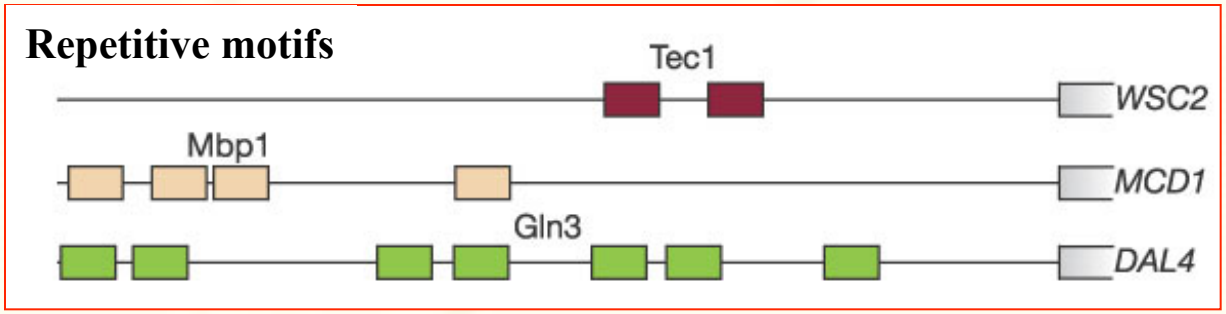
Two landmark papers

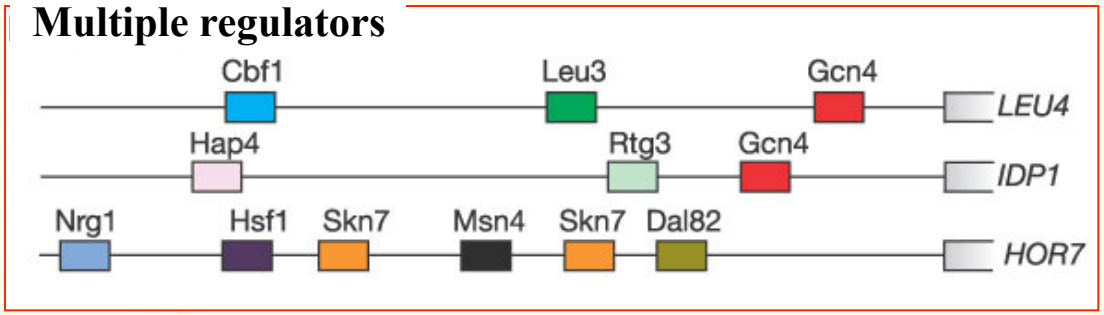
- Lee, … Frankel, Gifford, Young "**Transcriptional Regulatory Networks in Saccharomyces cerevisiae"** Science 2002
  - "We use this information to identify **network motifs**, the simplest units of network architecture, …. Our results reveal that eukaryotic cellular functions are **highly connected through networks of transcriptional regulators that regulate other transcriptional regulators."**

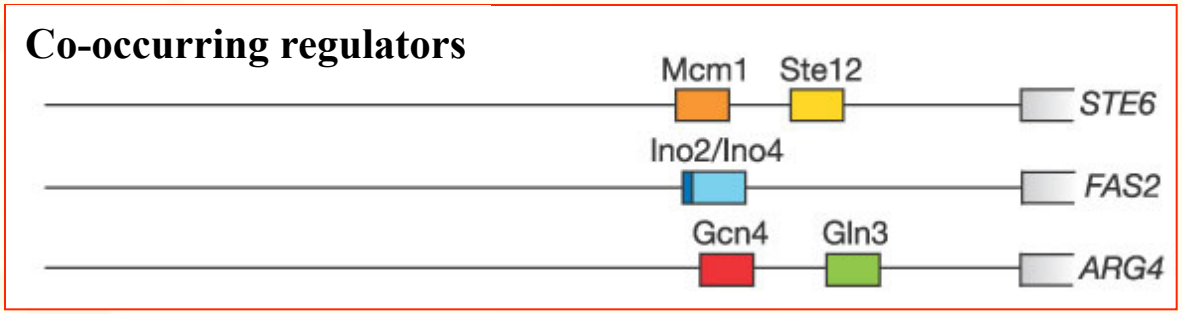
- Harbinson, … Frankel, Gifford, Young, "**Transcriptional regulatory code of a eukaryotic genome**" Nature 2004
  - "We have constructed an initial map of yeast's transcriptional regulatory code by identifying the sequence elements that are **bound by regulators under various conditions** and that **are conserved** among Saccharomyces species…We find that **environment-specific use** of regulatory elements predicts mechanistic models for the function of a large population of yeast's transcriptional regulators."

Yeast promoter architecture: different mode of regulation by transcription factors



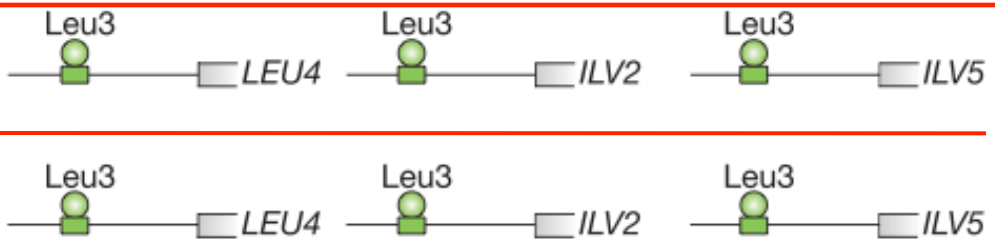






Harbinson, … R. Young,
Transcriptional regulatory code of a eukaryotic genome
Nature 2004

