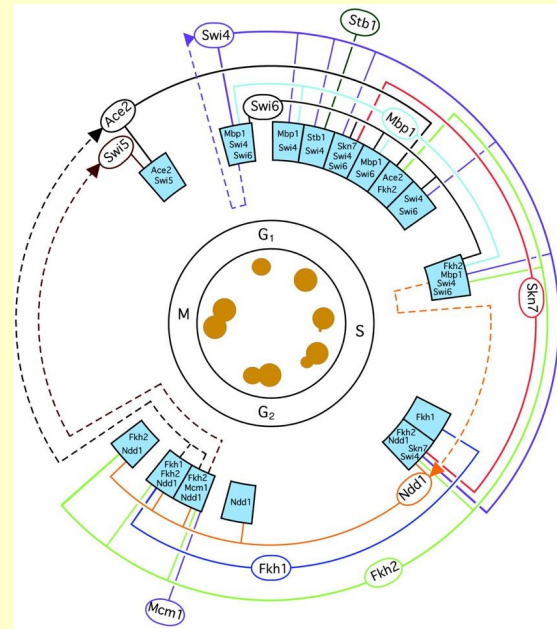


Gene Regulatory Networks



Doug Brutlag

Professor Emeritus

Biochemistry & Medicine (by courtesy)

Homework 7

Transcription Factor Motifs for Co-expressed Genes

- Choose a group of co-expressed genes from the Yeast Cell Cycle Database by first entering a single gene of interest to you and then clicking on select all co-expressed genes:

<http://genome-www.stanford.edu/cgi-bin/cellcycle/search>

- Collect all the names of the genes and enter them in Go Term Finder to discover common biological processes, common molecular mechanisms and common cellular components in the Saccharomyces Genome Database:

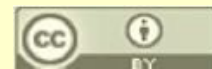
<http://www.yeastgenome.org/cgi-bin/GO/goTermFinder.pl>

- Using the Batch download link, collect all the gene sequences plus 1 KB upstream and enter the 1 KB upstream sequence from all the genes into MDScan and search for motifs 8, 10 or 12 bases long:

<http://mdscan.stanford.edu/>

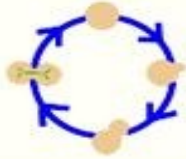
- You may have to remove simple sequences such as (tatatatata or ttttttttt or cacacacaca) by globally replacing them with an X character in order to ignore simple sequence repeats from the list of motifs found.
- Enter your list of co-expressed gene names in YeasTract web site and see if MDScan has discovered any known transcription binding sites:

<http://www.yeasttract.com/>



Yeast Cell Cycle Database

<http://genome-www.stanford.edu/cgi-bin/cellcycle/search>



Yeast Cell Cycle Analysis Project

Home

Cell Cycle
Analysis Home

[Home](#) | [Search](#) | [Figures](#) | [Data](#) | [Info](#) | [Links](#) |
[SGD](#) | [Microarray Homepage](#)

View Figures

View figures
from the paper

Search

Search the
complete dataset

Download

Data

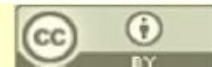
Download
images and
primary data
tables

Information

Information on
how to use this
site and scientific
methods

**Input the name (either ORF or common e.g
ACT1, YAL004W) for the gene you wish to
view**

**You may also use a wild-card to retrieve a
list of genes that match your expression, eg
cdc*.**

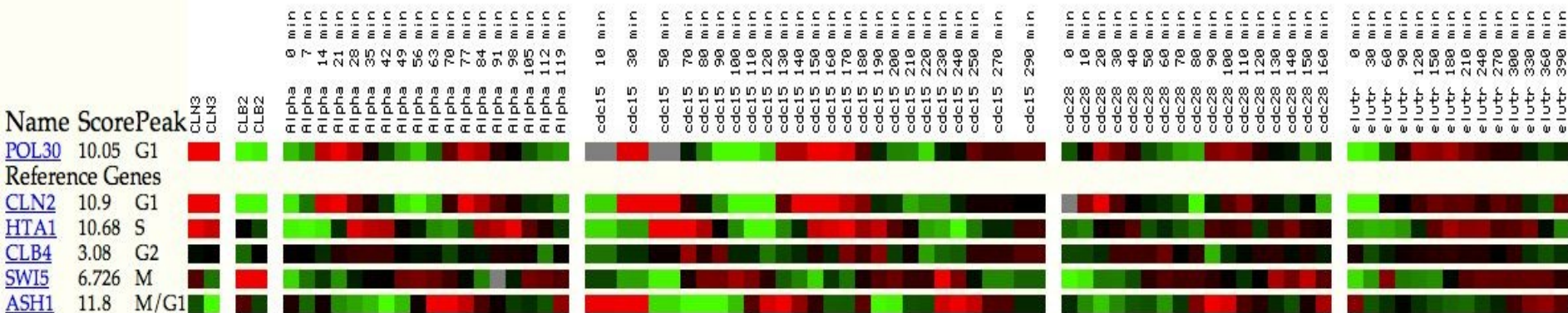


Yeast Cell Cycle Database

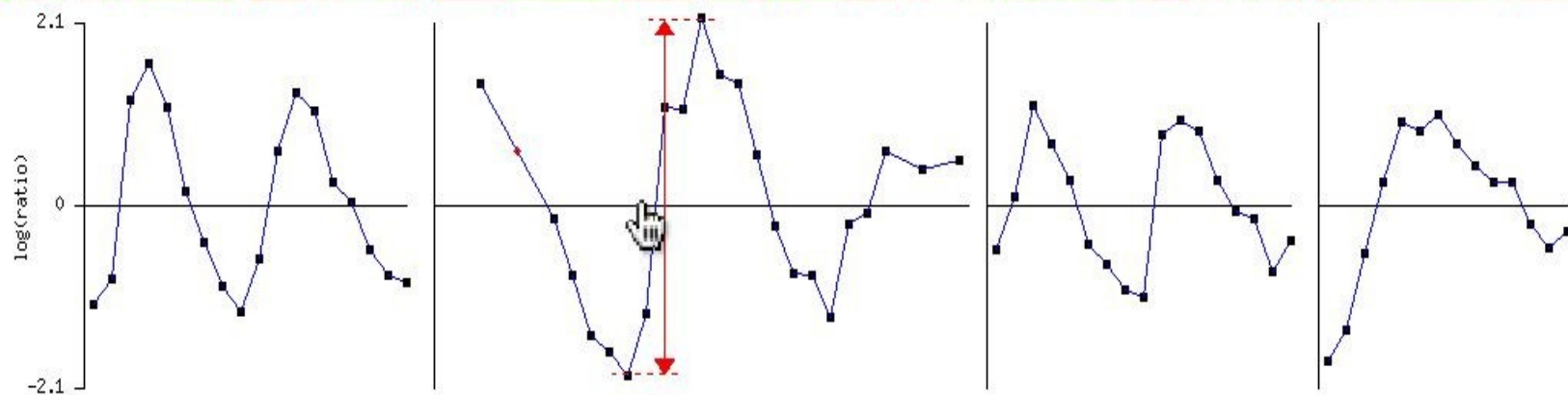


Yeast Cell Cycle Analysis Project

[Home](#) | [Search](#) | [Figures](#) | [Data](#) | [Info](#) | [Links](#) | [SGD](#) | [Microarray_Homepage](#)



Plot of POL30 (YBR088C)



Peak to trough ratio for POL30 in the cdc15 experiment is 17.38

Saccharomyces Genome Database

<http://www.yeastgenome.org/>



Saccharomyces Genome Database

[Site Map](#) | [Search Options](#) | [Help](#) | [Home](#) | |

[Community Info](#) | [Submit Data](#) | [BLAST](#) | [Primers](#) | [PatMatch](#) | [Gene/Seq Resources](#) | [Advanced Search](#) | [Community Wiki](#)

► Search Options

[Advanced Search](#), [Full-text Search \(Textpresso\)](#), [Search SGD web pages](#), [Global Gene Hunter](#), [Search Literature](#), and more.

► Help Resources

[Getting Started](#), [Sitemap](#), [FAQ](#), and more.

► Analysis & Tools

[BLAST](#), [Gene/Seq Resources](#), [Maps](#), and more.

► Homology & Comparisons

[PDB Homologs](#), [Protein Domains/Motifs](#), [Homologs](#), and more.

► Function & Expression

[Protein Info](#), [Pathways](#), [Expression Connection](#), and more.

► GO Resources

[GO Tutorial](#), [What is GO?](#), [GO Slim Mapper](#), [GO Term Finder](#), and more.

SGD™ is a scientific database of the molecular biology and genetics of the yeast *Saccharomyces cerevisiae*, which is commonly known as baker's or budding yeast.

New and Noteworthy

• **SGD Quarterly Newsletter - January 29, 2010**

SGD sends out its quarterly newsletter to colleagues designated as contacts in SGD. An [HTML version](#) of the newsletter is available. If you would like to receive this letter in the future please use the [Colleague Submission/Update](#) form to let us know.

• **SGD Curation News**

- [Genome Snapshot](#)
- [New papers](#) added to SGD this week.
- View [Genome-wide Analysis papers](#) in SGD.
- Do a [literature search](#).
- [Table of Gene Summary Paragraphs](#) (new ones highlighted yellow).

• **New genetic interaction data file available for download from SGD - January 22, 2010**

A large set of genetic interaction data from [Costanzo et al., The Genetic Landscape of a Cell, Science 327: 425-431 \(2010\)](#) is available for [download](#) from SGD's [Download Data Files](#) page and will be available via the web interface in the future. This study presents 74,984 genetic interactions, comprising a functional map of the cell. Thanks to the authors for providing their data to SGD.



Saccharomyces Genome Database Go Term Finder

<http://www.yeastgenome.org/>



[Site Map](#) | [Search Options](#) | [Help](#) | [Home](#) | [RSS](#)

[Community Info](#) | [Submit Data](#) | [BLAST](#) | [Primers](#) | [PatMatch](#) | [Gene/Seq Resources](#) | [Advanced Search](#) | [Community Wiki](#)

Gene Ontology Term Finder

The GO Term Finder (Version 0.83) searches for significant shared GO terms, or [parents](#) of those GO terms, used to describe the genes in your list to help you discover what they have in common. To map annotations of a group of genes to more general terms and/or to bin them in broad categories, use the [GO Slim Mapper](#).

Default Settings:

1. All genes/features that have GO annotations in the database
2. Manually curated and High-throughput annotation methods
3. Hits with $p\text{-value} < 0.01$ will be displayed on the results page

Optional: Add a title for your search results

Step 1: Query Set (Your Input)

Enter Gene/ORF names:
(separated by a return or a space)

OR

Upload a file of Gene/ORF names:

no file selected

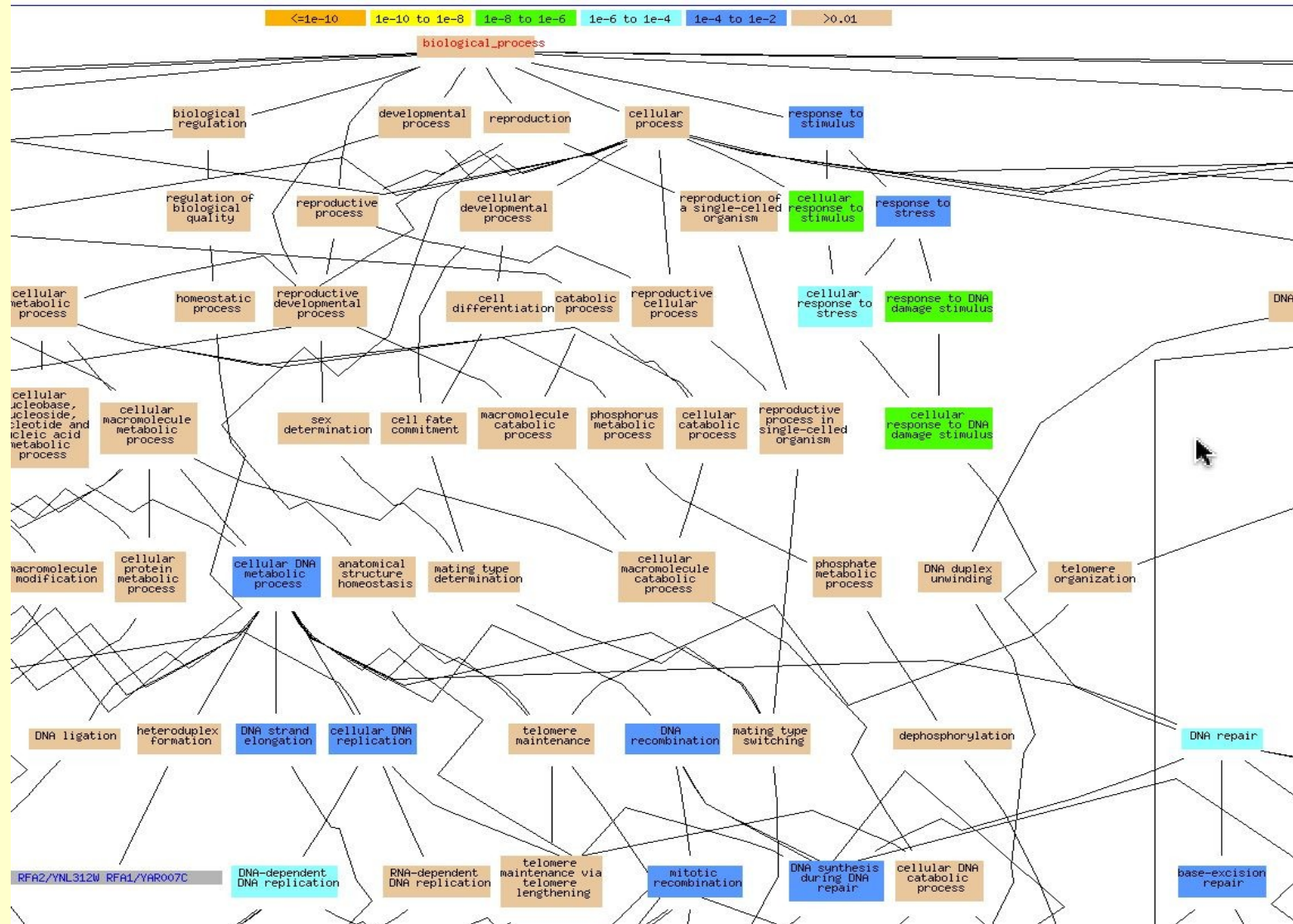
Step 2: Choose Ontology and Set Cutoff

- Process
 Function
 Component

Search using [default settings](#) or use Step 3, Step 4, and/or Step 5 below to customize your options.




Saccharomyces Genome Database Go Term Finder Results




Saccharomyces Genome Database Gene Sequence Resources

<http://www.yeastgenome.org/cgi-bin/seqTools>

Gene/Sequence Resources



[Site Map](#) | [Search Options](#) | [Help](#) | [Home](#) | 

[Community Info](#) | [Submit Data](#) | [BLAST](#) | [Primers](#) | [PatMatch](#) | [Gene/Seq Resources](#) | [Advanced Search](#) | [Community Wiki](#)

Gene/Sequence Resources

This resource allows retrieval of a list of options for accessing biological information, table/map displays, and sequence analysis tools for

1. a named gene or sequence,
2. a specified chromosomal region, or
3. a raw DNA or protein sequence.

If you would like to retrieve information simultaneously for multiple genes, please go to the [Batch Download page](#).

1. Enter a Name
or the first few characters followed by *:

Examples:
Gene - act1
ORF - YHR023W
SGDID - S000001855
GenBank Locus - YSCHELI
GenBank AccNo. - L00683
GenBank GI - gi:171655
Clone - 70353

If available, add flanking basepairs
upstream and downstream

Use the reverse complement

OR

2. Pick a chromosome:

Then enter coordinates (optional):
 to

The entire chromosome sequence
will be displayed if no coordinates are
entered.

Note: Enter coordinates in ascending
order for the Watson strand and
descending order for the Crick
strand.

Use the reverse complement

OR

**3. Type or Paste a
Sequence:**

DNA

The sequence **MUST** be provided in **RAW
format**, no comments (numbers are okay).

Use the reverse complement

 [Return to SGD](#)

[Send a Message to the SGD Curators](#)



Doug Brutlag 2010

Saccharomyces Genome Database Batch Download Tool

<http://www.yeastgenome.org/cgi-bin/batchDownload>

SGD Batch Download Tool

Step 1: Your Input

Option 1. Enter Feature/Standard Gene names (separate by return:)

POL30
MCD1
MSH6
CLN2

Examples:

Gene Name - PPN1
ORF - YDR452W
SGDID - S000002860

OR

Upload a file of Feature/Standard Gene names

no file selected

Option 2. Pick a chromosome

Then enter coordinates (optional):

to

If no coordinates are entered all the features in the selected chromosome will be retrieved.

OR

Upload a file of chromosomal regions (coordinates are optional) with the following tab-separated columns:

chr start_coordinate stop_coordinate

For example:

III 1356 20455

IV 11331 18001

VI 9856 100010

no file selected

Step 2: Choose the type of data that you want to retrieve (You can select multiple types) Please check the [help page](#) for details on the output file format.

• Sequence data

- Genomic DNA (DNA sequence with introns)
- Genomic DNA + 1 kb upstream and 1 kb downstream of flanking sequence
- Coding Sequence (DNA sequence without introns)
- ORF Translation (Protein Sequence)

• Other data


- Chromosomal Feature information including gene names, coordinates, description and SGDID (in [SGD_Feature.tab](#) file format)
- Gene Ontology (GO) Annotations (in [gene_association](#) file format)
- Phenotype
- Physical Interactions (in [Cytoscape](#) file format)
- Genetic Interactions (in [Cytoscape](#) file format)




Saccharomyces Genome Database Batch Download Tool

<http://www.yeastgenome.org/cgi-bin/batchDownload>

Download Data



[Site Map](#) | [Search Options](#) | [Help](#) | [Home](#) |  [RSS](#)

[Community Info](#) | [Submit Data](#) | [BLAST](#) | [Primers](#) | [PatMatch](#) | [Gene/Seq Resources](#) | [Advanced Search](#) | [Community Wiki](#)

Download Data


Downloading data for 21 Feature/Gene names...

This may take a while to run. For large gene lists, please download data from our [FTP site](#). Please wait...

Wait.....
Wait...

The following file(s) contain the data you requested (please refer to the [help page](#) for details on the file format). Please click on the file name to download.
Data will be available at this URL for **1 hour** after the request was made.

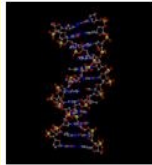
Type of Data Requested	Number of Features in the File	Name of Downloadable File	Size
Genomic DNA + 1kb upstream and 1kb downstream of flanking sequence	21	genomic1kb.fasta_643.gz	24.59 KB

[Return to SGD](#) [Send a Message to the SGD Curators](#) 



MDScan

<http://mdscan.stanford.edu/>



Overview

MDscan finds motif from ChIP-on-chip targets

Motif Finding

Search for interesting motifs in your sequences on our server

Input Format

How to specify the input parameters

Output

Explanations

How to look at the output file we send you

Motif Regressor

An improved MDscan for download.

Reference

Nat Biotechnol. 2002 20(8):835-9.

Contacts

People behind the project

SeqMotifs

See other motif finding algorithms we have developed.

MDscan

A Fast and Accurate Motif Finding Algorithm With Applications To Chromatin Immunoprecipitation Microarray Experiments

[Xiaole Shirley Liu](#), [Douglas L. Brutlag](#), [Jun S. Liu](#)
[Stanford Medical Informatics](#), [Stanford University](#)

While chromatin immunoprecipitation followed by cDNA microarray (ChIP-on-chip) has become a popular procedure for studying genome-wide protein-DNA interactions and transcription regulation, it can only map the probable protein-DNA interaction loci within 1-2kb resolution. To pinpoint the interaction sites down to the base pair level, we introduce a novel computational method, Motif Discovery scan (MDscan), that examines the ChIP-array selected sequences and searches for DNA sequence motifs representing the protein-DNA interaction sites. MDscan combines the advantages of two widely adopted motif search strategies, word enumeration and position-specific weight matrix updating, and incorporates the ChIP enrichment information to accelerate the search and enhance its success rate. The intuition is to first search for similar words appearing in the sequences more likely to contain the motif (highly ChIP-enriched sequences) because these sequences have higher signal to noise ratio. Words in each similarity group can initialize a position specific motif matrix and the motif can be updated and refined with the whole input sequences (all ChIP-selected targets). The method showed both speed and accuracy advantages compared to several established motif-finding algorithms in both simulation and published yeast ChIP-on-chip experiments. MDscan can be used not only with the ChIP experiments, but also to find DNA motifs in other experiments in which a subgroup of the sequences can be inferred to contain relatively more abundant motif sites.

We recently developed a new program, [Motif Regressor](#), to better utilize mRNA expression level or ChIP-on-chip enrichment information to improve the performance of MDscan. Motif Regressor first identifies a set of non-redundant candidate motifs using MDscan, and scans the promoter region of every gene in the genome with each candidate motif to measure how good a promoter matches a motif (in terms of both the number of sites and the strength of matching). It then uses linear regression analysis to select motifs whose promoter matching scores are significantly correlated with ChIP-on-chip enrichment or downstream gene expression values. When ranking motifs by linear regression p-value, Motif Regressor automatically picks the best motif and optimal motif width. Due to its computational intensity, Motif Regressor is not currently available as a web server. However, for interested users to explore the program locally, the program is available for download at: <http://www.math.umass.edu/~conlon.mr.html>

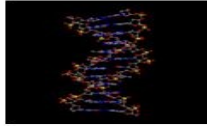
Obtaining a local copy of MDscan:

MDscan is free-of-charge to academia. Please check out: [Brutlag Bioinformatics Group Software Download](#) and [Academic License Instructions](#) for details.



MDScan Input

<http://mdscan.stanford.edu/>



Overview

MDscan finds motif from ChIP-on-chip targets

Motif Finding

Search for interesting motifs in your sequences on our server

Input Format

How to specify the input parameters

Output

Explanations

How to look at the output file we send you

Motif Regressor

An improved MDscan for download.

Reference

Nat Biotechnol. 2002 20(8):835-9.

Contacts

People behind the project

SeqMotifs

See other motif finding algorithms we have developed.

*Suggestions, comments, bugs to: [Xiaole Shirley Liu](#)
Last updated 4/15/2004.*

MDscan Search

Please don't submit more than one job at a time!!

Submit another job only if one of the following is true:

1. You have received answer from your previous submission.
2. Your previous job was submitted more than one day ago.

User information

Please specify your email so we can send the search result to you

Run information

Please specify a run name (will be included in the result email)

Input Sequences

You can either specify a file

no file selected

Or paste your sequences below:

```
>YKR013W PRY2 SGDID:S000001721, Chr XI from 462602-465591, Verified ORF
TCAAAAATTCACGGCGGCAAAAAGGTCTGTAACACTACCTAAACATTTAAAGCCTTTGAGCTA
TAAATCATTATGTACACTGAGAATGCCACATGCTAAAACAAGGAAGTTTTACGCCGAAGTA
GCACATTAAC TGACCACTCAATAGAAAGAAAACAGAAAGTGGAAATGGATGGCCACTTTC
A
```

Background Model

You can use the input sequence as background if you don't specify any of the following.

Or you can specify a background sequence file

no file selected

Or paste background sequences below:

Or use the precomputed genome background model

Motif Model

Motif width

Number of top sequences to look for candidate motifs

Number of candidate motifs for scanning the rest sequences

Report the top final motifs found



MDScan Output

```
*****
*
*           MDscan Search Result
*
*****
```

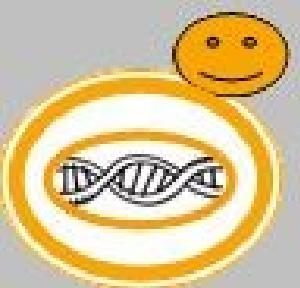
Pm 0.2639 Minimum match (6/8)
The highest scoring 5 motifs are:

Motif 1: Wid 8; Score 4.276; Sites 68; Con AAACGCGT; RCon ACGCGTTT

	A	C	G	T	Con	rCon	Deg	rDeg
1	47.78	7.85	6.44	37.93	A	T	W	W
2	47.78	7.85	27.57	16.80	A	T	R	Y
3	97.08	0.81	0.81	1.31	A	T	A	T
4	1.31	96.58	0.81	1.31	C	G	C	G
5	1.31	0.81	96.58	1.31	G	C	G	C
6	1.31	96.58	0.81	1.31	C	G	C	G
7	1.31	0.81	96.58	1.31	G	C	G	C
8	28.07	0.81	0.81	70.32	T	A	W	W

```
| YPL256C CLN2 SGDID:S000006177, Chr XVI from 67614-63977, reverse complement, Verified ORF Len 1020 Site #1 r 36
TGACGCGA
| YML027W YOX1 SGDID:S000004489, Chr XIII from 220406-223563, Verified ORF Len 1000 Site #1 f 494
AAACGCGT
| YML027W YOX1 SGDID:S000004489, Chr XIII from 220406-223563, Verified ORF Len 1000 Site #2 r 496
TTACGCGT
| YML027W YOX1 SGDID:S000004489, Chr XIII from 220406-223563, Verified ORF Len 1000 Site #3 f 555
AGACGCGA
| YML027W YOX1 SGDID:S000004489, Chr XIII from 220406-223563, Verified ORF Len 1000 Site #4 f 760
AAACGCGA
| YER070W RNR1 SGDID:S000000872, Chr V from 297948-302614, Verified ORF Len 1020 Site #1 r 695
TTACGCGT
| YER070W RNR1 SGDID:S000000872, Chr V from 297948-302614, Verified ORF Len 1020 Site #2 r 630
TGACGCGT
| YER070W RNR1 SGDID:S000000872, Chr V from 297948-302614, Verified ORF Len 1020 Site #3 r 559
TTACGCGT
| YER070W RNR1 SGDID:S000000872, Chr V from 297948-302614, Verified ORF Len 1020 Site #4 r 509
AAACGCGT
| YER070W RNR1 SGDID:S000000872, Chr V from 297948-302614, Verified ORF Len 1020 Site #5 f 507
TTACGCGT
| YER070W RNR1 SGDID:S000000872, Chr V from 297948-302614, Verified ORF Len 1020 Site #6 f 557
AAACGCGT
| YER070W RNR1 SGDID:S000000872, Chr V from 297948-302614, Verified ORF Len 1020 Site #7 f 628
AAACGCGT
| YER070W RNR1 SGDID:S000000872, Chr V from 297948-302614, Verified ORF Len 1020 Site #8 f 693
```





YeasTract Transcription Factor Database

<http://www.yeasttract.com/>



[Tutorial](#)

Quick search... [Search](#)

DISCOVERER NEW

Regulatory Associations:

- [Search for TFs](#)
- [Search for Genes](#)
- [Search for Associations](#)

Group genes:

- [Group by TF](#)
- [Group by GO](#)

Pattern Matching:

- [Search by DNA Motif](#)
- [Find TF Binding Site\(s\)](#)
- [Search Motifs on Motifs](#)

Utilities:

- [ORF List ↔ Gene List](#)
- [IUPAC Code Generation](#)
- [Generate Regulation Matrix](#)

Retrieve:

- [TF-Consensus List](#)
- [Upstream Sequence](#)
- [Flat files](#)

About YeasTract:

- [Contact Us](#)
- [Cite YEASTRACT](#)
- [Acknowledgments](#)
- [Credits](#)

Support & suggestions:
yeastract@kdbio.inesc-id.pt

[Home](#)



Welcome to YEASTRACT

[Contact Us](#) - [Tutorial](#) - 



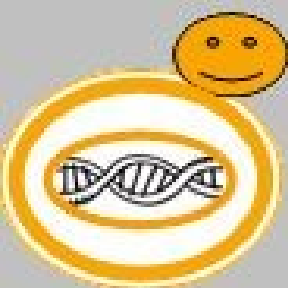
YEASTRACT (**Y**east **S**earch for **T**ranscriptional **R**egulators **A**nd **C**onsensus **T**racking) is a curated repository of more than 34469 regulatory associations between transcription factors (TF) and target genes in *Saccharomyces cerevisiae*, based on more than 1000 bibliographic references. It also includes the description of 291 specific DNA binding sites for more than a hundred characterized TFs. Further information about each Yeast gene has been extracted from the Saccharomyces Genome Database (SGD) as per the latest release, version 1.1438, from Apr 11, 2009. For each gene the associated Gene Ontology (GO) terms and their hierarchy in GO was obtained from the GO consortium as per the OBO flat file from Apr 24, 2009. Currently, YEASTRACT maintains a total of 27250 terms from GO. The nucleotide sequences of the promoter and coding regions for Yeast genes were obtained from Regulatory Sequence Analysis Tools (RSAT). All the information in YEASTRACT will be updated regularly to match the latest data from SGD, GO consortium, RSA Tools and recent literature on yeast regulatory networks.

YEASTRACT now includes [DISCOVERER](#), a set of tools that can be used to identify complex motifs found to be over-represented in the promoter regions of co-regulated genes. DISCOVERER is based on the MUSA and RISO algorithms. These algorithms take as input a list of genes and identify over-represented motifs, which can then be compared with transcription factor binding sites described in the YEASTRACT database.

Facilities are also provided to enable the exploitation of the gathered data when solving a number of biological questions, as exemplified in the Tutorial. YEASTRACT allows the identification of documented or potential transcription regulators of a given gene and of documented or potential regulons for each transcription factor. It also renders possible the comparison between DNA motifs and the transcription factor binding sites described in the literature. The system also provides a useful mechanism for grouping a list of genes (for instance a set of genes with similar expression profiles as revealed by microarray analysis) based on their regulatory associations with known transcription factors.

YEASTRACT provides a set of queries to search and retrieve important biological information from the gathered data and to predict transcription regulation networks in yeast from data emerging from gene-by-gene analysis or global approaches.





YeastTract Transcription Factor Search

<http://www.yeasttract.com/formfindregulators.php>



 [Tutorial](#)

Quick search... [Search](#)

DISCOVERER NEW

Regulatory Associations:

- [Search for TFs](#)
- [Search for Genes](#)
- [Search for Associations](#)

Group genes:

- [Group by TF](#)
- [Group by GO](#)

Pattern Matching:

- [Search by DNA Motif](#)
- [Find TF Binding Site\(s\)](#)
- [Search Motifs on Motifs](#)

Utilities:

- [ORF List ⇌ Gene List](#)
- [IUPAC Code Generation](#)
- [Generate Regulation Matrix](#)

Retrieve:

- [TF-Consensus List](#)
- [Upstream Sequence](#)
- [Flat files](#)

About Yeasttract:

- [Contact Us](#)
- [Cite YEASTRACT](#)
- [Acknowledgments](#)
- [Credits](#)

[Home](#) > [Search Transcription Factors](#)

[Contact Us](#) - [Tutorial](#) - 



Search Transcription Factors

By regulated genes

Search Type	Regulated Gene(s)
<input checked="" type="radio"/> Documented	POL30
<input type="radio"/> Potential	MCD1
<input type="radio"/> Image	MSH6
<input type="radio"/> Both	CLN2
	TOS4
	RAD27
	RFA1
	CSI2

[Search](#)

[Clear](#)



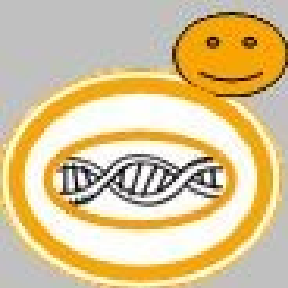
By keyword

Keyword

[Search](#)

[Clear](#)





YeastTract Transcription Factor Search Results

<http://www.yeasttract.com/formfindregulators.php>

 [Tutorial](#)

Quick search... [Search](#)

DISCOVERER NEW

Regulatory Associations:

- Search for TFs
- Search for Genes
- Search for Associations

Group genes:

- Group by TF
- Group by GO

Pattern Matching:

- Search by DNA Motif
- Find TF Binding Site(s)
- Search Motifs on Motifs

Utilities:

- [ORF List ⇌ Gene List](#)
- [IUPAC Code Generation](#)
- [Generate Regulation Matrix](#)

Retrieve:

- [TF-Consensus List](#)
- [Upstream Sequence](#)
- [Flat files](#)

About Yeasttract:

- [Contact Us](#)
- [Cite YEASTRACT](#)
- [Acknowledgments](#)
- [Credits](#)

[Home](#) > [Search Transcription Factors](#) > [Result](#)

[Contact Us](#) - [Tutorial](#) - 



Search Transcription Factors



ORF/Gene Name	Documented Transcription Factor - Reference
YBR088c/POL30 Back to top ▲	Mbp1p - Reference + Sfp1p - Reference Swi4p - Reference
YDL003w/MCD1 Back to top ▲	Cin5p - Reference + Mbp1p - Reference Sfp1p - Reference Swi4p - Reference
YDR097c/MSH6 Back to top ▲	Dot6p - Reference + Mbp1p - Reference Sfp1p - Reference Swi6p - Reference
YPL256c/CLN2 Back to top ▲	Gcn4p - Reference + Gcr1p - Reference Mbp1p - Reference Mcm1p - Reference Rme1p - Reference