Condition invariant



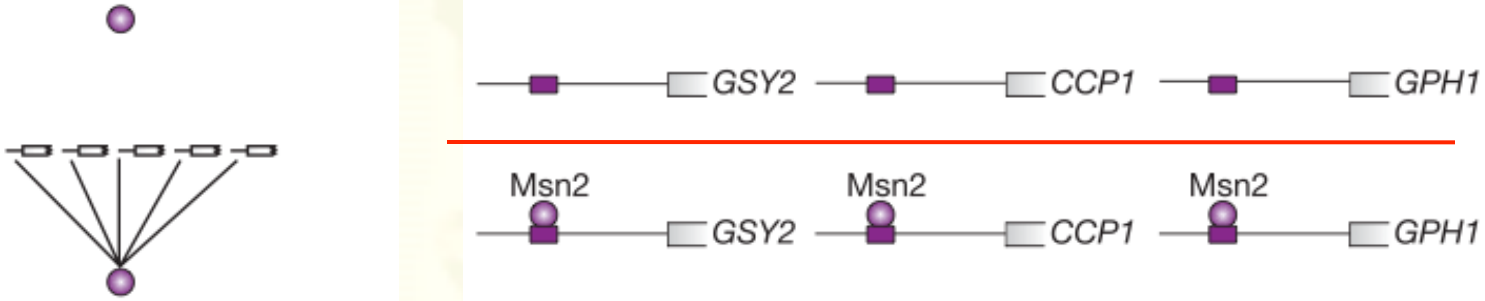
Condition enabled

Condition expanded

Condition altered

Condition invariant

Condition enabled



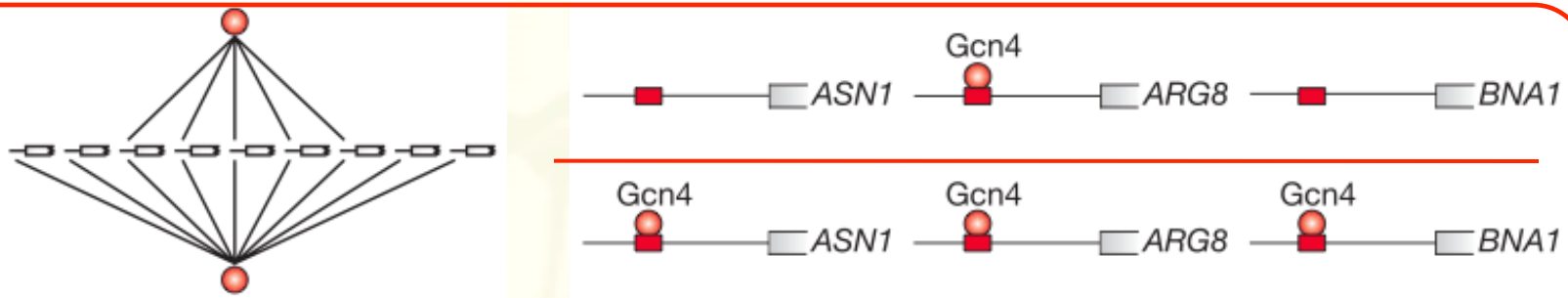
Condition expanded

Condition altered

Condition invariant

Condition enabled

Condition expanded



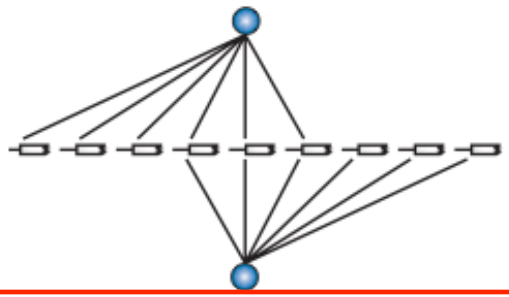
Condition altered

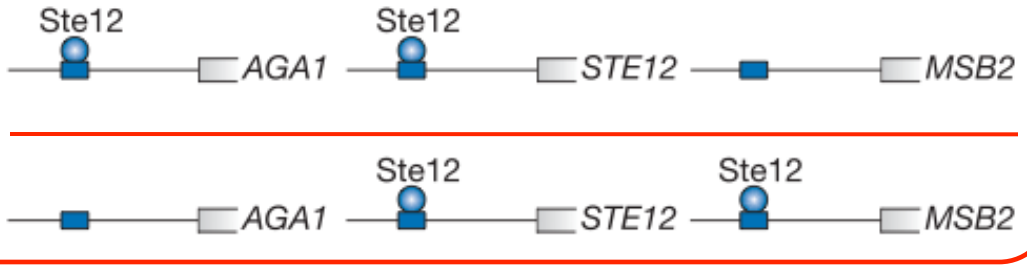
Condition invariant

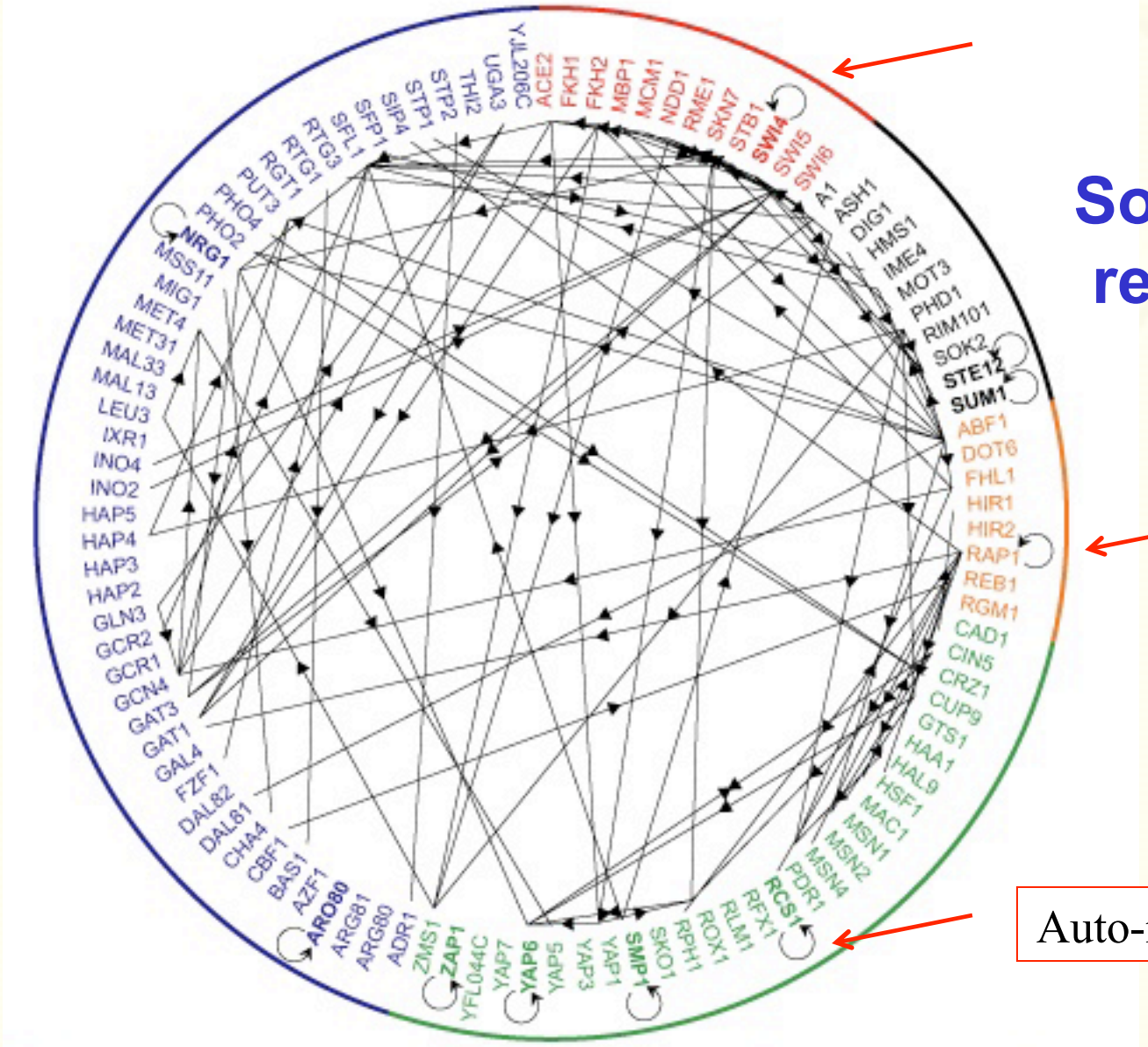
Condition enabled

Condition expanded

Condition altered