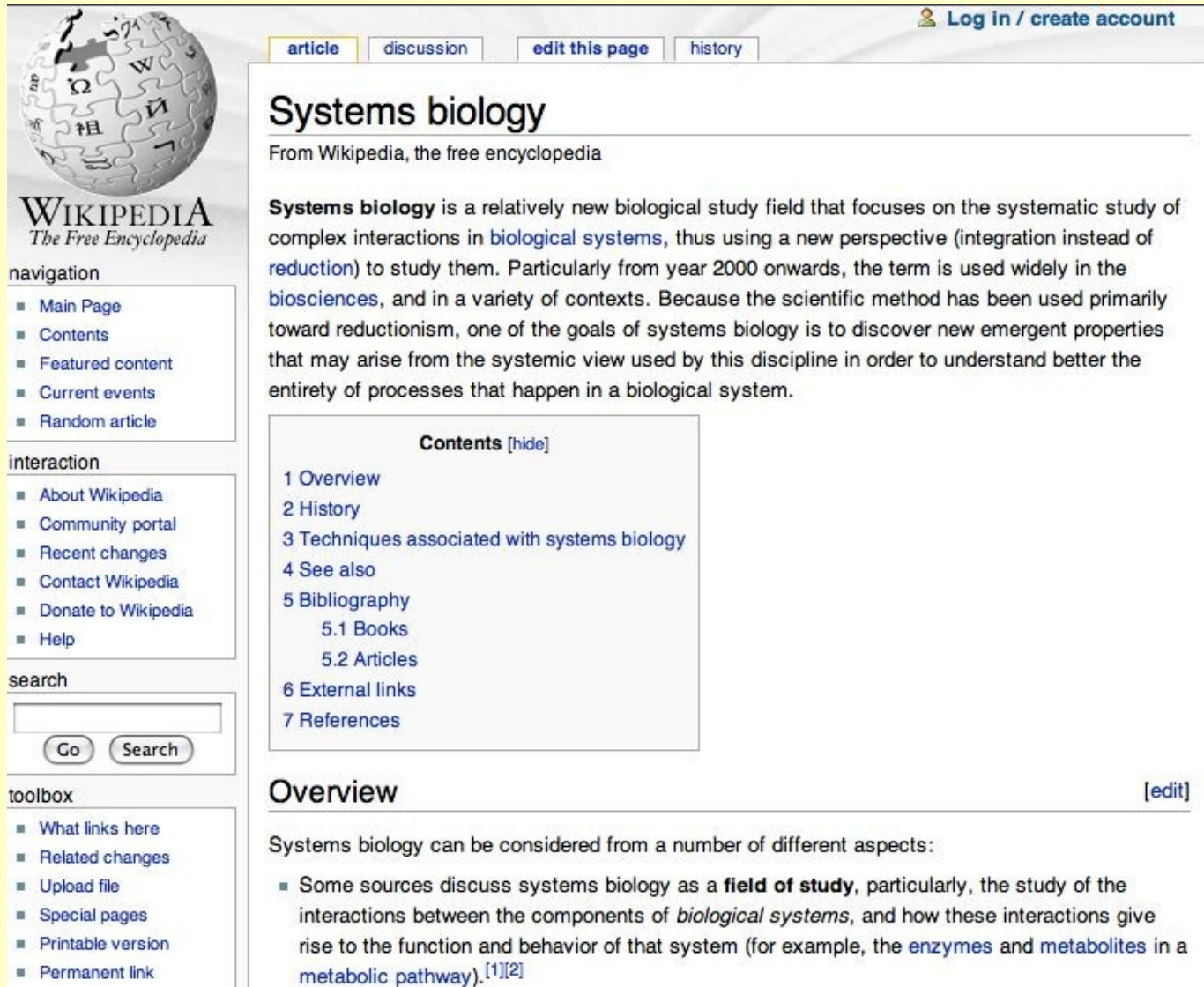

Jump to ▼

- [YBR088c - POL30](#)
- [YDL003w - MCD1](#)
- [YDR097c - MSH6](#)
- [YPL256c - CLN2](#)
- [YLR183c - TOS4](#)
- [YKLI13c - RAD27](#)
- [YAR007c - RFA1](#)
- [YOL007c - CSI2](#)
- [YOL090w - MSH2](#)
- [YGR152c - RSR1](#)
- [YML027w - YOX1](#)
- [YGR151c](#)
- [YLL022c - HIF1](#)
- [YILO66c - RNR3](#)
- [YOL017w - ESC8](#)
- [YKLO45w - PRI2](#)
- [YDL164c - CDC9](#)
- [YKRO13w - PRY2](#)
- [YNL312w - RFA2](#)
- [YIL140w - AXL2](#)
- [YERO70w - RNR1](#)



What is Systems Biology?

http://en.wikipedia.org/wiki/Systems_biology



The screenshot shows the Wikipedia article for "Systems biology". At the top right, there are links for "Log in / create account". Below that are tabs for "article", "discussion", "edit this page", and "history". The main heading is "Systems biology", followed by the text "From Wikipedia, the free encyclopedia". The main body of the article begins with a paragraph: "Systems biology is a relatively new biological study field that focuses on the systematic study of complex interactions in biological systems, thus using a new perspective (integration instead of reduction) to study them. Particularly from year 2000 onwards, the term is used widely in the biosciences, and in a variety of contexts. Because the scientific method has been used primarily toward reductionism, one of the goals of systems biology is to discover new emergent properties that may arise from the systemic view used by this discipline in order to understand better the entirety of processes that happen in a biological system." Below this is a "Contents" section with a "[hide]" link, listing: 1 Overview, 2 History, 3 Techniques associated with systems biology, 4 See also, 5 Bibliography (with sub-sections 5.1 Books and 5.2 Articles), 6 External links, and 7 References. The "Overview" section is partially visible, starting with "Systems biology can be considered from a number of different aspects:" followed by a list item: "Some sources discuss systems biology as a field of study, particularly, the study of the interactions between the components of biological systems, and how these interactions give rise to the function and behavior of that system (for example, the enzymes and metabolites in a metabolic pathway).^{[1][2]}". On the left side of the article, there is a navigation menu with links for Main Page, Contents, Featured content, Current events, Random article, interaction (About Wikipedia, Community portal, Recent changes, Contact Wikipedia, Donate to Wikipedia, Help), search (with a search box and "Go" and "Search" buttons), and toolbox (What links here, Related changes, Upload file, Special pages, Printable version, Permanent link).

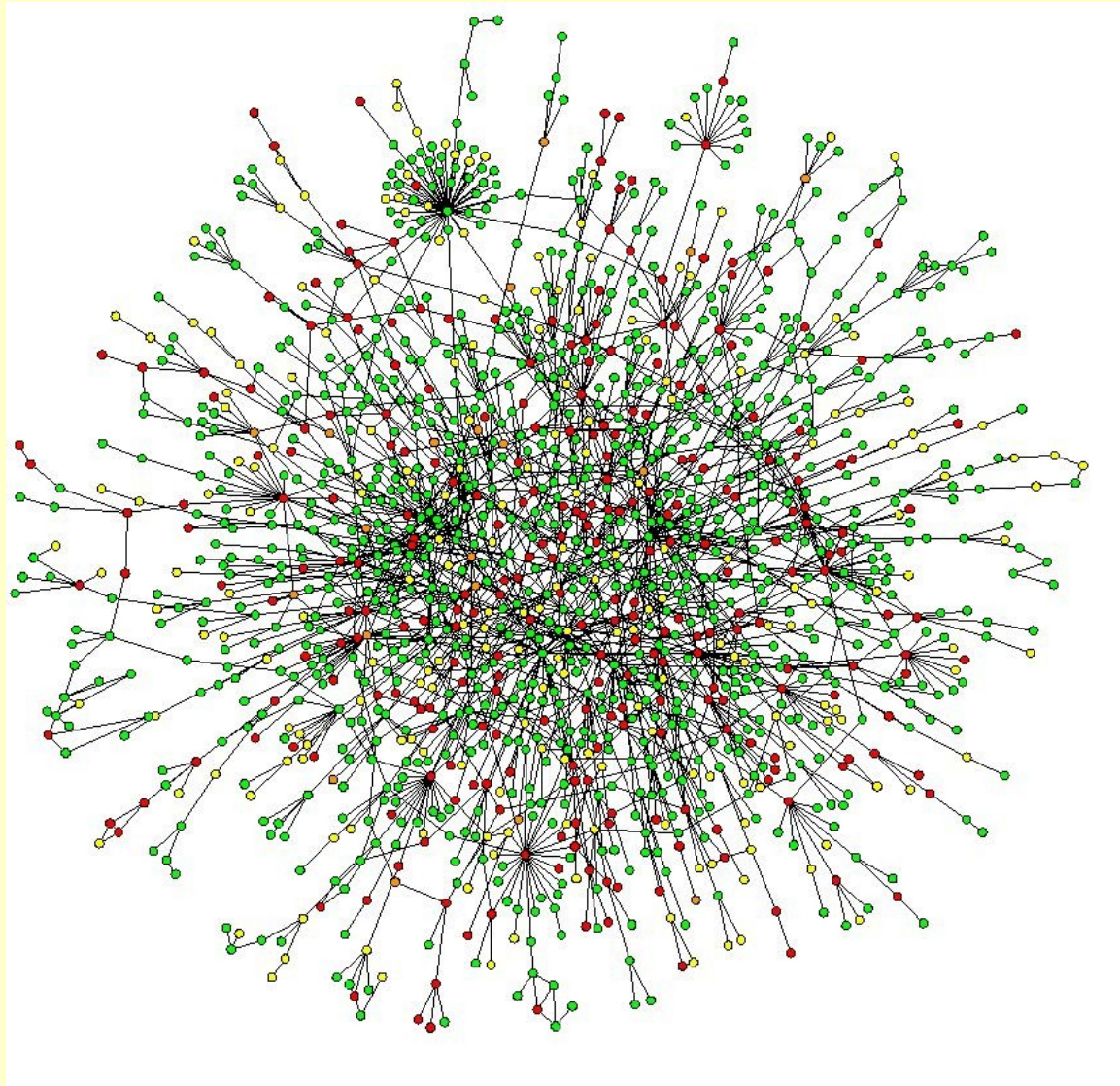


What is Systems Biology?

- Systematic approach to biology
 - Protein networks involving all protein-protein interactions
 - Metabolic networks involving all enzymes & pathways
 - Gene regulation involving all regulatory interactions
 - Cell behavior involving all environmental responses
 - Population biology involving all selectable traits.
- Challenges of System Biology
 - Identifying all selectable traits, all regulatory interactions, all protein complexes, all metabolic pathways, all regulatory pathways, all selectable traits
 - Computationally representing and simulating cell structures, metabolism, regulation, behavior and selection
 - Usually approached by simplification, studying one protein complex, one metabolic pathway, one regulatory network or one trait at a time. Then combining individual examples to represent the entire cell or organism.

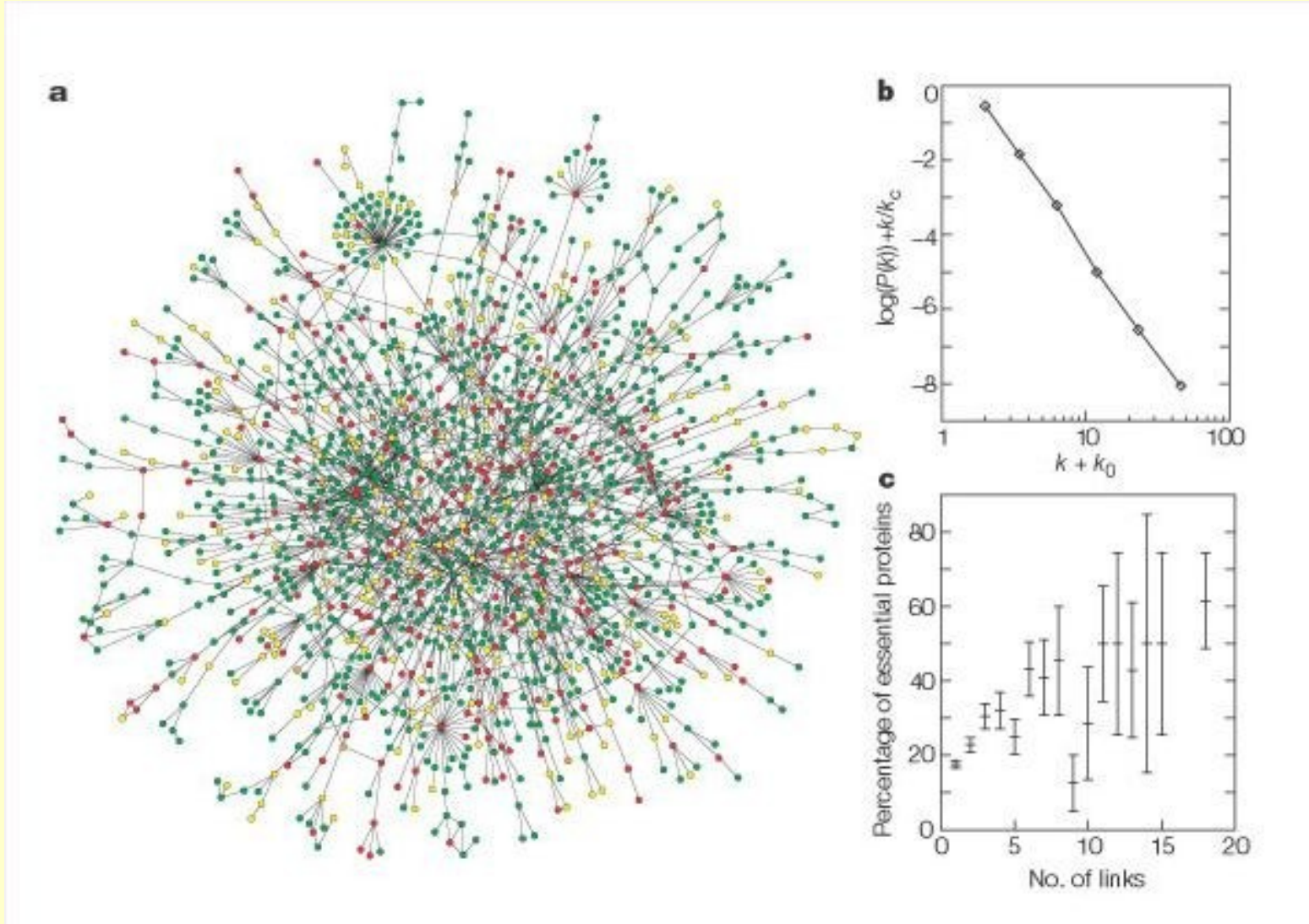
Yeast Protein Interaction Network

H. Jeong, et al., Nature, Vol 411, p41 (2001)



Yeast Protein Interaction Network

H. Jeong, et al., Nature, Vol 411, p41 (2001)



Evolutionary Rate in the Protein Interaction Network

Hunter B. Fraser,^{1*}† Aaron E. Hirsh,^{2*} Lars M. Steinmetz,³
Curt Scharfe,³ Marcus W. Feldman²

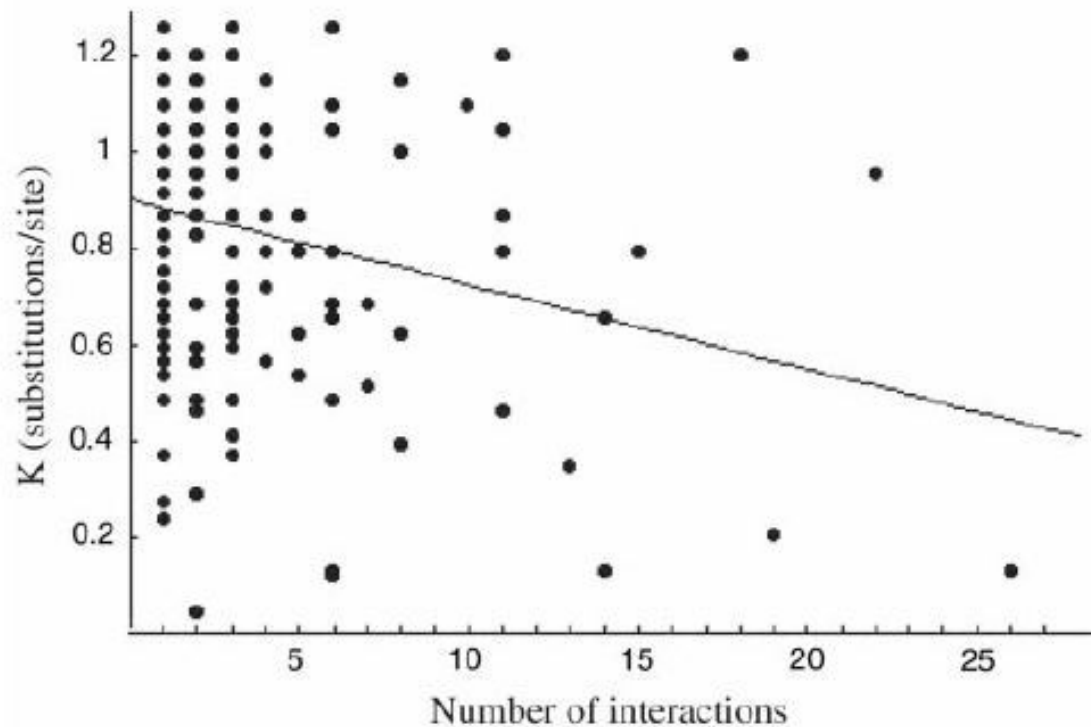
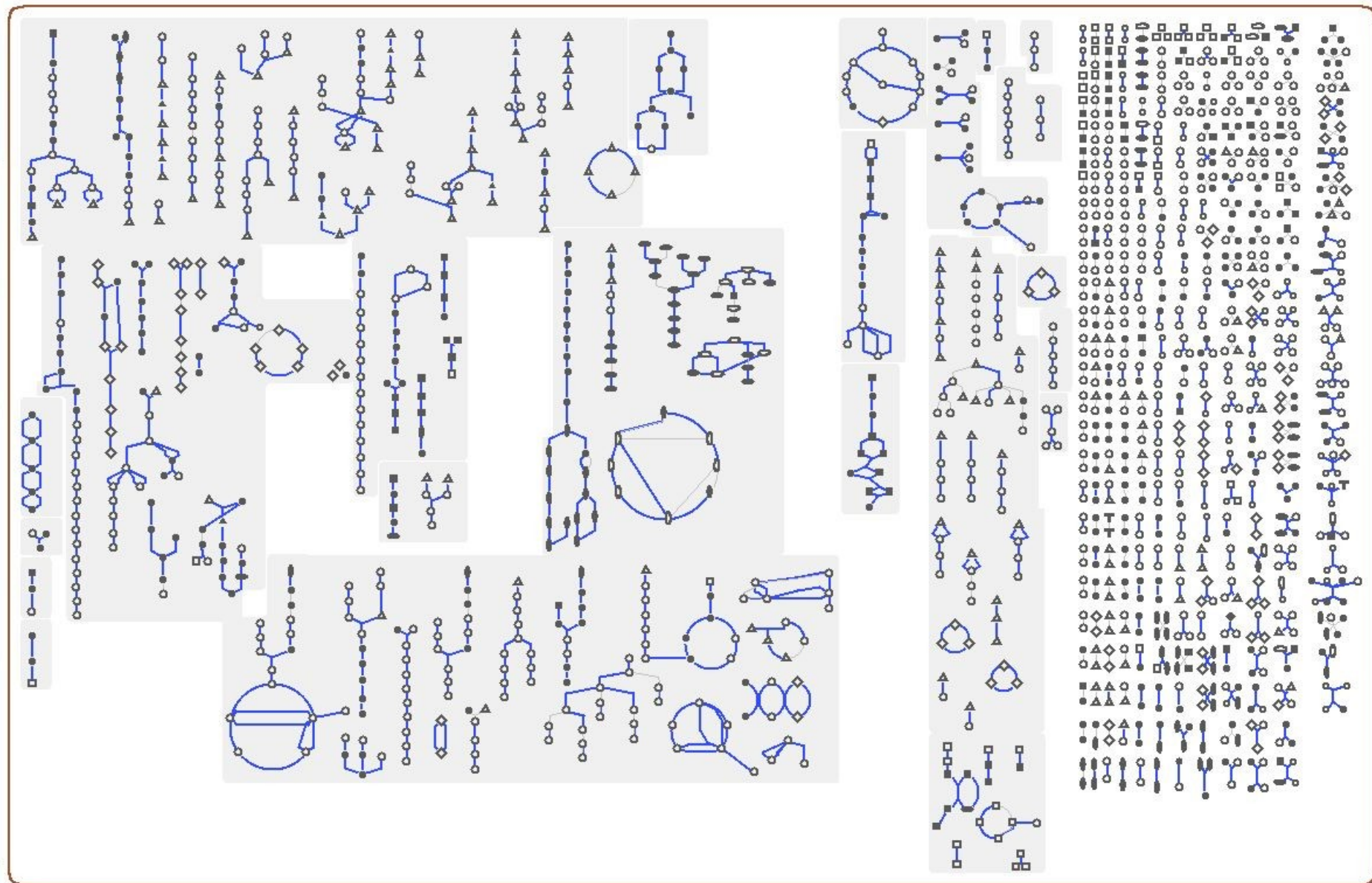