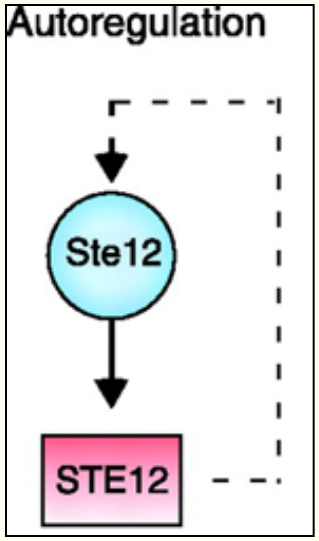


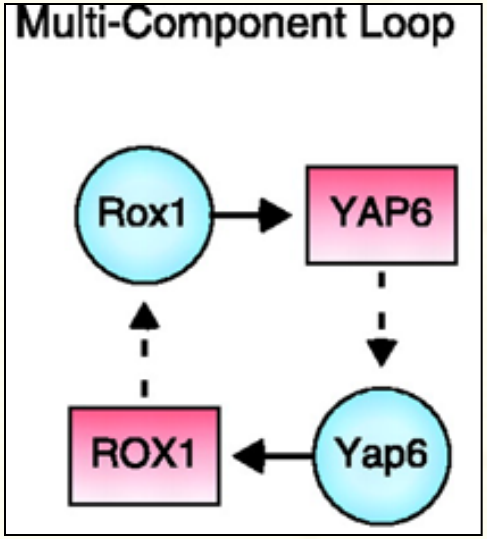


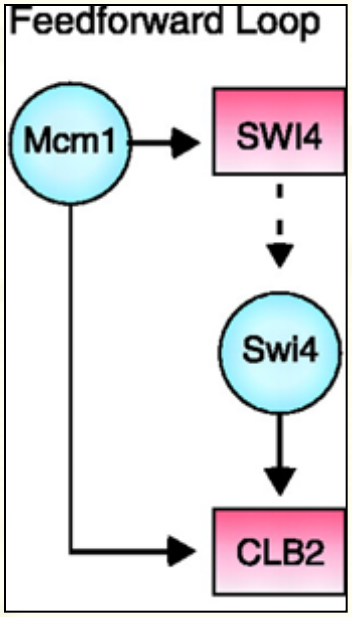


Some TFs are regulated by other TFs

Auto-regulatory

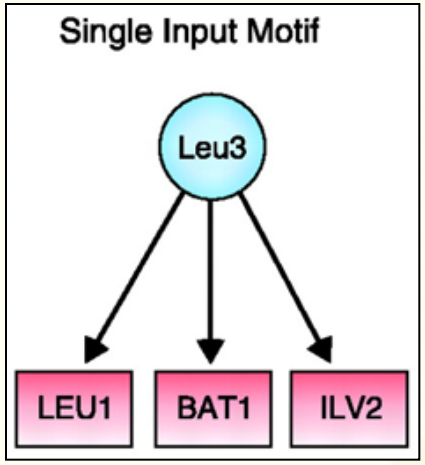


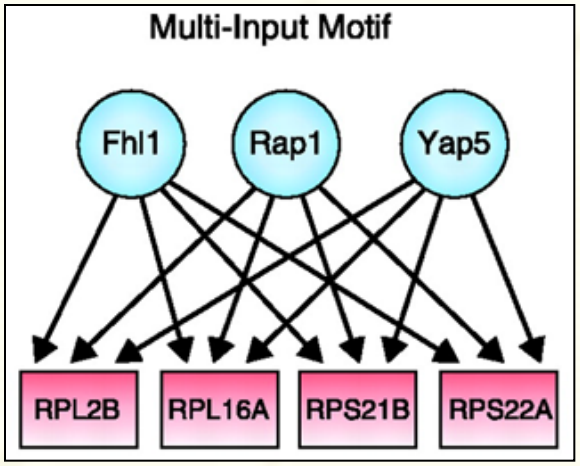




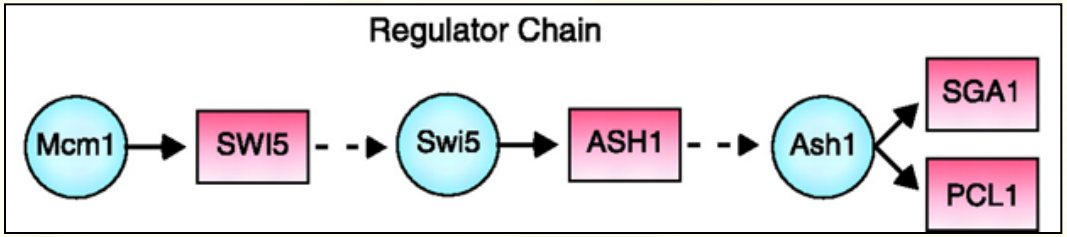
TF

Target gene



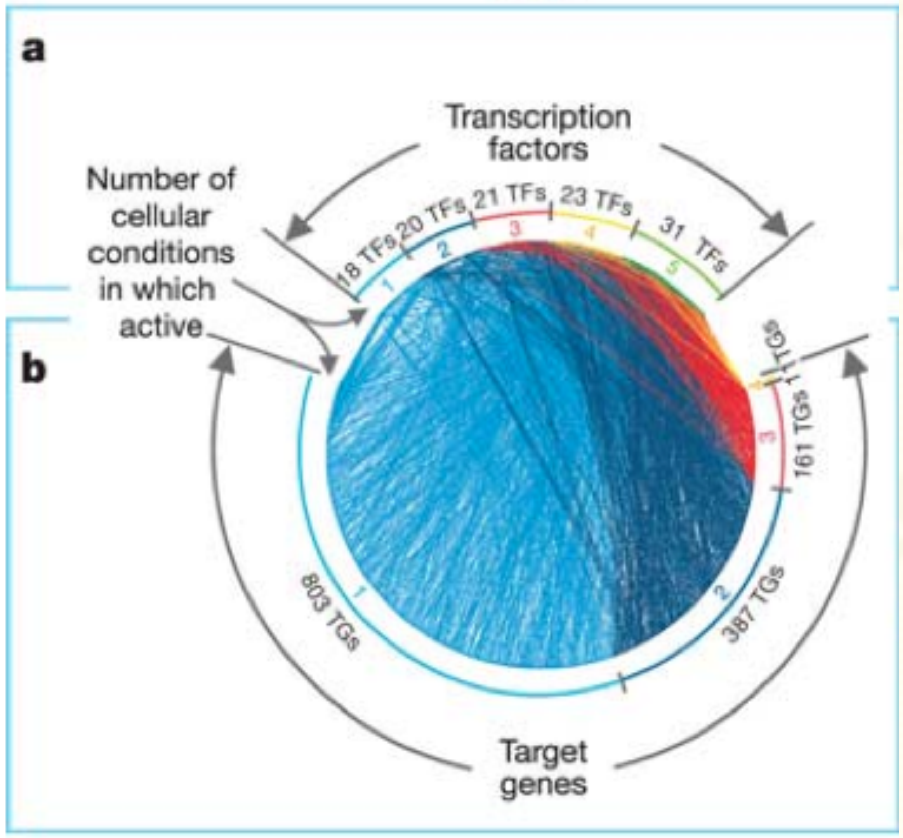


Regulatory network motifs



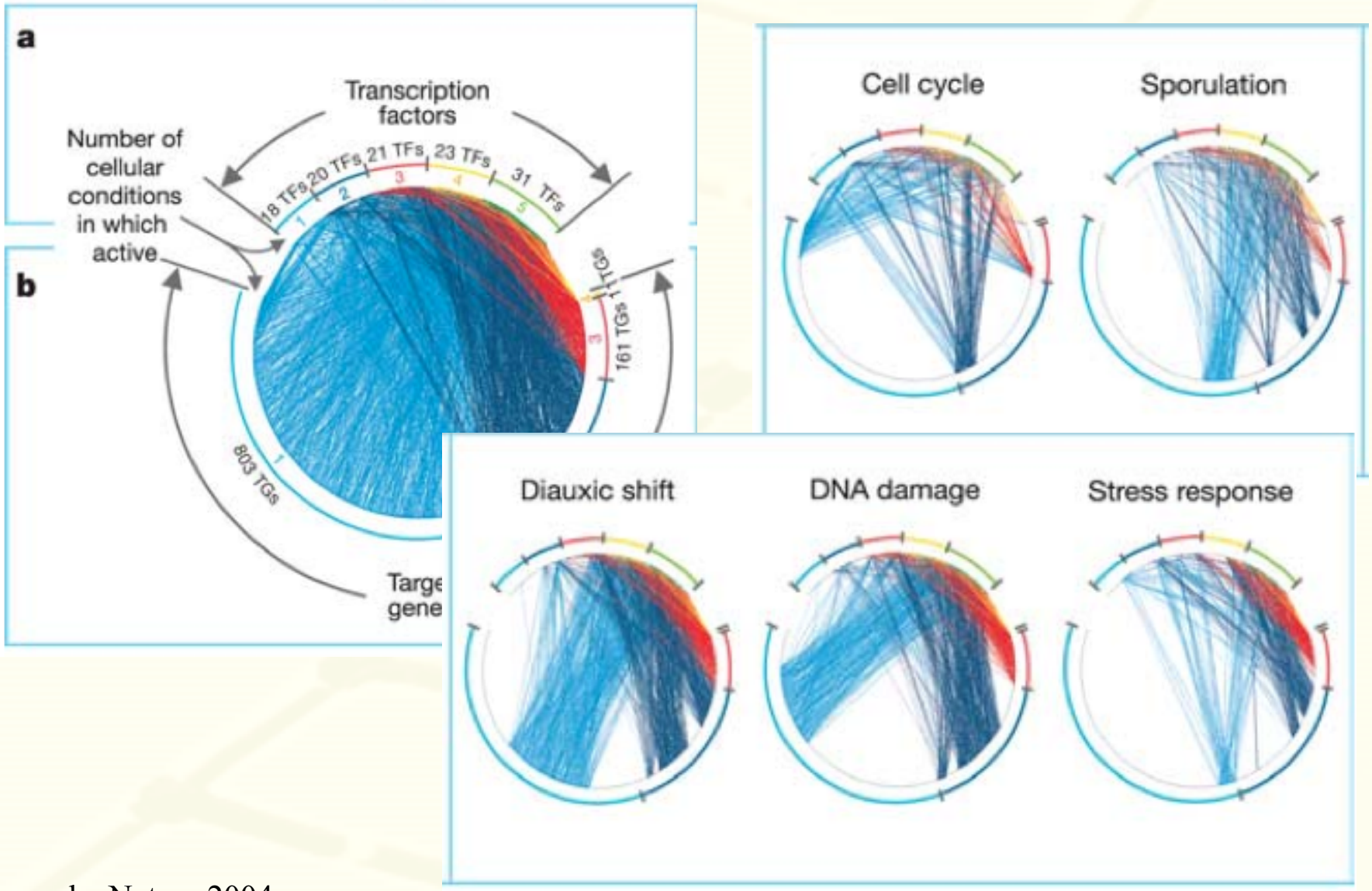
Lee, … Young Science 2002,
Transcriptional Regulatory Networks in S. cerevisiae

Combine regulatory network with gene expression – dynamics of the network



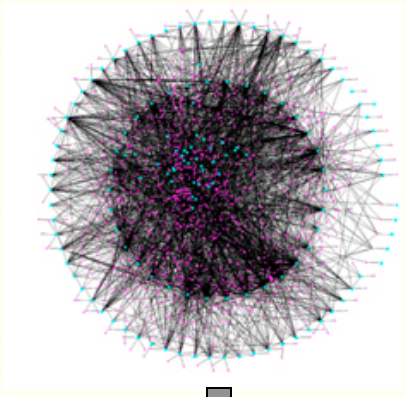
Luscombe Nature 2004

Combine regulatory network with gene expression – dynamics of the network

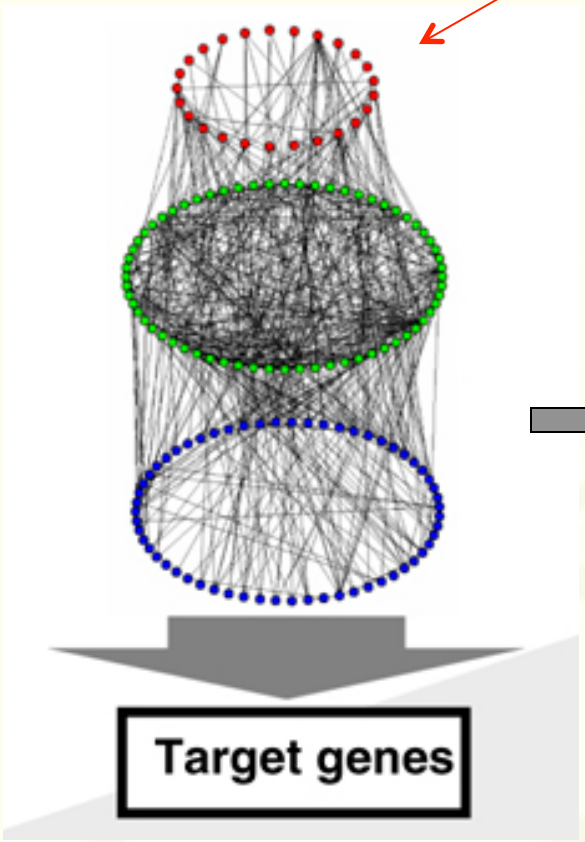


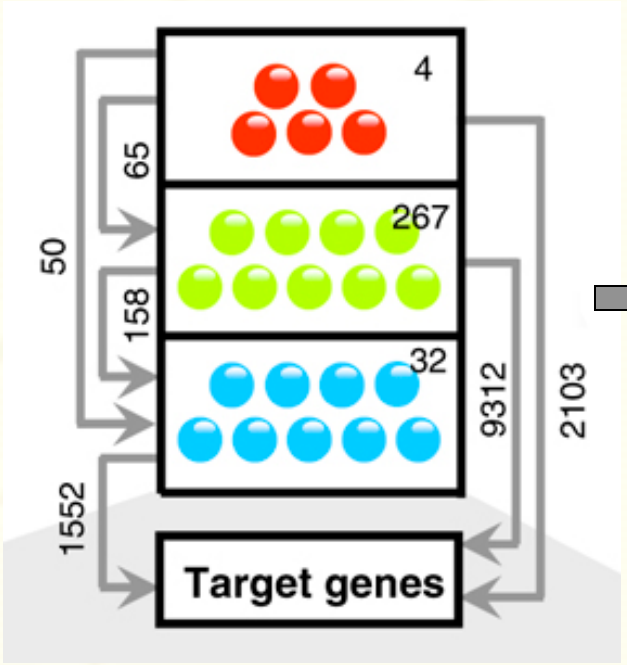
Luscombe Nature 2004

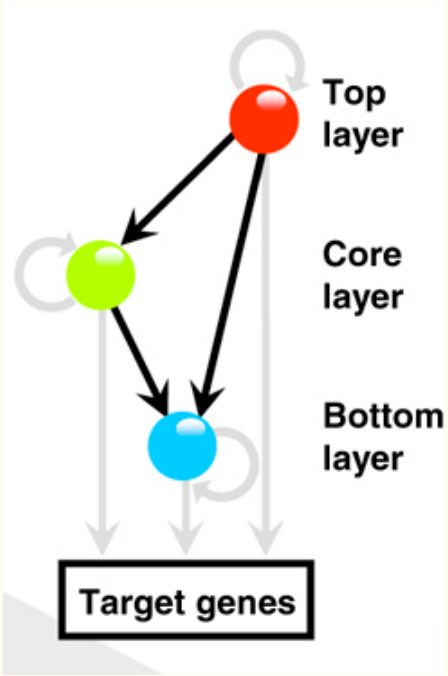
Hierarchical structure of regulatory network (?)



Master regulators







Jothi et al Mol Sys. Biol 2009

Some evolutionary questions

- For the TFs that have many targets and those that have few targets, do they have different evolutionary constraints on their sequence and mRNA expression level ?

- For the target genes that have many regulators, and those have just one regulator, do they have different evolutionary constraints on their expression and sequence ?

- What happens if a regulator is duplicated ?
- What happens if a target is duplicated ?
- What happens if the entire genome is duplicated ?

more evolutionary questions

- How did a target gene become regulated by a TF ?
  - Two alternative ways:
  - By creating a *cis*- motif (binding site) that can be recognized an existing TF, or
  - By changing the TF protein sequence and structure, which leads to recognition of an existing DNA motif

- Are changes in *trans*- elements (TF) more important than *cis*- elements (binding sites) for the evolution of expression and the organism ? (see next slide)

- How conserved are the regulatory mechanisms for orthologous genes in different organisms ? (see next slide)

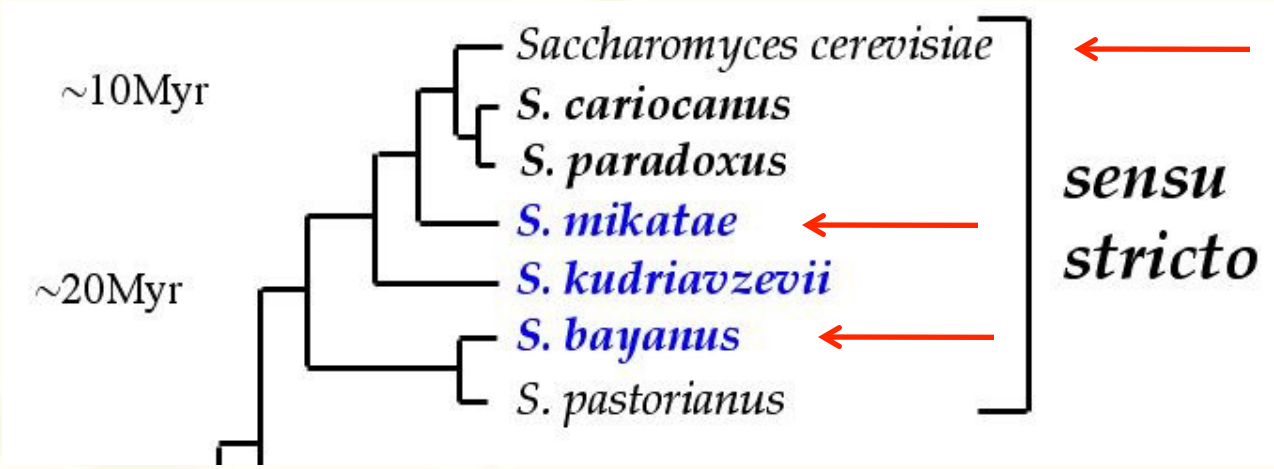
Divergence of TF binding sites between species

Science 10 August 2007:

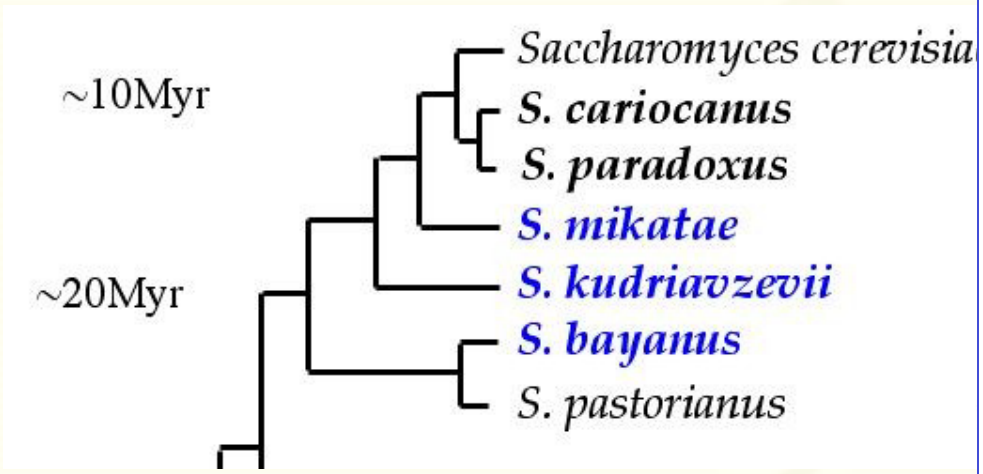
Divergence of Transcription Factor Binding Sites Across Related Yeast Species

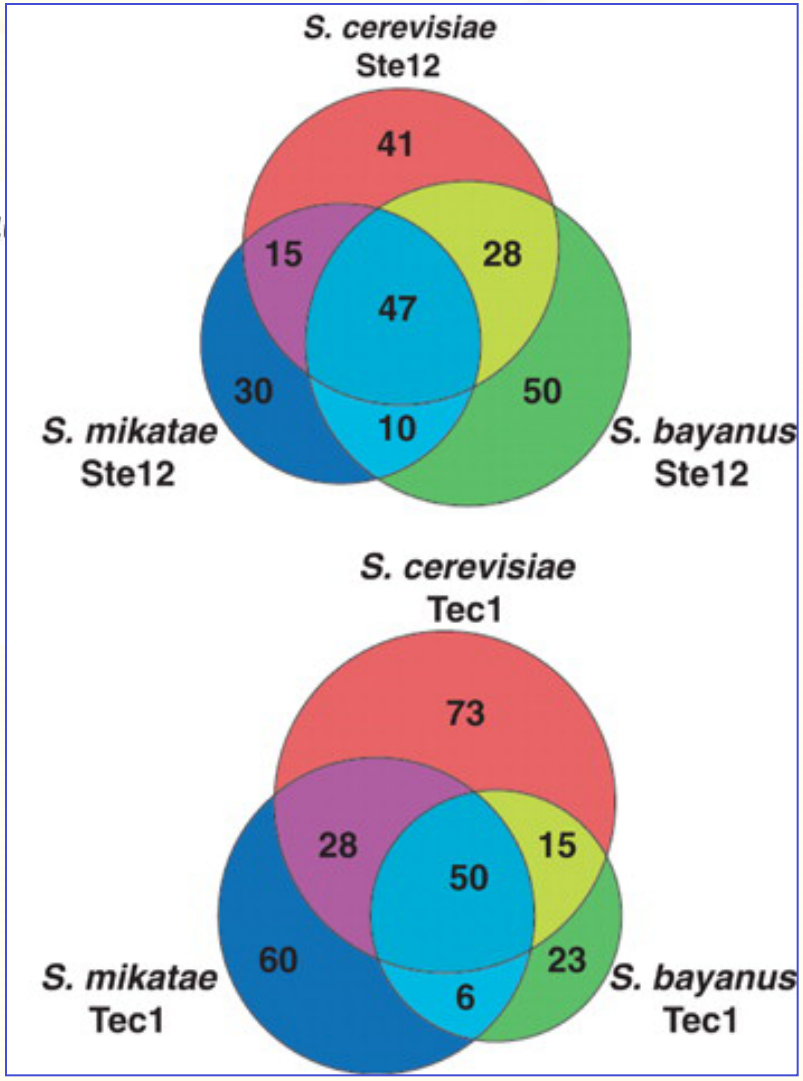
Anthony R. Borneman[1,*], Tara A. Gianoulis[2], Zhengdong D. Zhang[3], Haiyuan Yu[3], Joel Rozowsky[3], Michael R. Seringhaus[3], Lu Yong Wang[4], Mark Gerstein[2,3,5] and Michael Snyder[1,2,3,†]