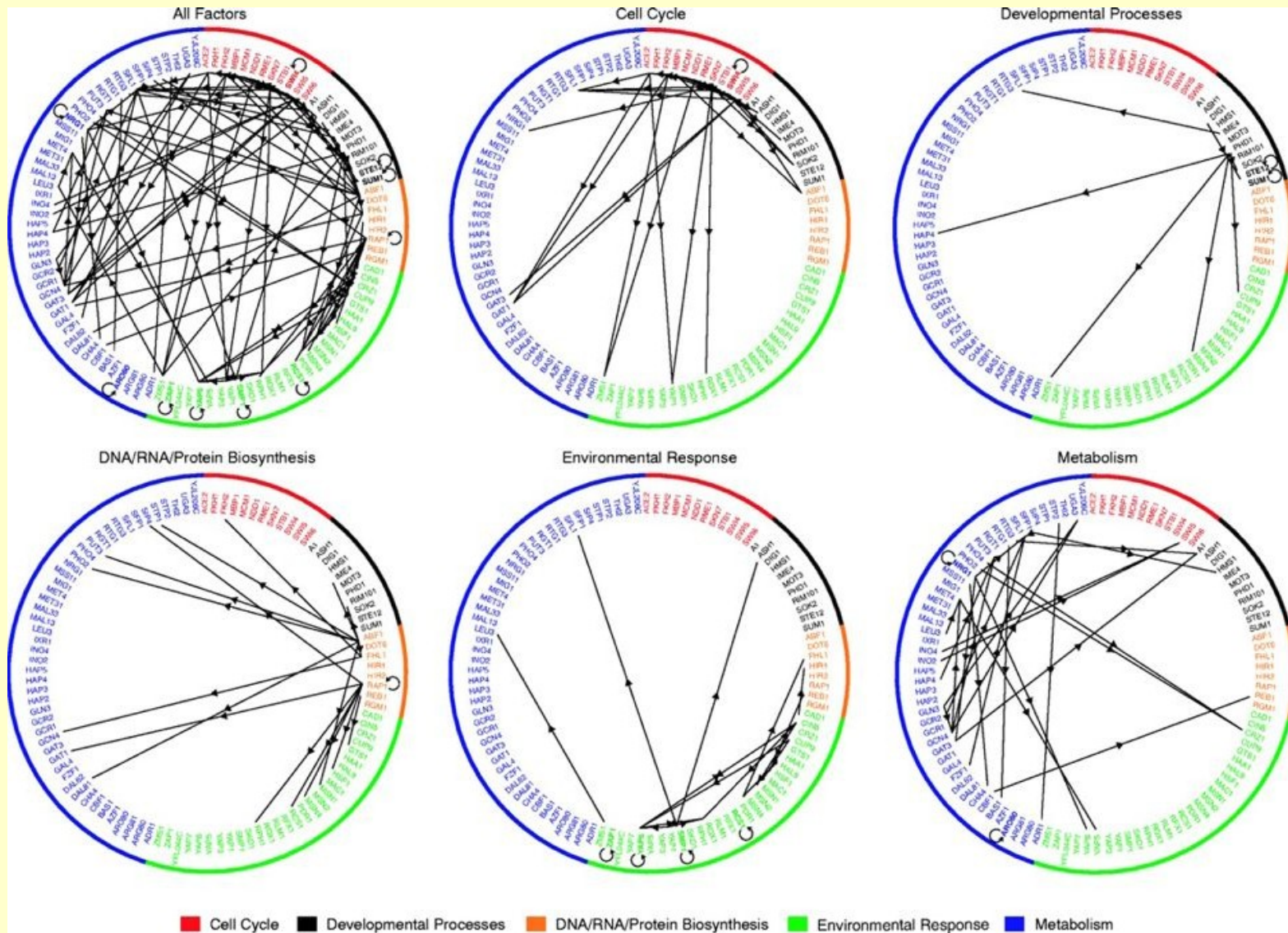
Fig. 1. The relation between the number of protein-protein interactions (I) in which a yeast protein participates and that protein's evolutionary rate, as estimated by the evolutionary distance (K) to the protein's well-conserved ortholog in the nematode *C. elegans*.

Yeast Metabolic Pathways

<http://biocyc.org/>



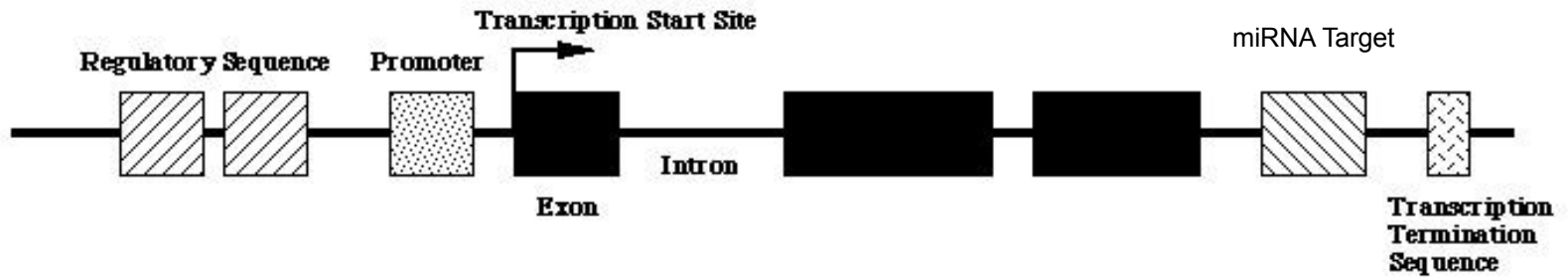
Yeast Gene Regulatory Networks



Genomics Provides

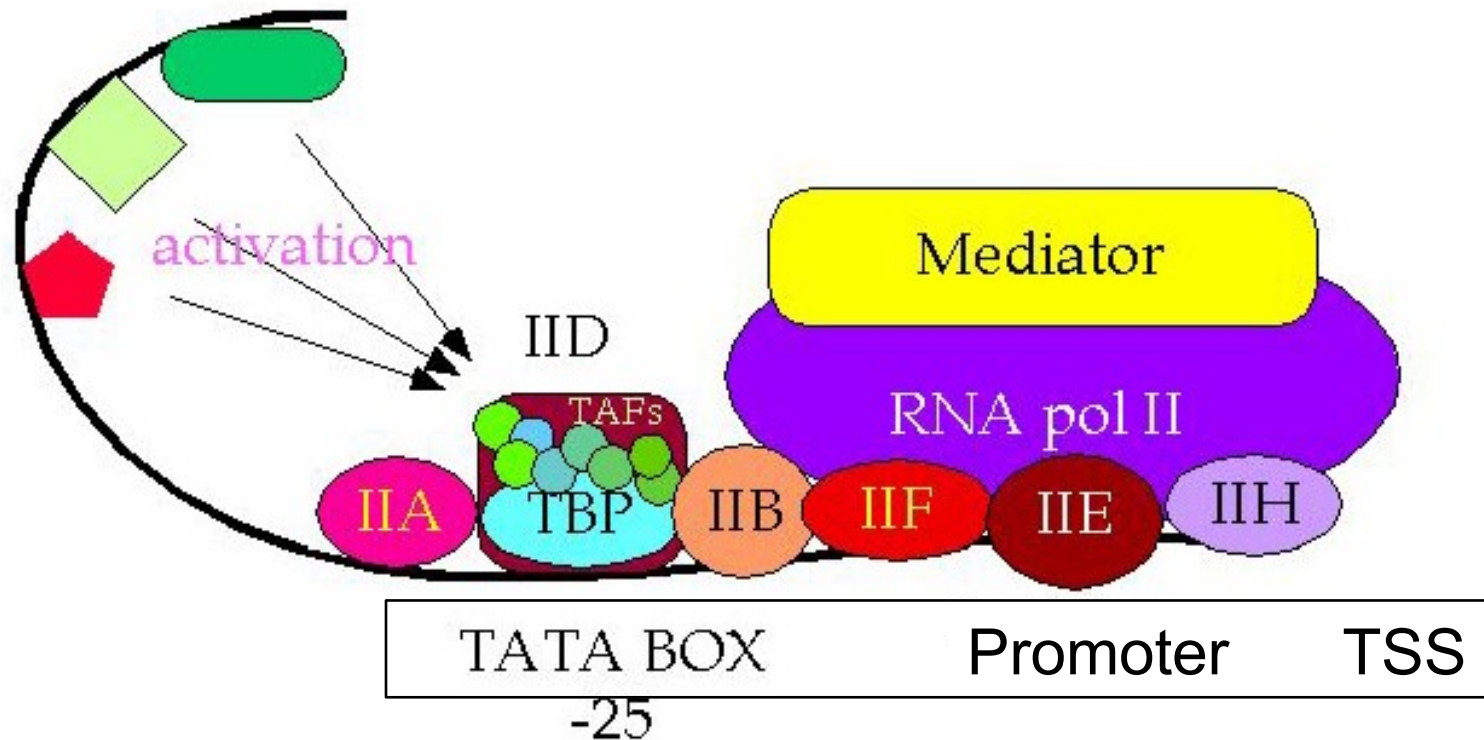
- Location and number of genes
 - By mapping messenger RNAs (cDNAs or ESTs)
 - By computational gene finding programs
 - Glimmer
 - GenScan
 - Grail & Grail EXP
 - FGENESH
 - GeneMark & GenMarkHMM
 - By homology and conservation
- Gene expression and regulation information
 - From tissue specific mRNAs (cDNAs or ESTs)
 - From DNA chips or microarrays
 - Gene probes and tiling arrays
 - ChIP-Chip experiments
 - [Stanford Microarray Database](#)
 - [NCBI GEO](#)
 - [EBI ArrayExpress](#)
- Gene regulatory sequences
 - Promoter signals near beginning (5' end) of transcripts
 - Transcription factor binding sites
 - Transcription start sites (TSS)
 - Transcription termination signals (TTS)
 - RNA processing signals
 - Splice donor and acceptor sites
 - 5' Cap site
 - 3' polyAdemylation

Eukaryotic Gene Structure

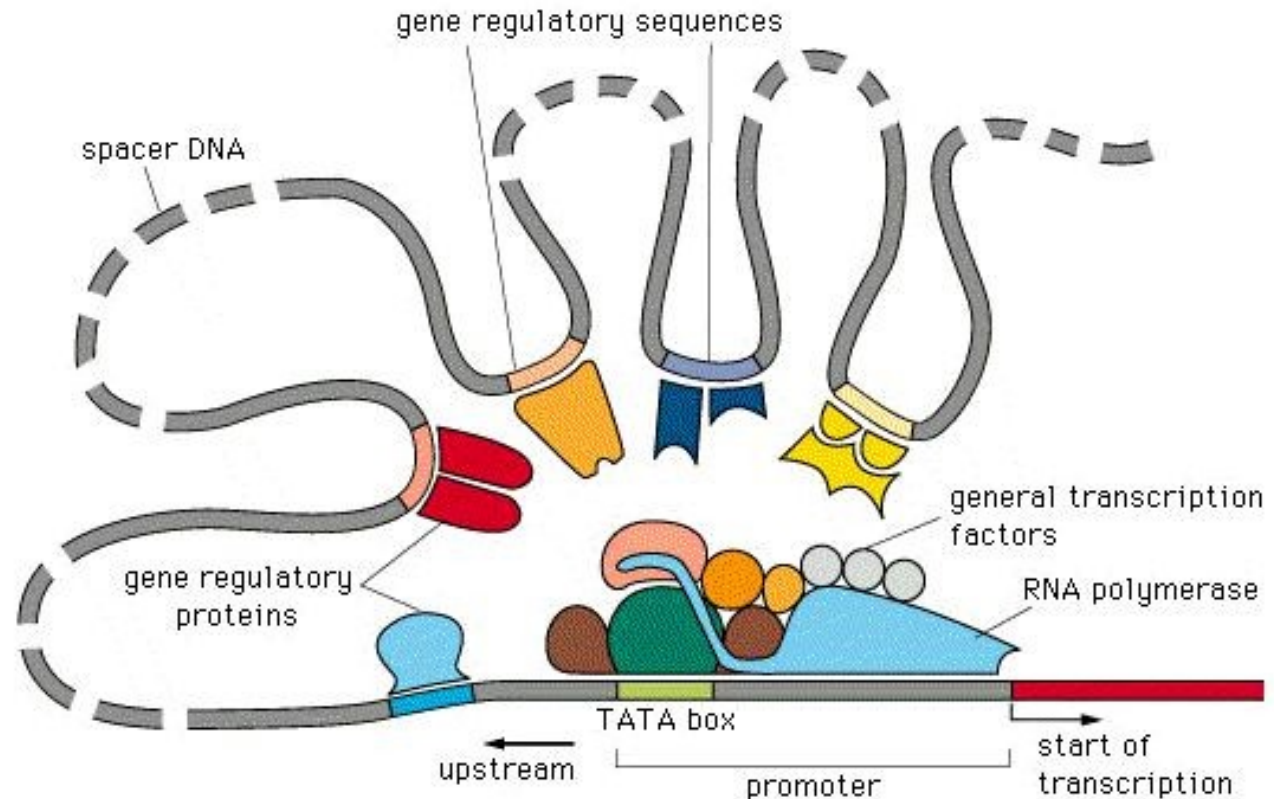


RNA Polymerase Promoter Binding

Enhancer

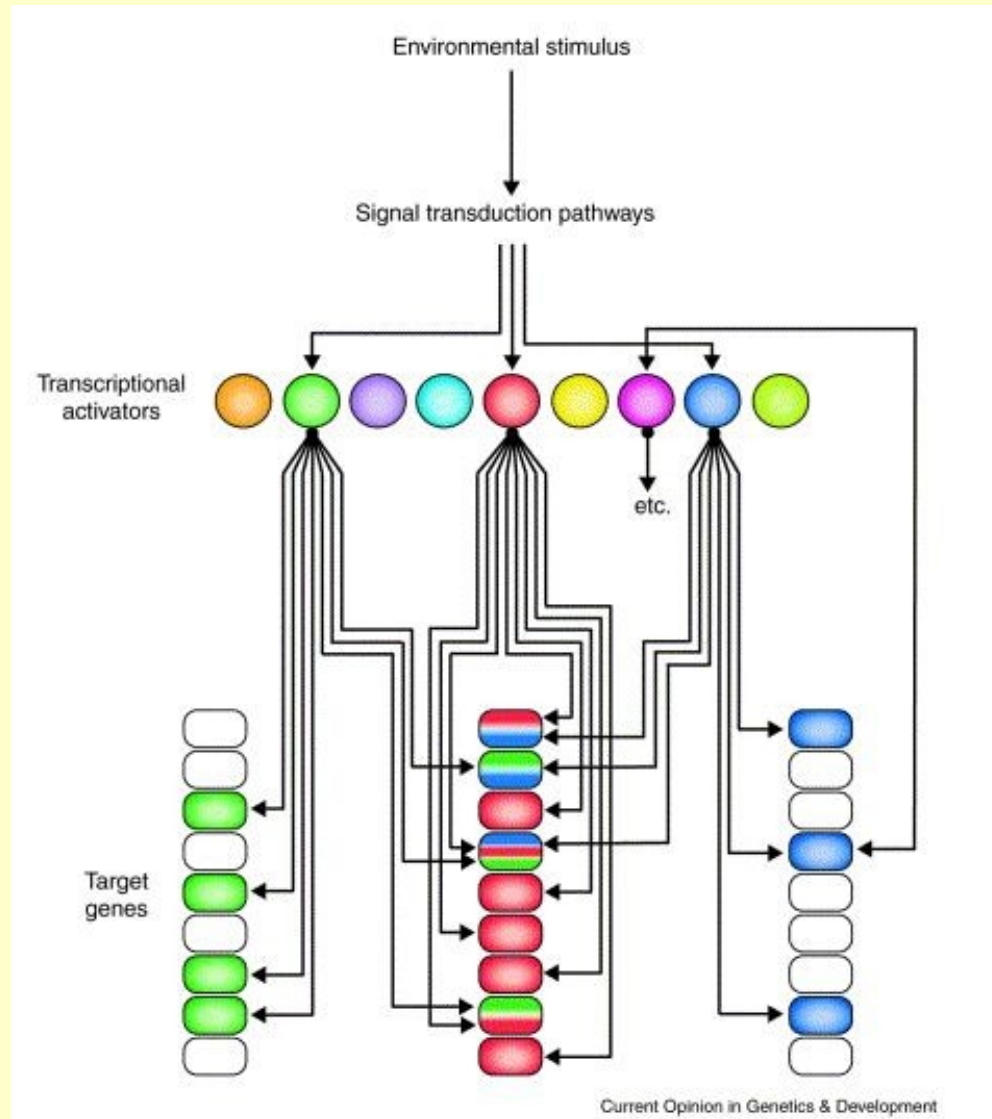


Multiple Enhancer Sequences



©1996 GARLAND PUBLISHING

Gene Expression Regulatory Network



Enhanceosome

http://www.rcsb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/pdb122_1.html

February 2010 **Molecule of the Month** by David Goodsell
Previous Features
doi: [10.2210/rcsb_pdb/mom_2010_2](https://doi.org/10.2210/rcsb_pdb/mom_2010_2)

Enhanceosome

keywords: transcription factor, gene expression, CBP, CREB-binding protein, transcriptional enhancers, enhanceosome

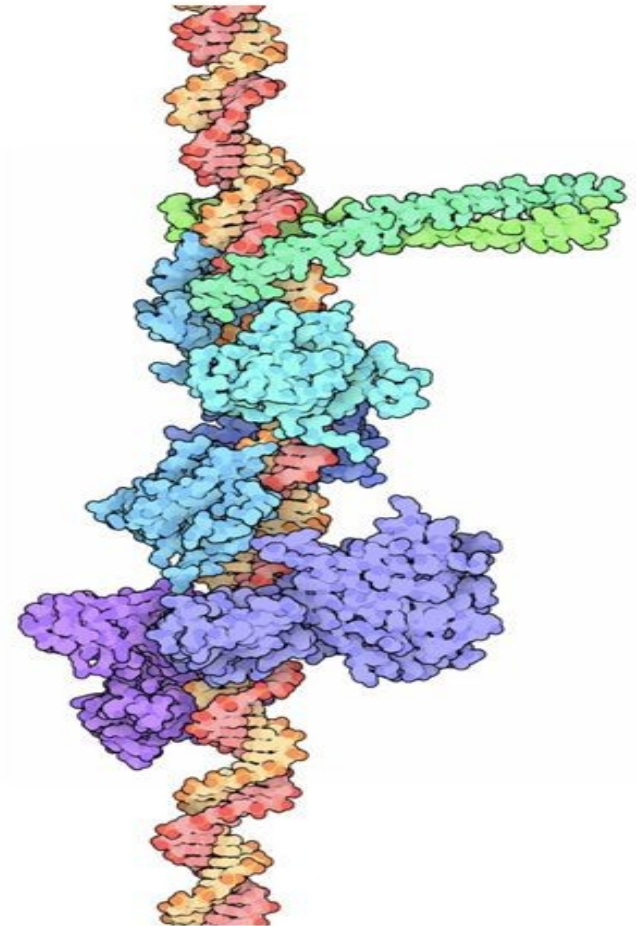
Take a moment to ponder the form of your body: the shape of your face, the color of your eyes, the length of your fingers, the perfect articulation of your bones and muscles, the way your hair grows curly or straight. Now let your imagination travel inward, and think of the complex shapes and functions of your different cells, and the teeming molecular world inside each one. Remarkably, this amazing structure and form and function is specified by information in the genome, which encodes a mere 20,000-25,000 protein-coding genes. One of the great puzzles being pieced together by scientists is the mechanism by which these genes, and the methods used to control their expression, specify all of these different aspects of life.

Combinatorial Control

In order to specify which gene will be expressed in a given situation, your cells use a diverse collection of DNA-binding proteins to control access to the DNA. Surprisingly, there are relatively few of these proteins: by some estimates, the human genome encodes about 2,600 of them. But then, the capabilities of this limited set are greatly expanded by using them in combination, by requiring two or more to bind simultaneously to activate a gene. In this way, each protein may be used in many ways and the spectrum of responses is far more varied.

Enhancing Transcription

The assembly of DNA and proteins pictured here is a transcriptional enhanceosome (PDB entries **1t2k**, **2pi0**, **2o6g** and **2o61**) that controls expression of interferon-beta, an important protein for fighting viral infection. When the cell is infected by viruses, several different DNA-binding proteins are produced, including ATF-2/c-Jun (in green at the top), interferon response factors (IRF, shown in turquoise at the center), and nuclear factor kB (NF-kB, shown in blue and magenta at the bottom). Individually, each one is not sufficient to activate the gene, and each one also plays other roles in the activation of other genes (for instance, Nf-kB is also important in immune responses, inflammation, apoptosis, and many other processes). But when they all bind together, they activate the gene and interferon is made.

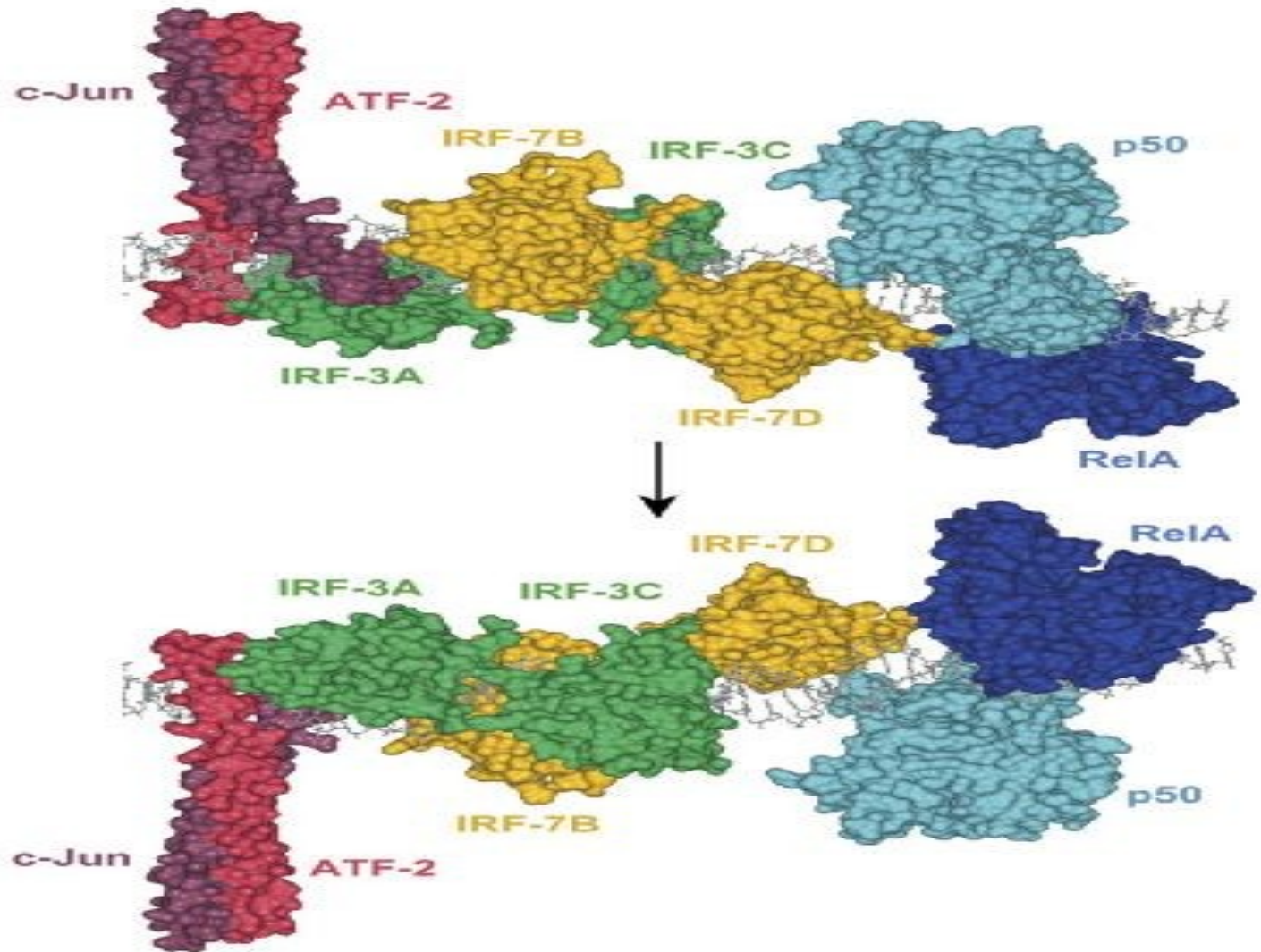


Next: **Integrating the Signal**



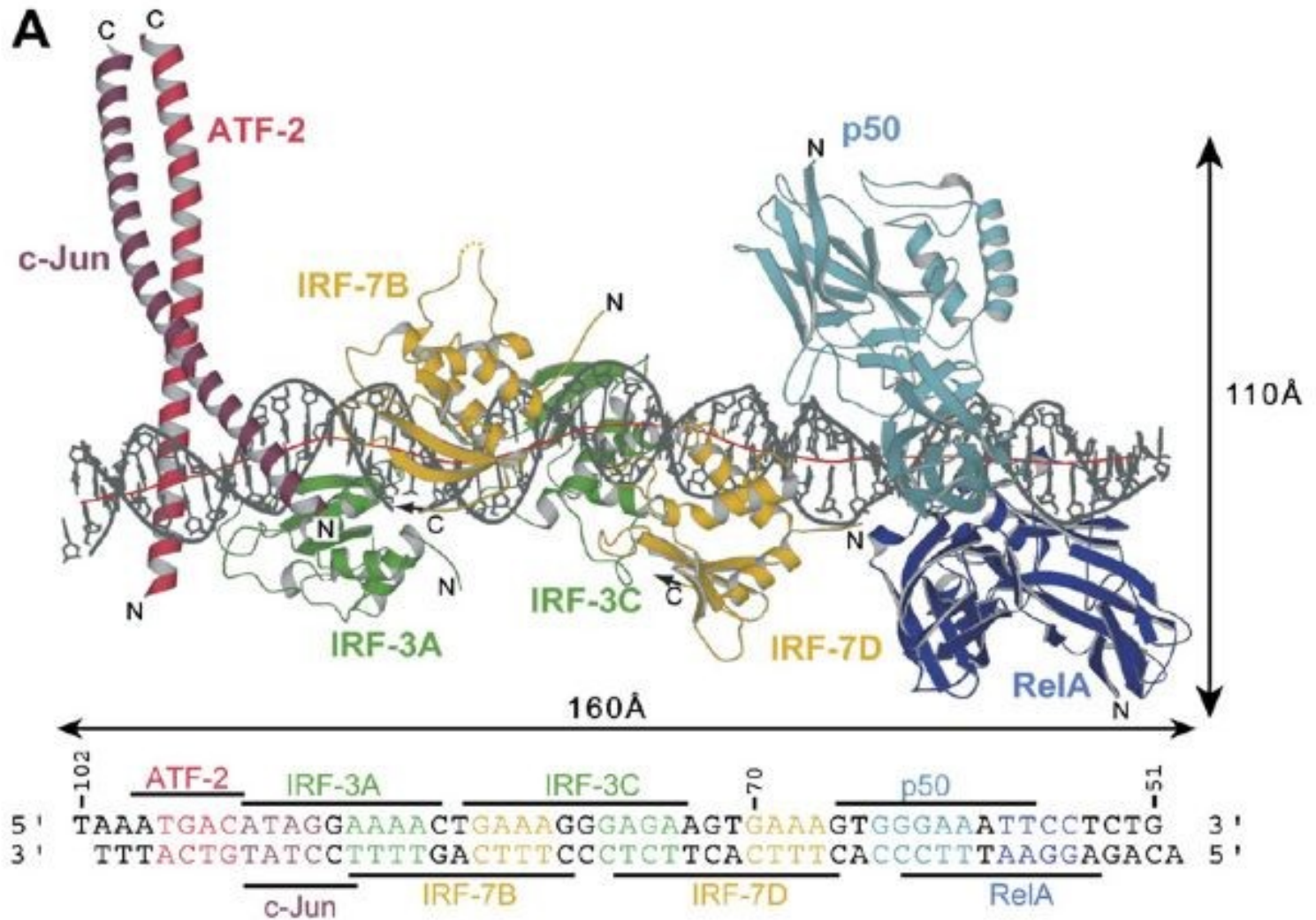
Enhanceosome

<http://www.cell.com/retrieve/pii/S0092867407006563>



Enhanceosome

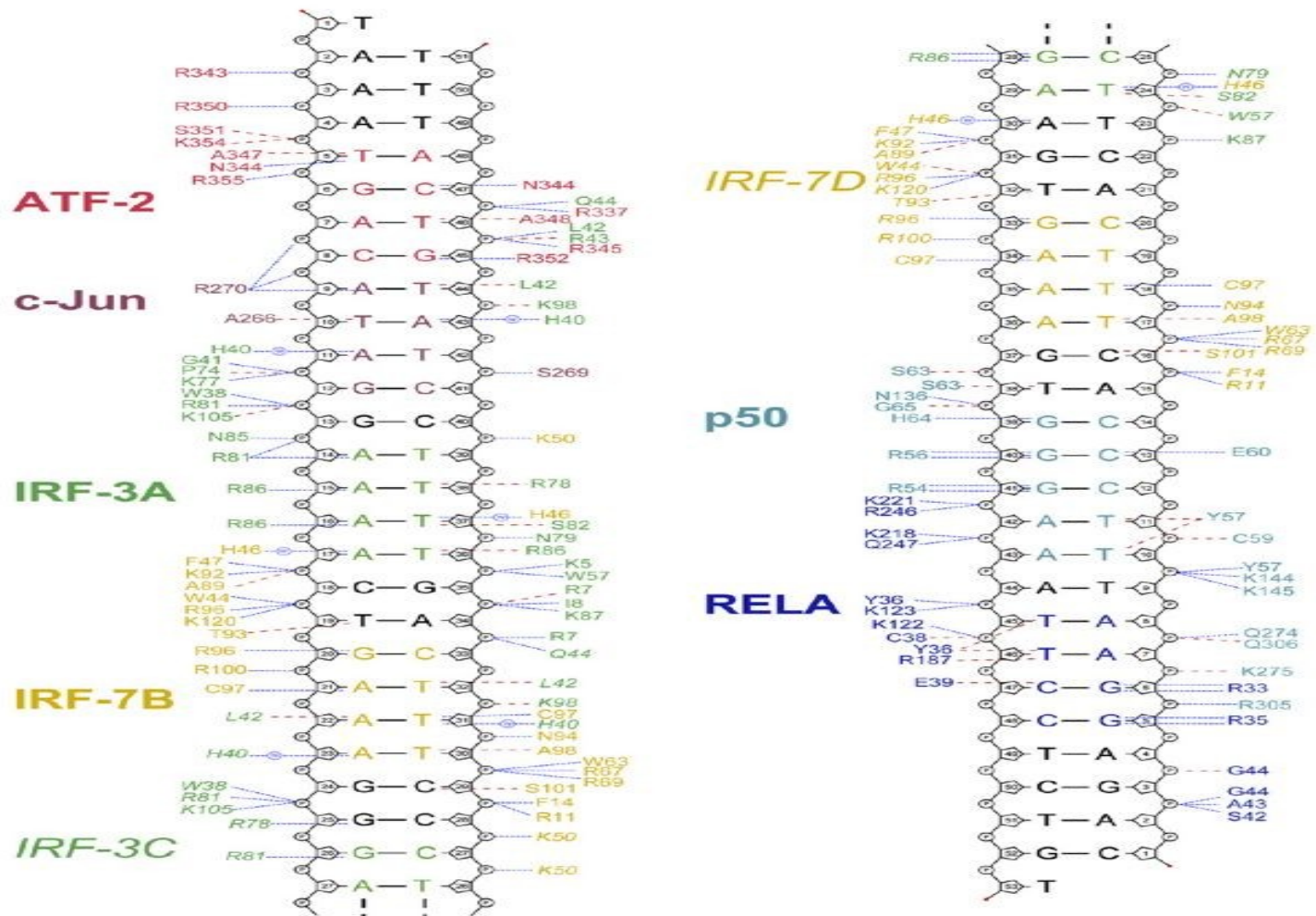
<http://www.cell.com/retrieve/pii/S0092867407006563>



IFN-beta Enhancer

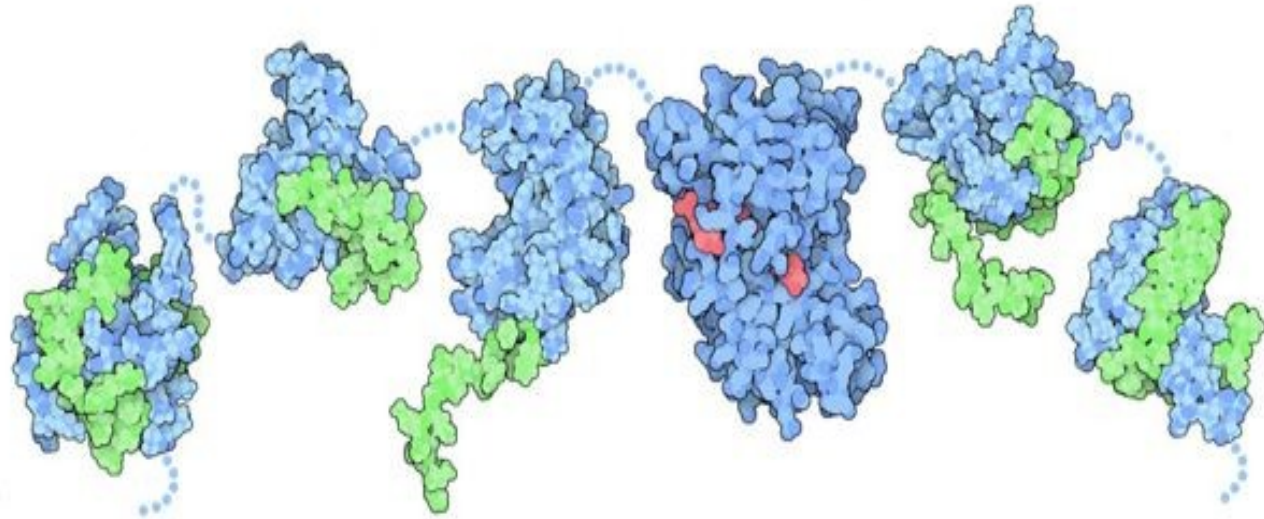
<http://www.cell.com/retrieve/pii/S0092867407006563>

IFN- β enhancer



Enhanceosome: Integrating the Signal

http://www.rcsb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/pdb122_2.html



Integrating the Signal

Once the transcription factors bind to the different sites in the enhancer DNA sequence, the signal must somehow be sensed and used to activate the gene. In many cases, this is performed using CREB-binding protein or the similar protein p300. This protein is composed of many connected domains, (PDB entries **1l8c**, **1kdx**, **1jsp**, **3biy**, **2ka6** and **1kbh**), which bind to different proteins in the assembled enhanceosome. Then, a large domain in the center acts as a histone acetyltransferase, modifying histones in nucleosomes and causing them to disassemble and reveal the gene. In the interferon- β gene, a nucleosome normally covers the start site of transcription, blocking transcription. Assembly of the enhanceosome leads to removal of this nucleosome, allowing the gene to be expressed.