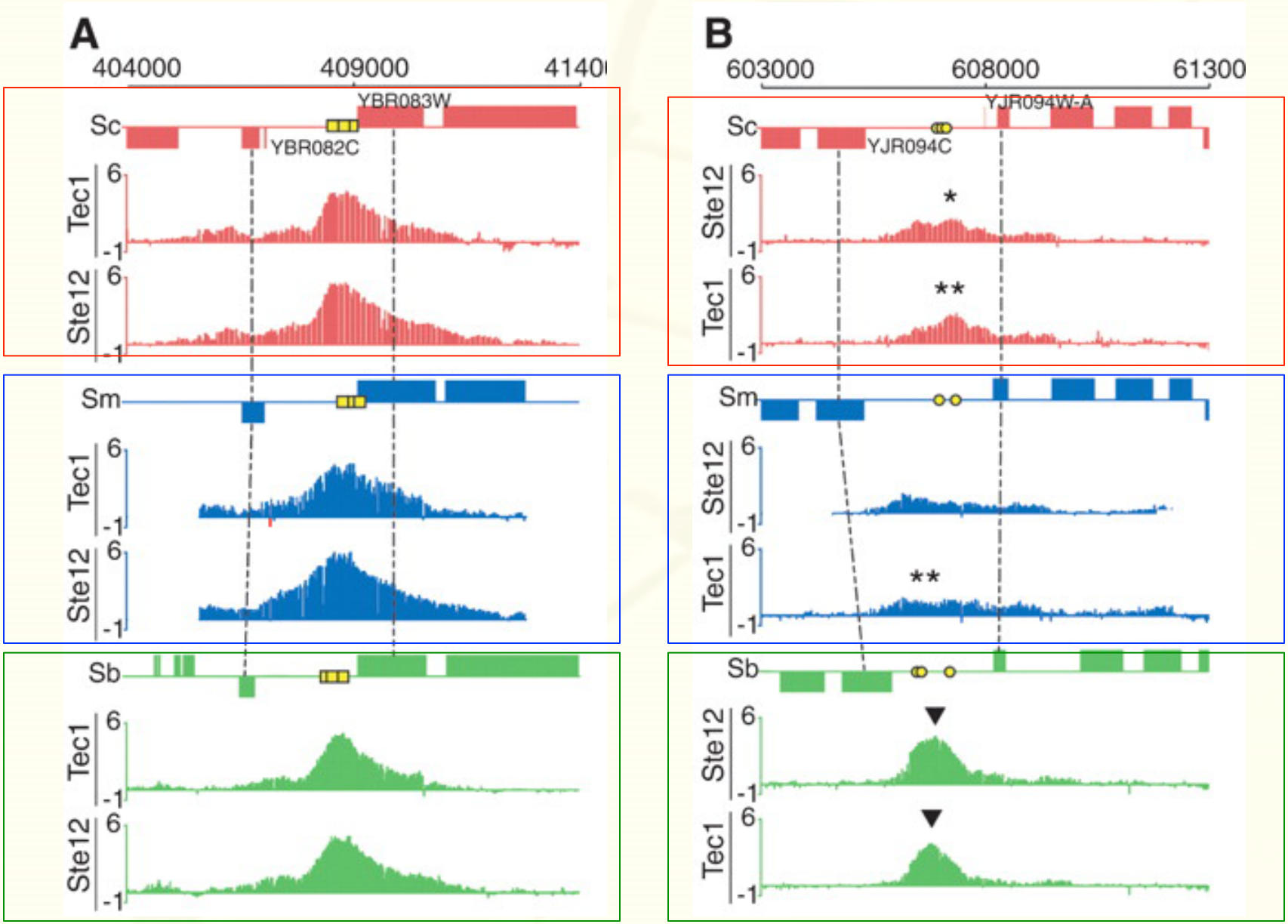
- The authors used ChIP-chip and determined the binding sites of Ste12 and Tec1 in 3 yeasts. These TFs bind collaboratively to targets.



Divergence of TF binding sites between species

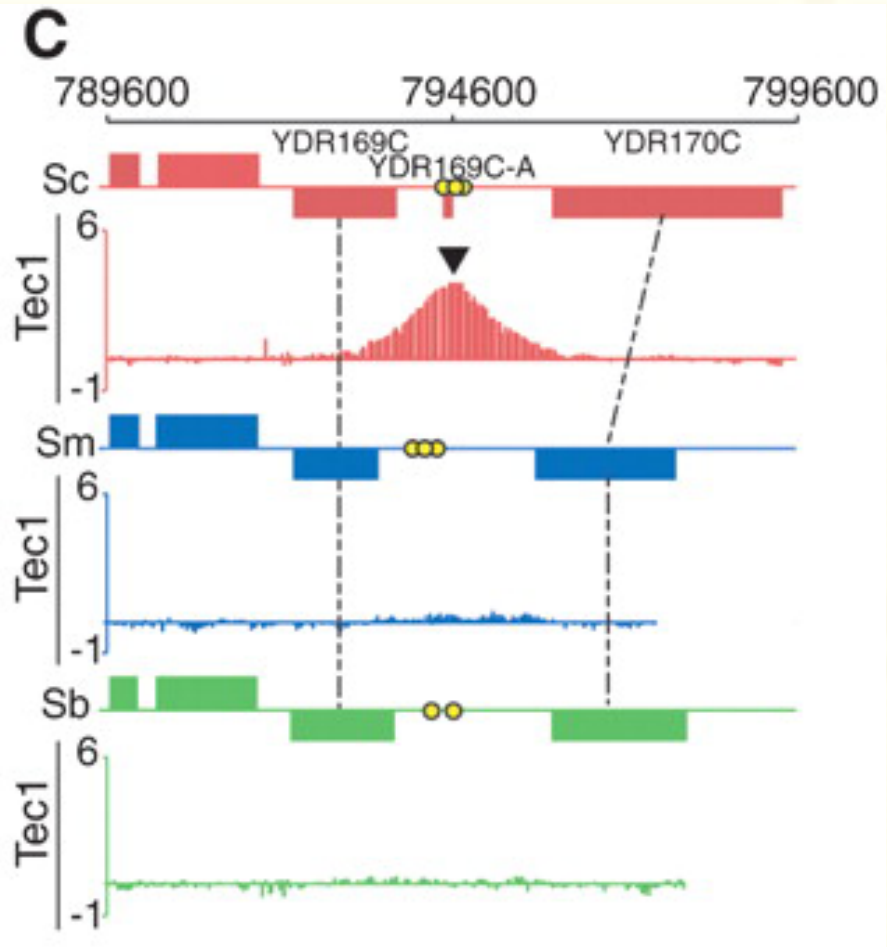


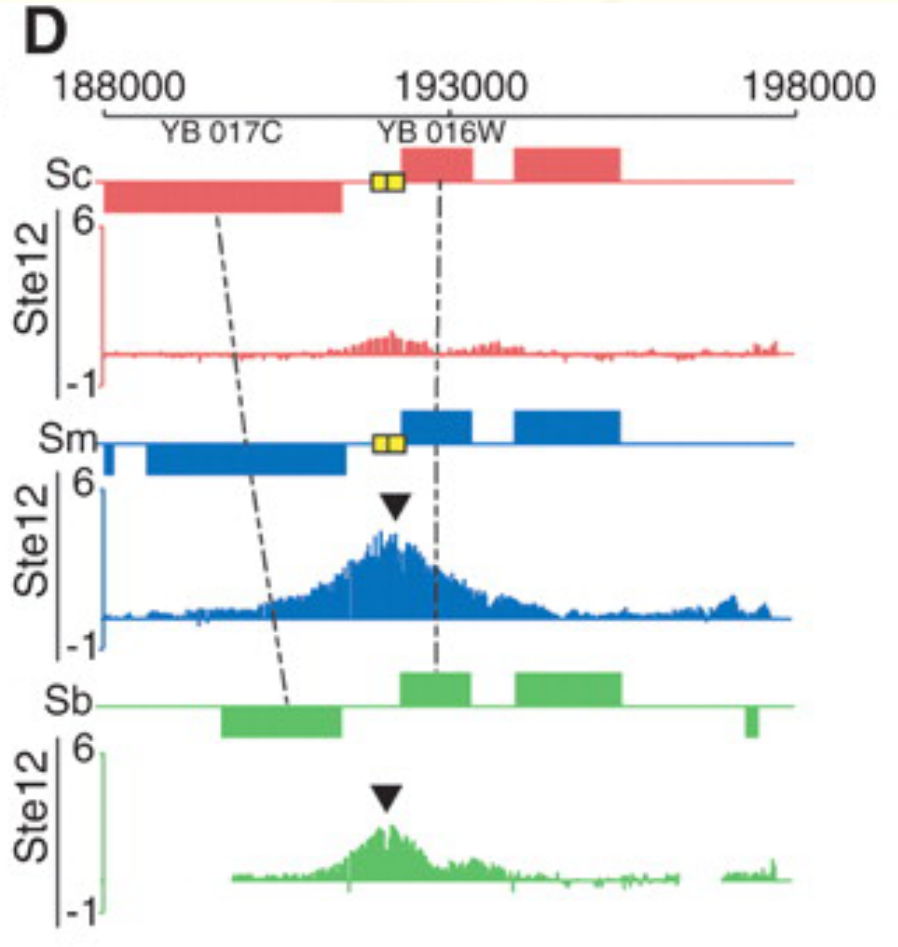




Conserved site, conserved binding

Conserved site, different binding intensity





Conserved site, no binding

No site, observed binding

Divergence of TF binding sites between species

- **Summary of observations:**
- Orthologous genes often regulated by the same transcription factors, but their binding sites are often too diverged to be recognized by sequence.