Previous: [Enhanceosome](#)

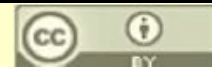
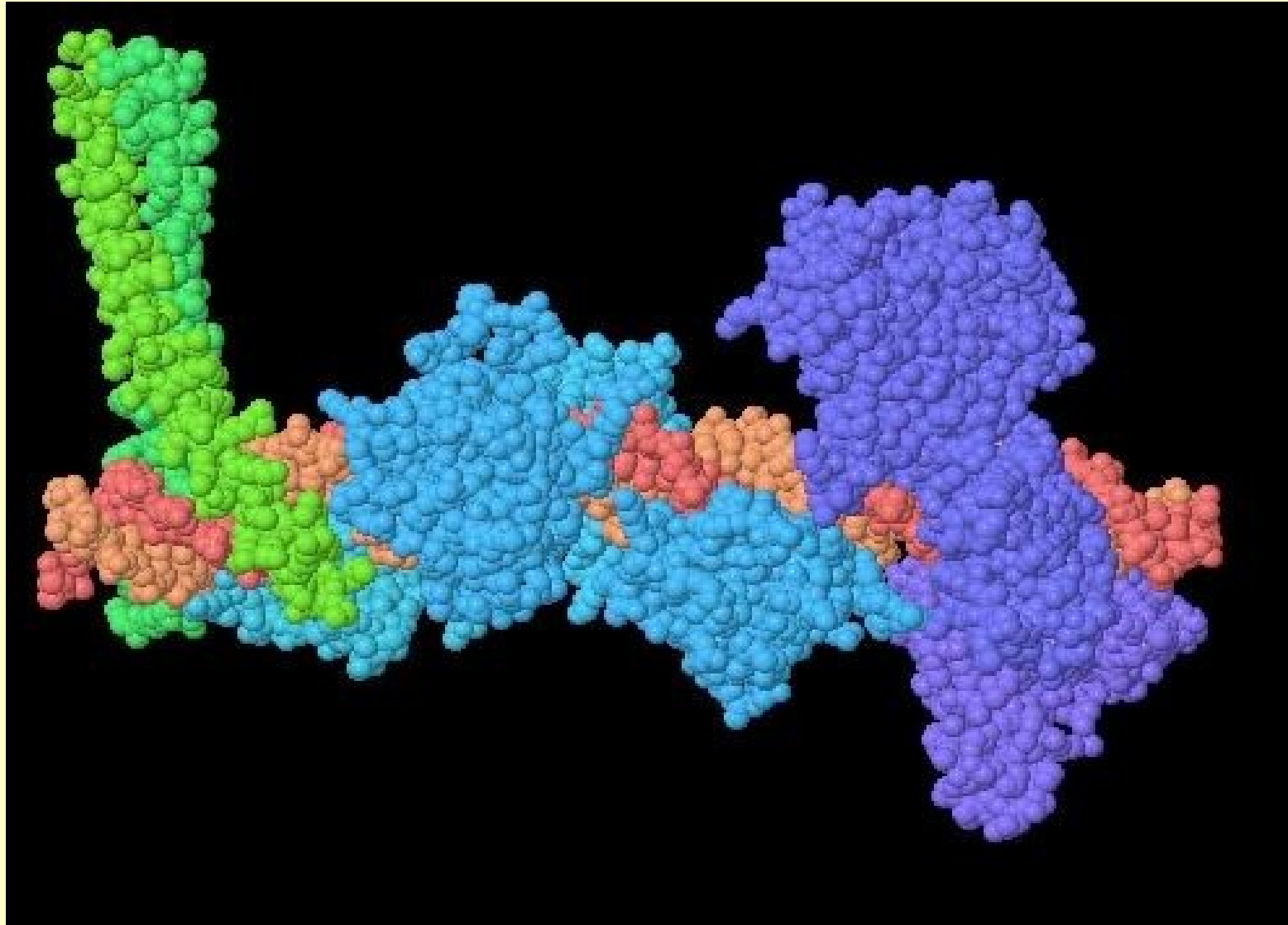
Home: [Enhanceosome](#)

Next: [Exploring the Structure](#)



Enhanceosome: Exploring the Structure

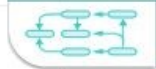
http://www.rcsb.org/pdb/static.do?p=education_discussion/molecule_of_the_month/pdb122_2.html



ht

PATHWAYS ▶ The information-processing pathway at the IFN-beta enhancer

asp


 Submitted by: [Guru](#)
[COMMENT ON THIS PATHWAY](#) | [DESCRIPTION](#) | [CONTRIBUTORS](#) | [SAVE THIS LINK](#) | [SUBMIT](#) | [LEGEND](#)

PRODUCT INDEX

PRODUCT SEARCH

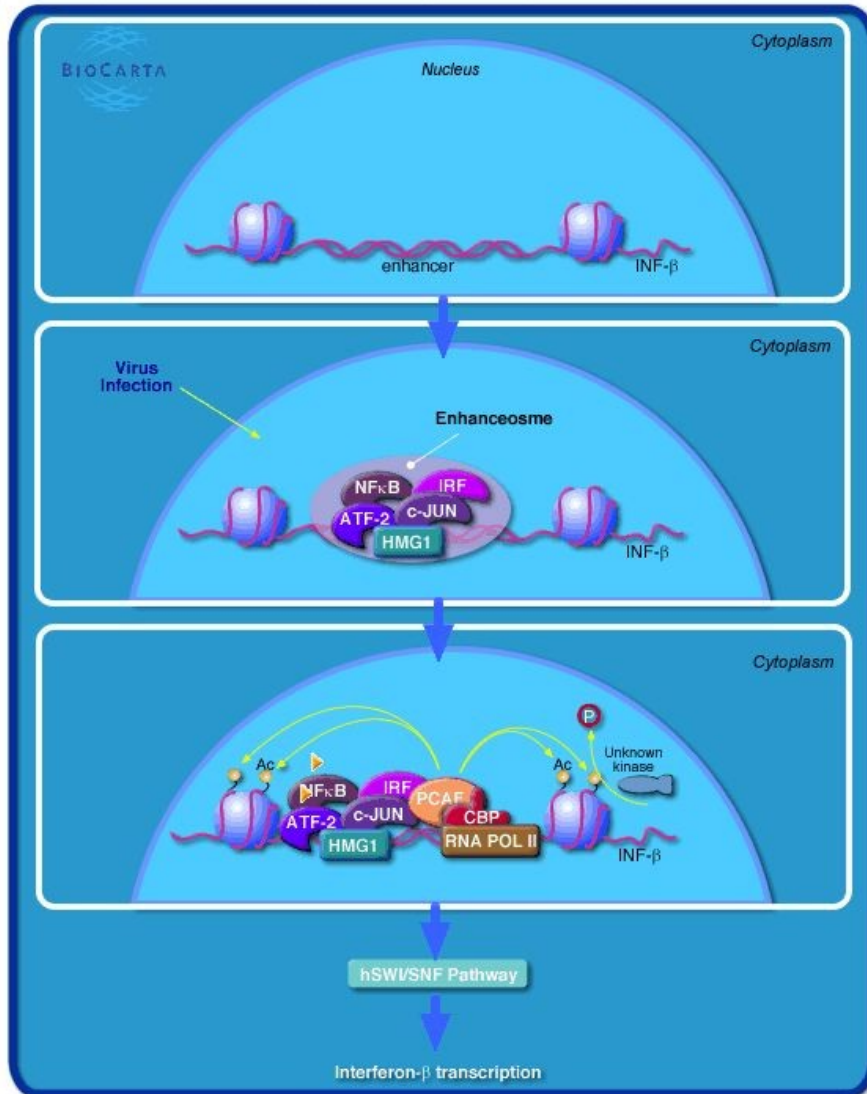
SEARCH

Contains Exact [Advanced Search](#)

PRODUCT HIGHLIGHT ON OFF

PROTEIN LIST

REQUEST A CATALOG



This Pathway:

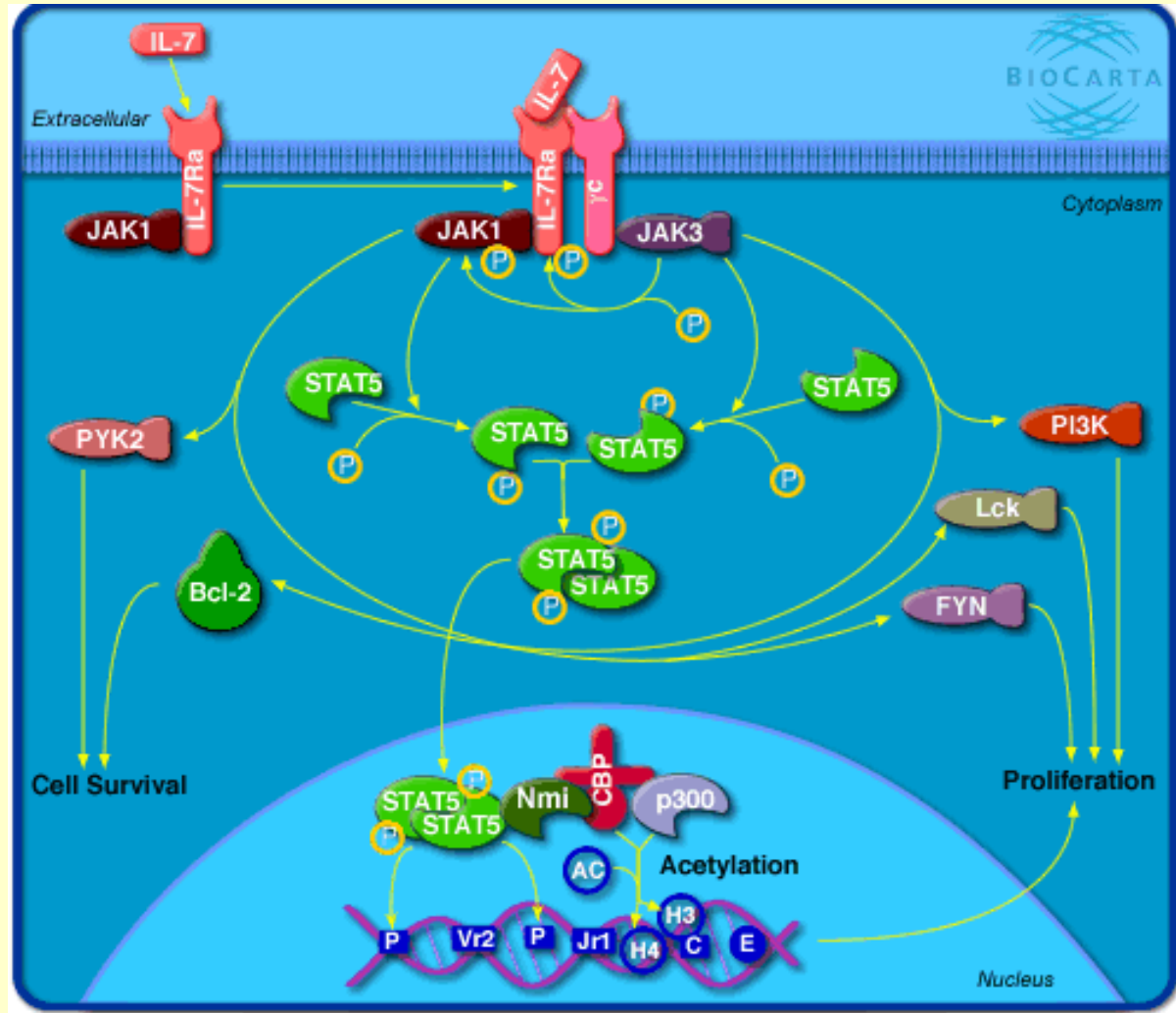


Other Species:



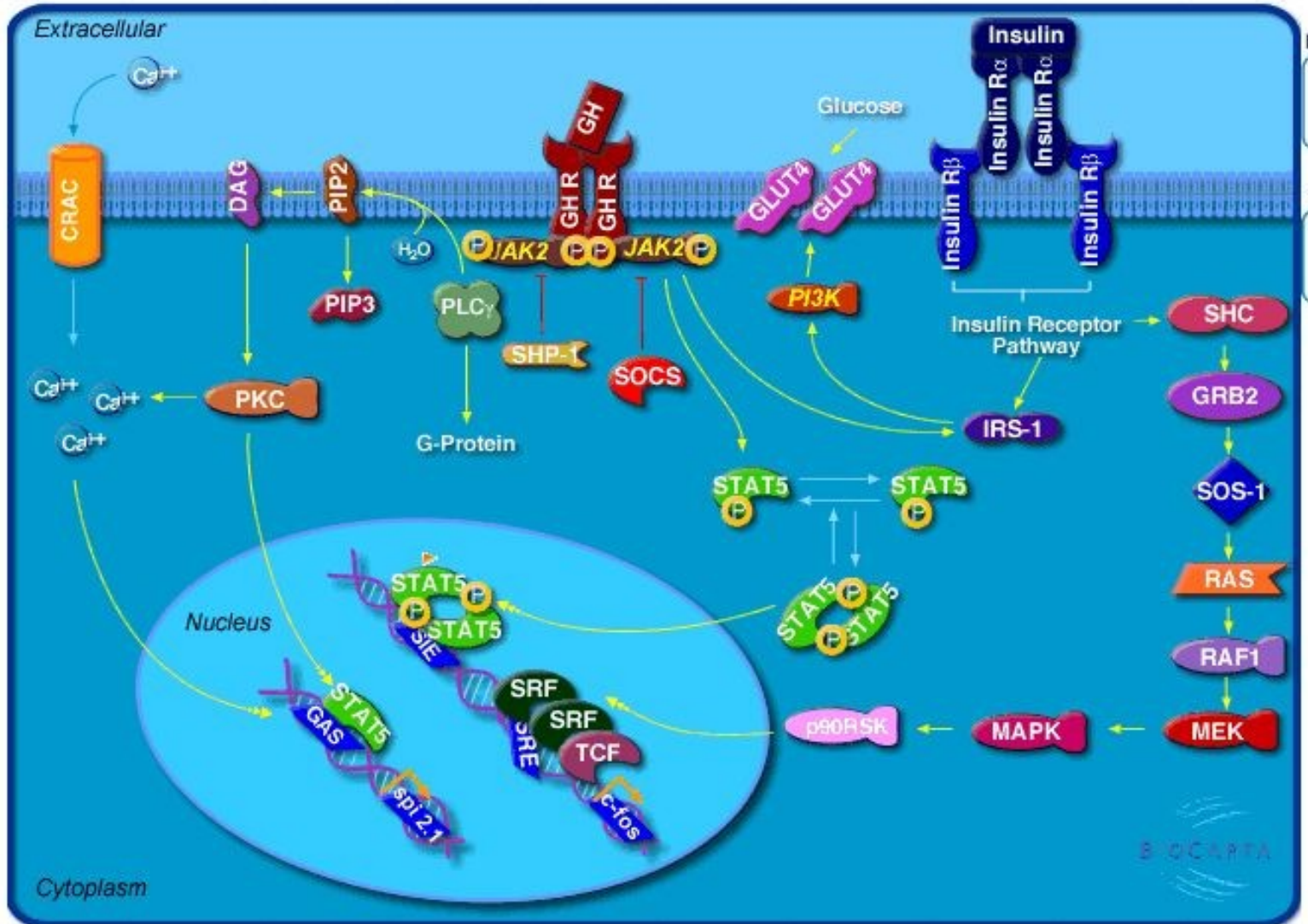
IL7 Regulatory Pathway

http://www.biocarta.com/pathfiles/h_il7Pathway.asp/



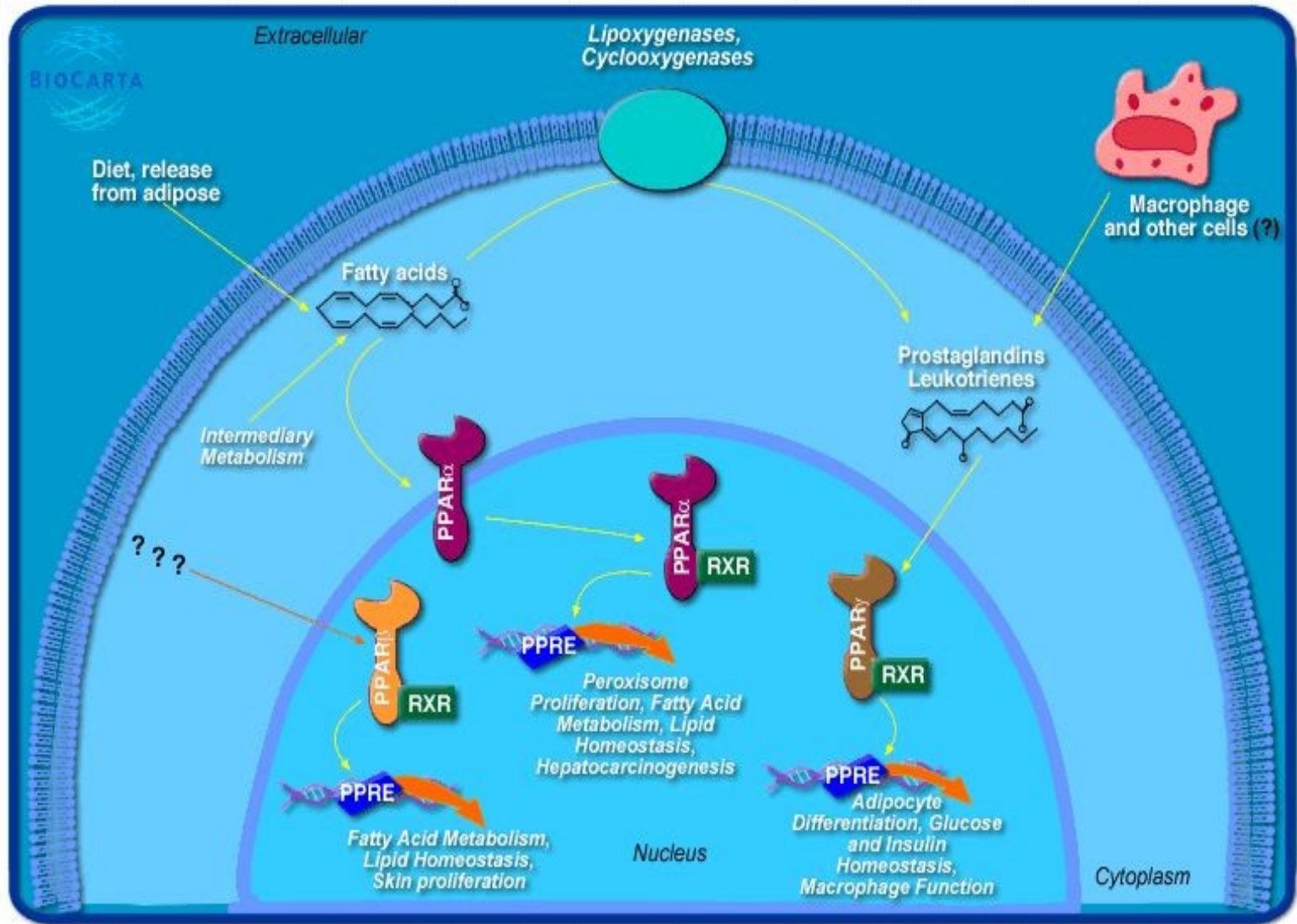
Growth Hormone Receptor Pathway

http://www.biocarta.com/pathfiles/h_ghPathway.asp/



PPAR-Mechanism of Action

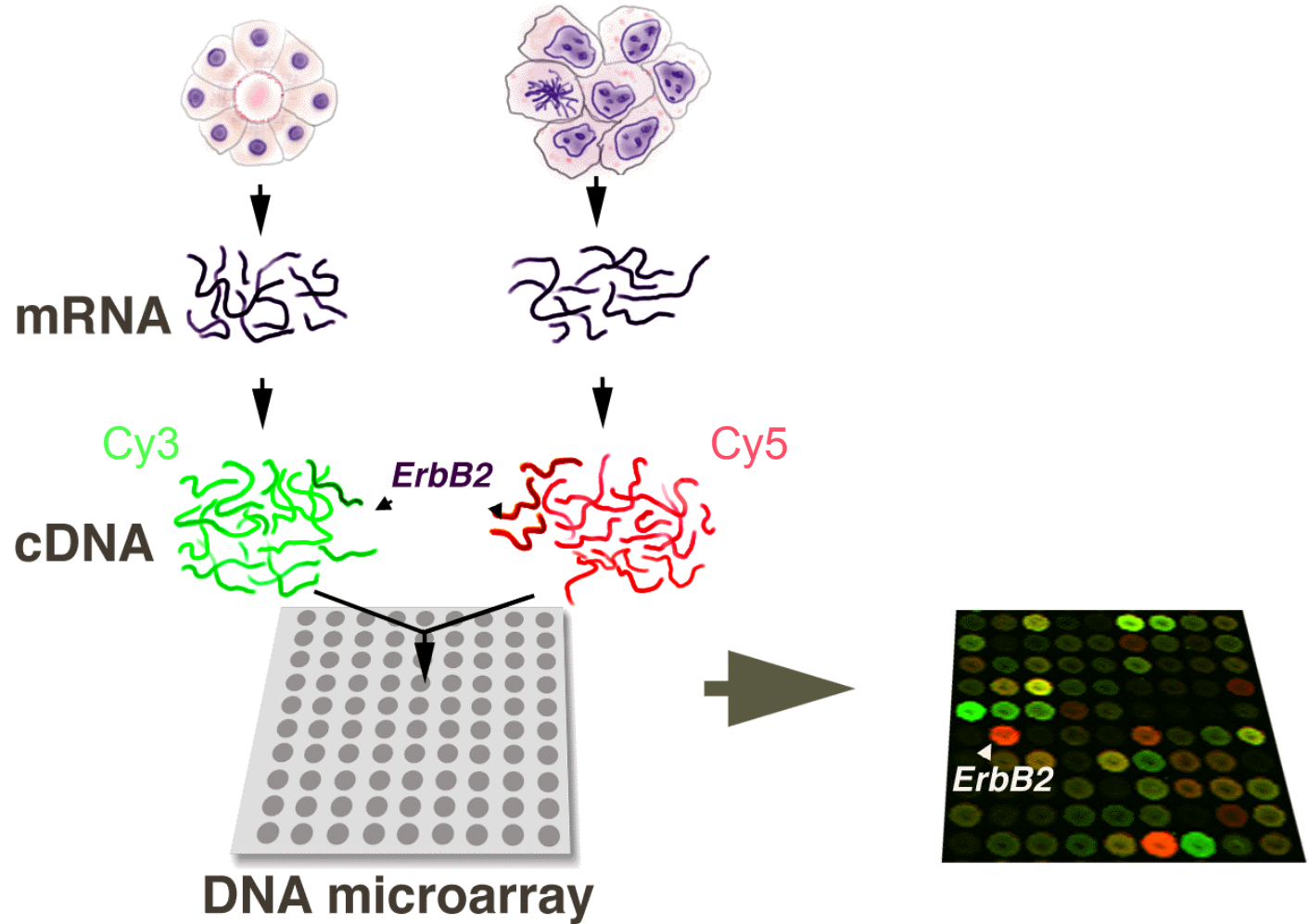
http://www.biocarta.com/pathfiles/h_pparPathway.asp



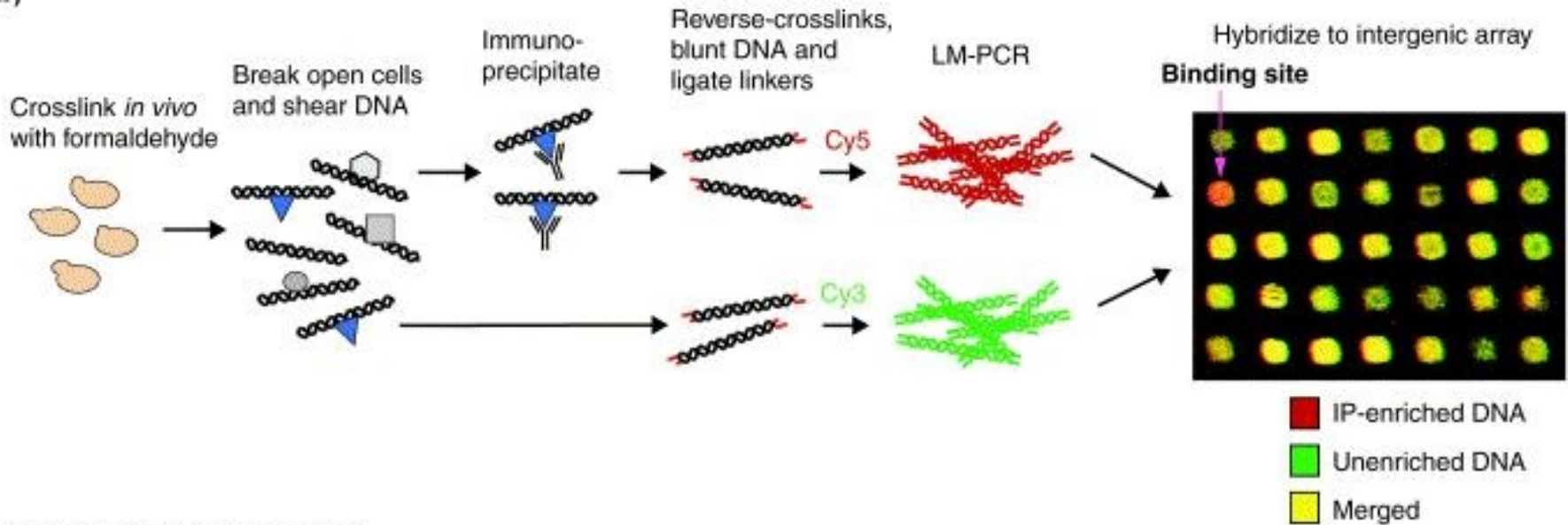
Gene Regulatory Mechanisms

- Transcriptional Mechanisms
 - Type of promoters & RNA polymerase
 - Control of Transcription
 - Constitutive
 - Inducible
 - Repressible
 - Transcription Factors and TFBS
- RNA processing
 - Capping
 - Splicing and Alternative Splicing
 - Poly-Adenylation
 - RNA export to cytoplasm
 - RNA degradation rates
- Translational Mechanisms
 - Micro RNAs (miRNAs)
 - Silencer RNAs (siRNAs or RNAi) degrading mRNA
- Epigenetic Mechanisms
 - Chromatin remodeling
 - Histone acetylation
 - DNA methylation

Measuring Gene Expression with DNA Microarrays and cDNA Labeling

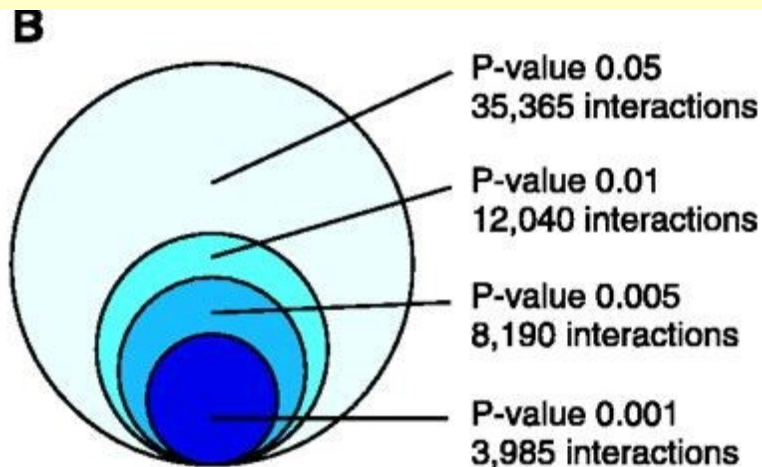
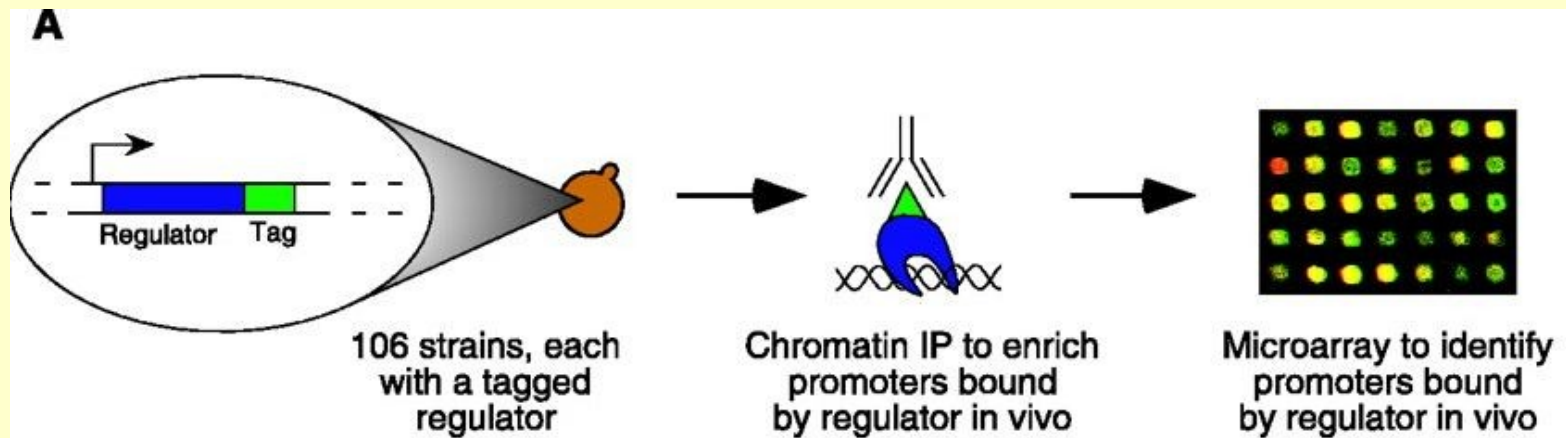