- "Transcription factor binding sites have therefore diverged substantially faster than gene sequence. Thus, gene regulation resulting from transcription factor binding is likely to be a major cause of divergence between related species."

- Conclusion: Changes in TF binding sites are more common and more important in gene regulation than changes in TFs.

Variation of TF binding sites between strains

Nature **464**, 1187-1191 (22 April 2010) | doi:10.1038/nature08934; Received 13 December 2009; Accepted 19 February 2010; Published online 17 March 2010

Genetic analysis of variation in transcription factor binding in yeast

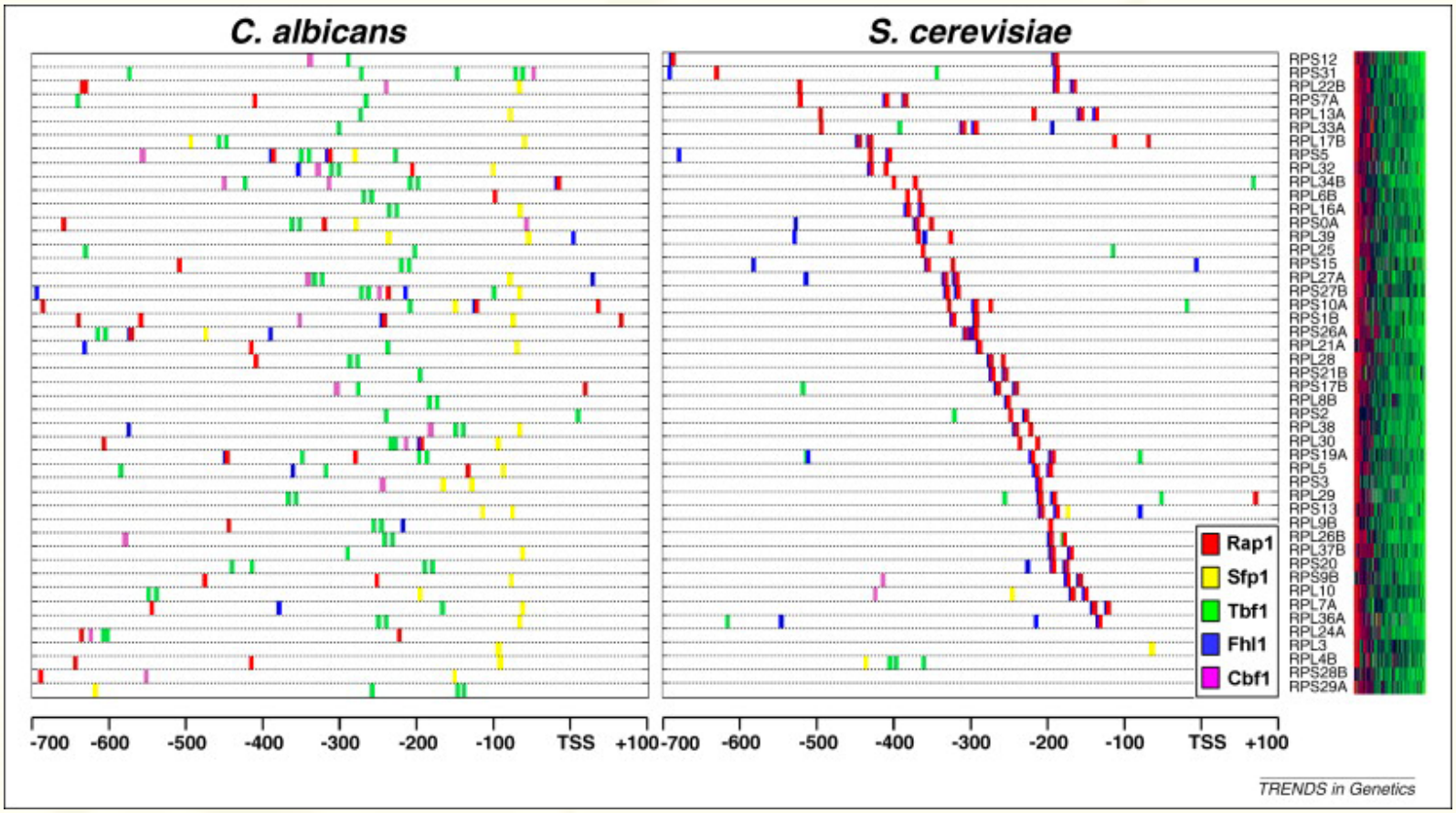
Wei Zheng[1,6], Hongyu Zhao[2,3], Eugenio Mancera[4], Lars M. Steinmetz[4] & Michael Snyder[1,5]

- The same authors looked at variation among different S. crevisiae strains instead of different species, and used ChIP-seq instead of ChIP-chip.
- *"We showed that most transcription factor binding variation is cis- linked, and that many variations are associated with polymorphisms residing in the binding motifs of Ste12."*

But there are exceptions …

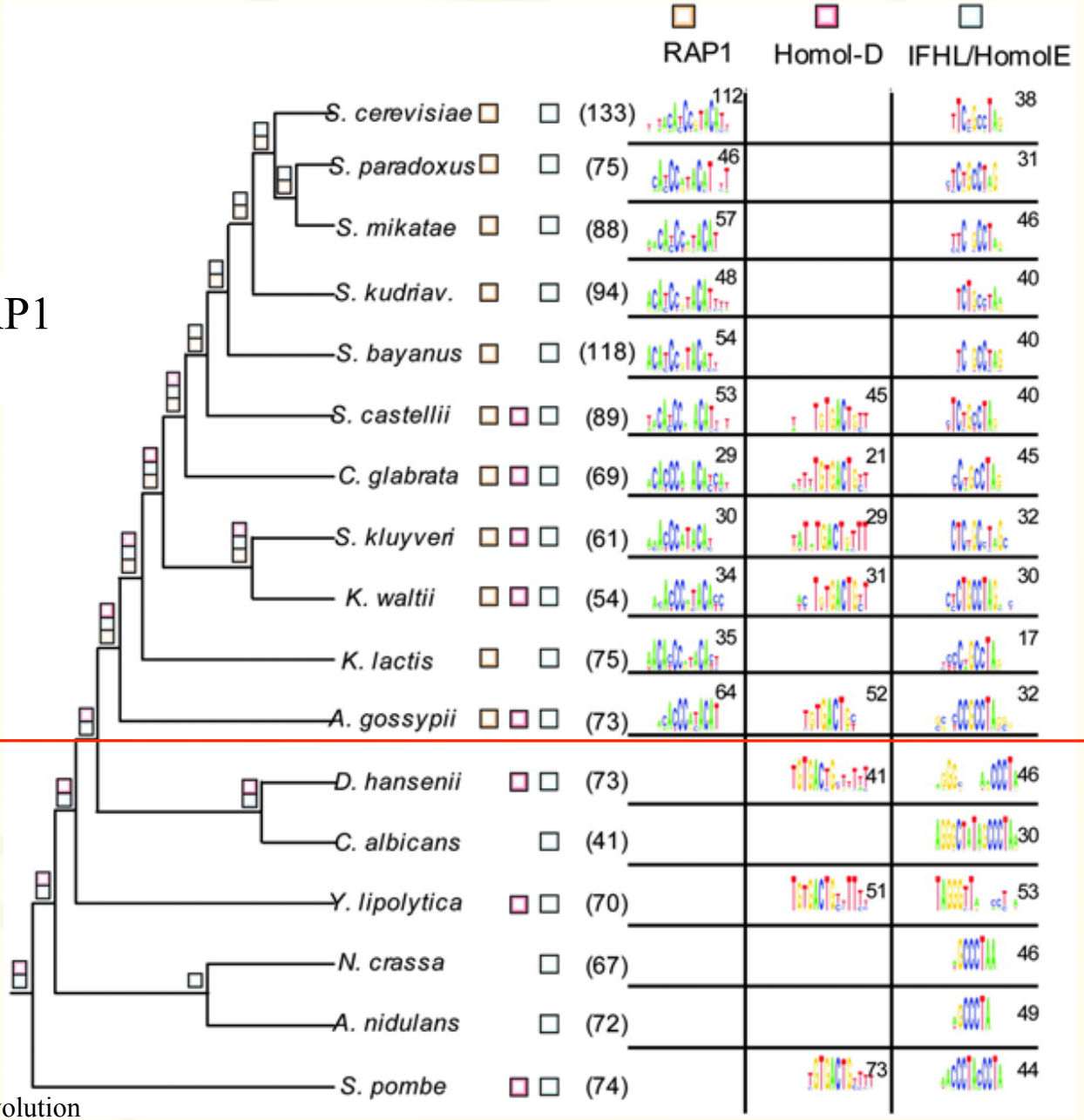
Regulation of ribosomal proteins: an example of transcription factor switching in evolution

- Ribosomal proteins are highly regulated, however they are primarily regulated by **Rap1** in *S. cerevisiae* and **Tbf1** in *C. albicans*.



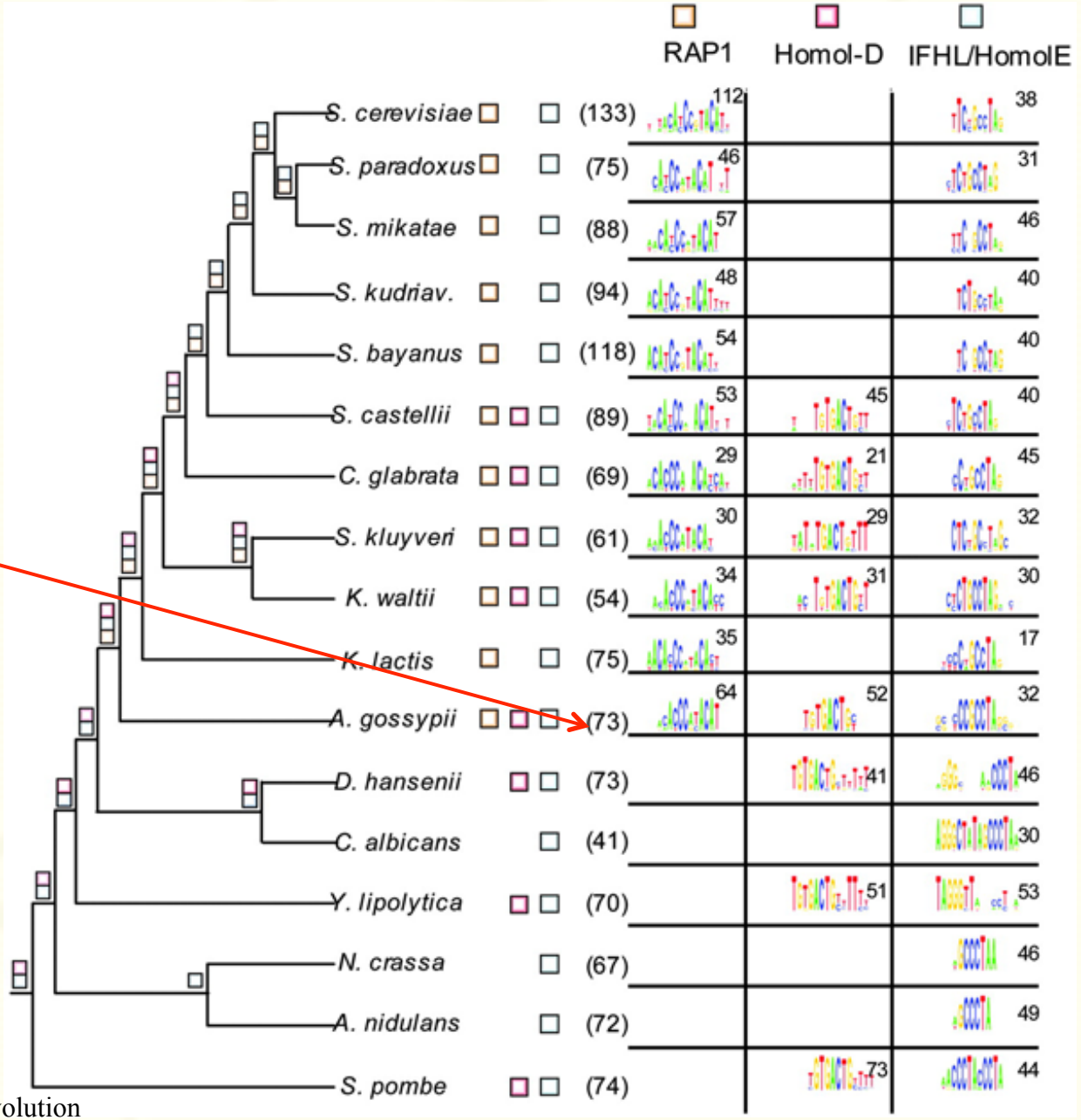
Weirauch, Hughes Trends in Genetics 2010

Binding matrix for RAP1 and Tbf1

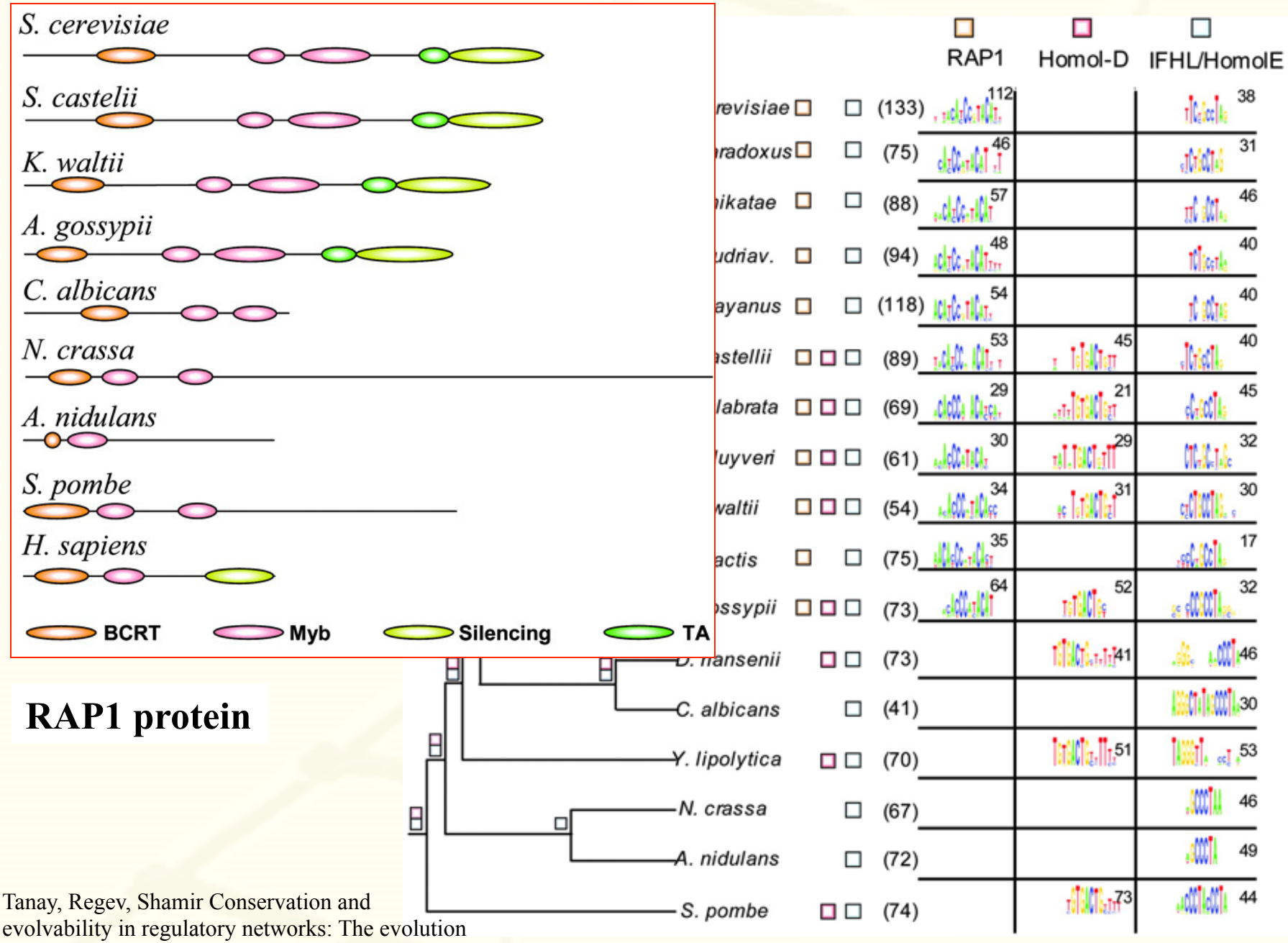


Tanay, Regev, Shamir Conservation and evolvability in regulatory networks: The evolution of ribosomal regulation in yeast PNAS 2005

What happened here ?

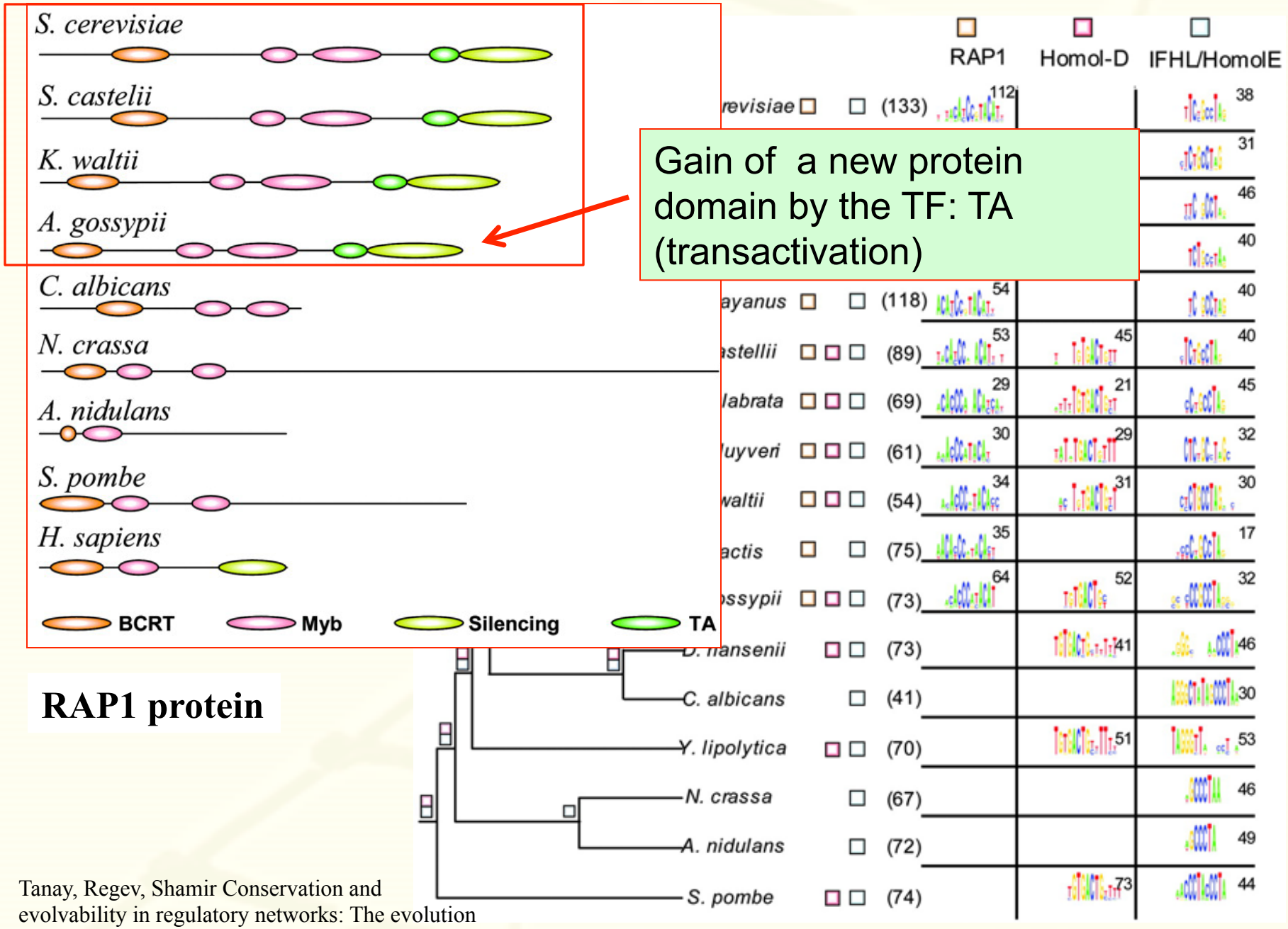


Tanay, Regev, Shamir Conservation and evolvability in regulatory networks: The evolution of ribosomal regulation in yeast PNAS 2005



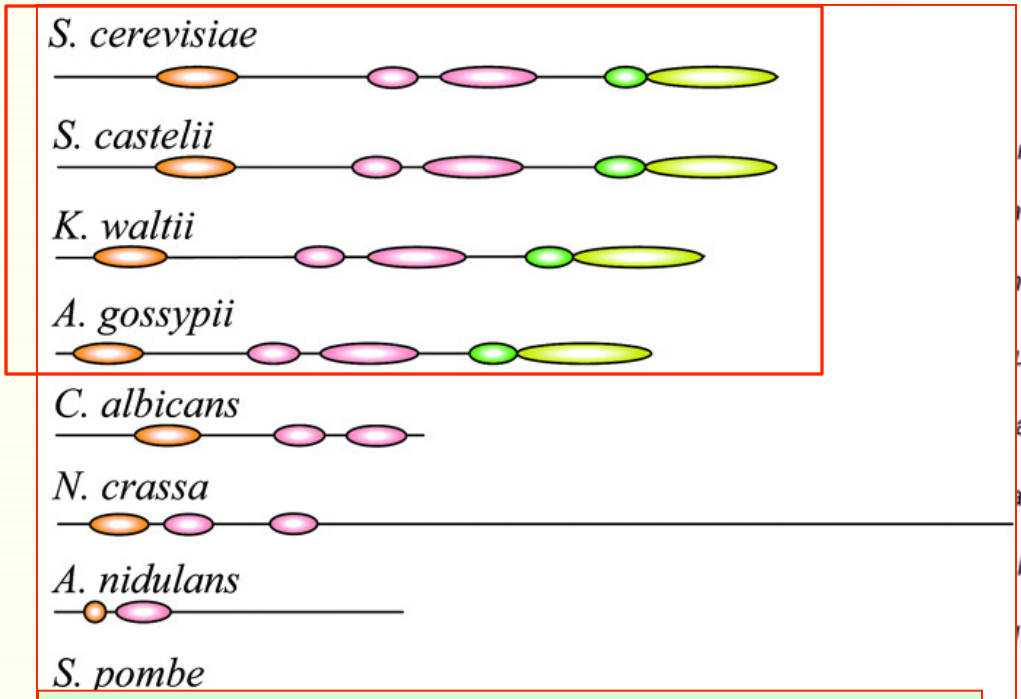
RAP1 protein

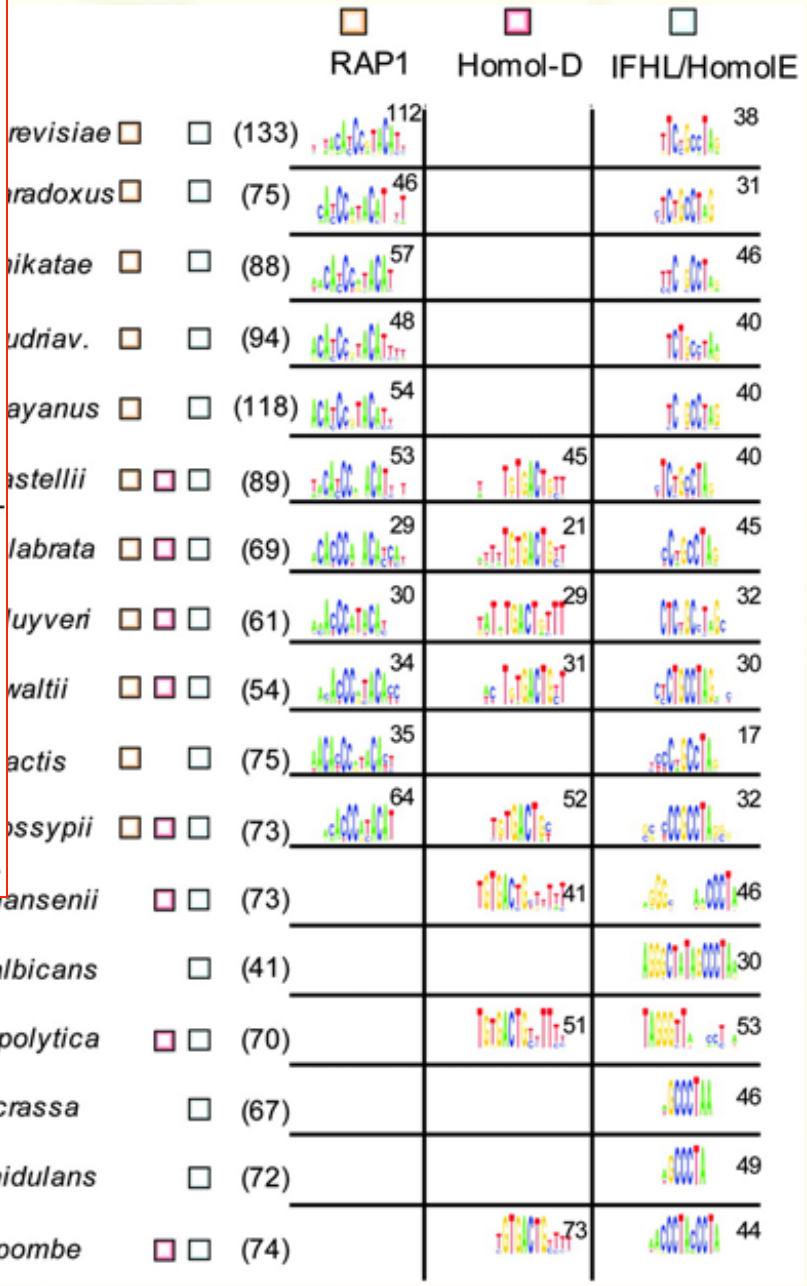
Tanay, Regev, Shamir Conservation and evolvability in regulatory networks: The evolution of ribosomal regulation in yeast PNAS 2005



RAP1 protein

Tanay, Regev, Shamir Conservation and evolvability in regulatory networks: The evolution of ribosomal regulation in yeast PNAS 2005





Here is likely what happened:

Step 1: Rap1 gained a TA domain and becomes a transcription factor.

Step 2: Ribosomal protein genes gained a binding site for Rap1

Step 3: Cbf1 binding sites are lost from ribosomal protein genes

A word of caution

- What we see for ribosomal proteins may be an exception. It is likely that the majority of the orthologs are under the regulation of the same transcription factors in related organisms.

- But the spatial distribution of these binding sites could be less conserved than previously thought.

- Also the sequence of these binding sites could have diverged beyond recognition. It would be interesting to investigate the co-variation or compensation between the protein sequence of TF and the actual binding sites.

We will come back with more interesting stories on regulatory evolution and variation

- Evolutionary Dynamics of Transcription Factor Binding in 5 vertebrates
- “humanized mouse”: moving a human chromosome into mouse cells
- “Population genomics of human gene expression”, cis vs trans effects, and SNP vs CNV
- “Variation in Transcription Factor Binding Among Human individuals”

End of lecture 3

- Next: We will talk about protein-protein interactions, protein complexes, biological networks etc in the next lecture.