ChIP-Chip Methodology

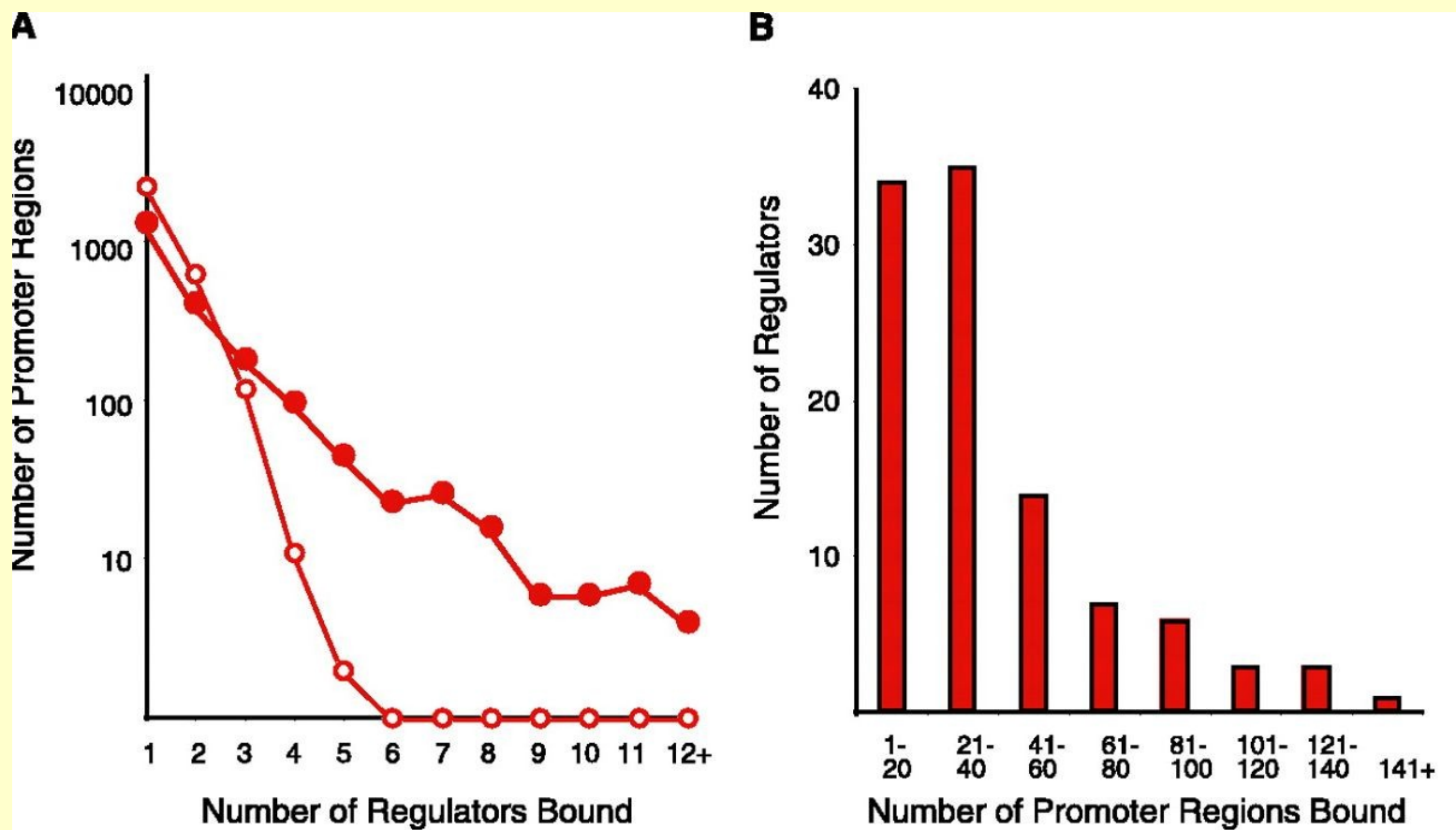


Current Opinion in Genetics & Development

ChIP-Chip Identification of Transcription factor Binding Sites

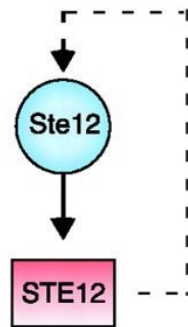


Distribution of TFBS Among 6300 Promoters

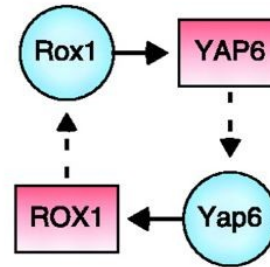


Network Motifs in Yeast Regulatory Networks

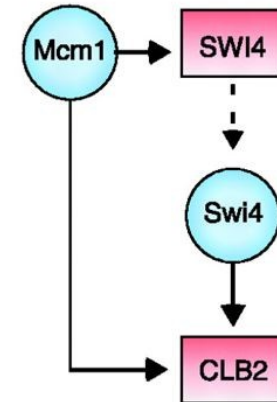
Autoregulation



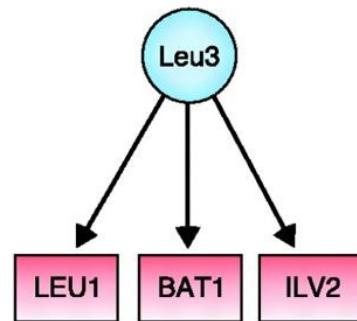
Multi-Component Loop



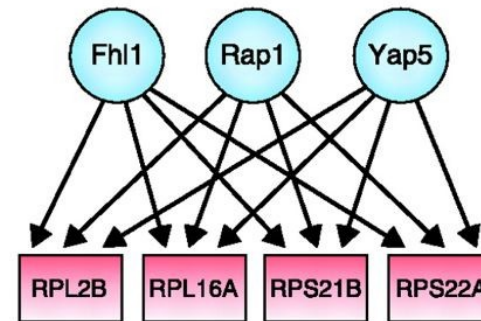
Feedforward Loop



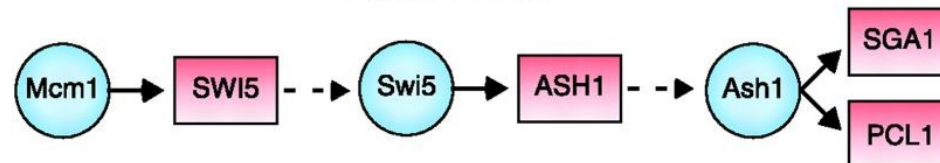
Single Input Motif



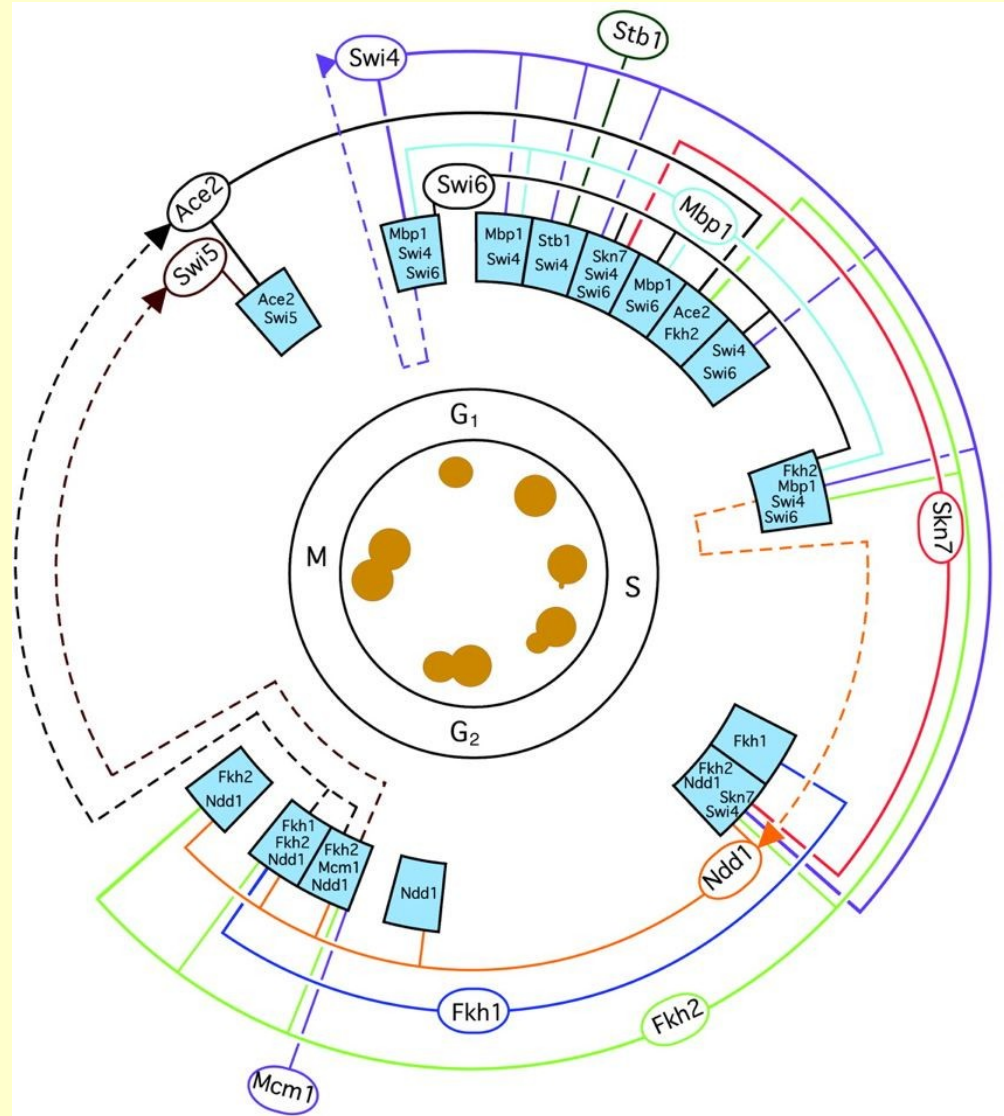
Multi-Input Motif



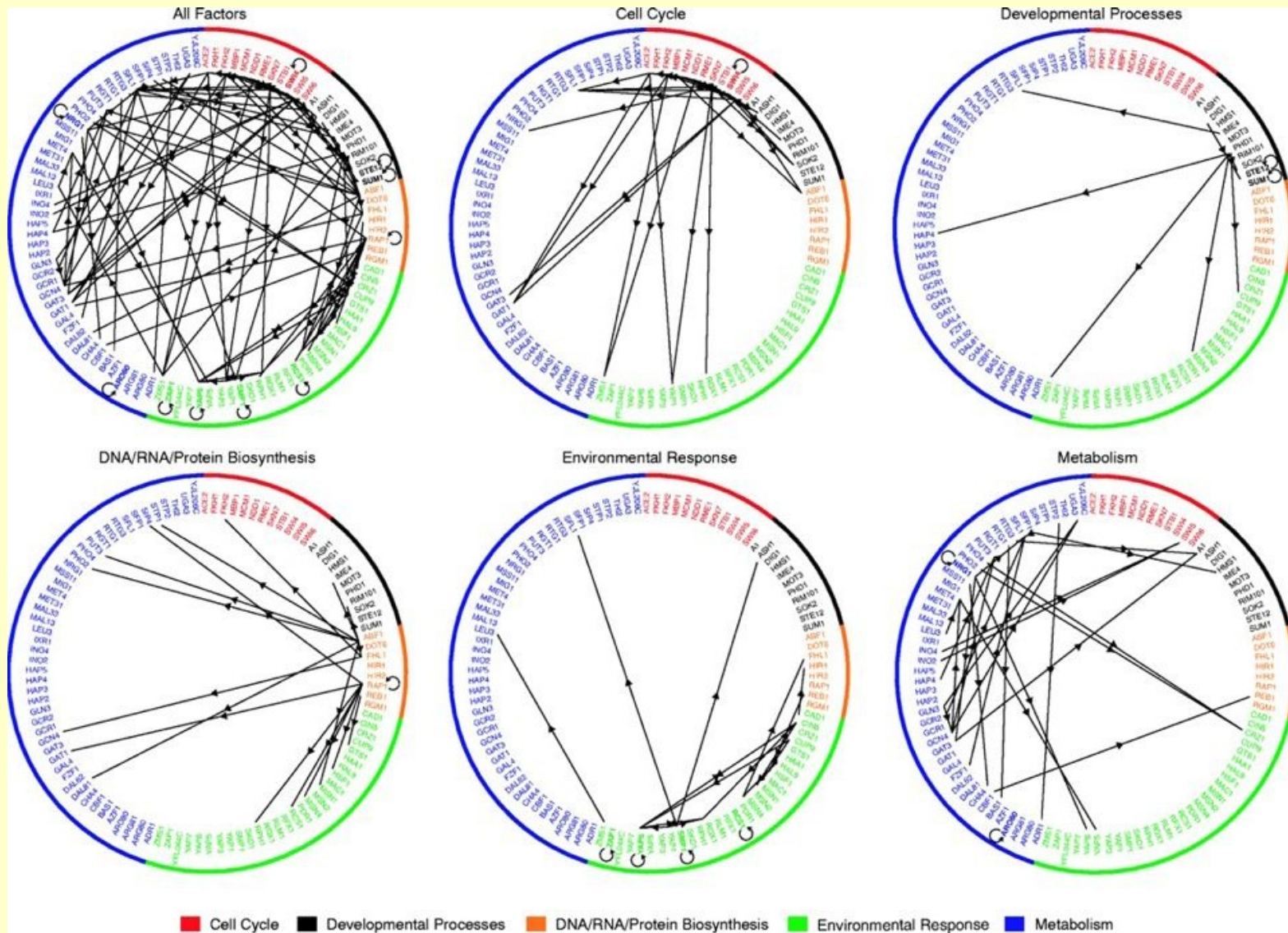
Regulator Chain



Yeast Cell Cycle Regulatory Motifs

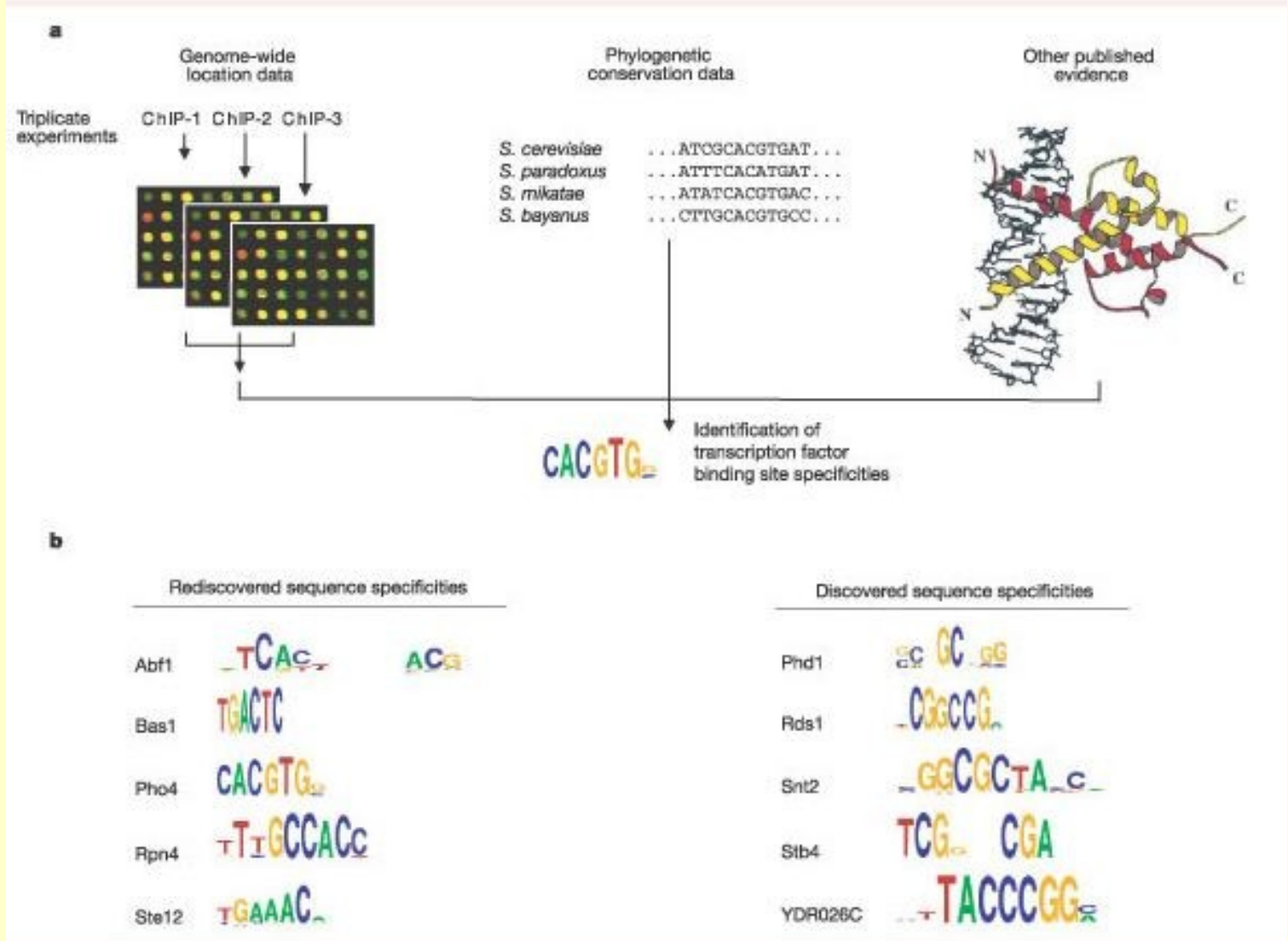


Yeast Regulatory Motifs



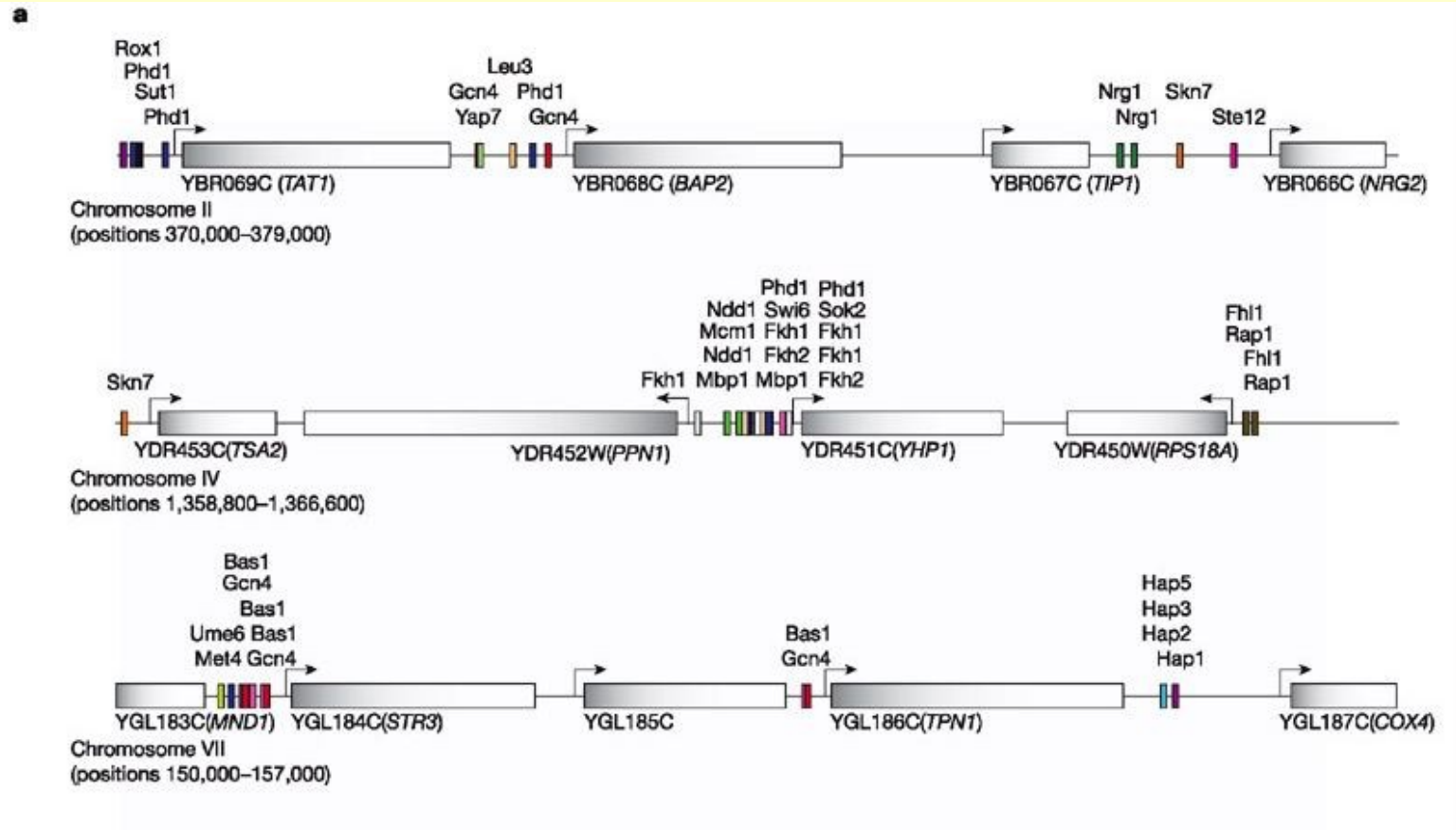
Transcription Factor Binding Site Specificities

http://web.wi.mit.edu/young/regulatory_code/



Yeast Regulatory Map

http://web.wi.mit.edu/young/regulatory_code/



Yeast Regulatory Map

http://web.wi.mit.edu/young/regulatory_code/

b

All sequence matches to DNA binding specificities



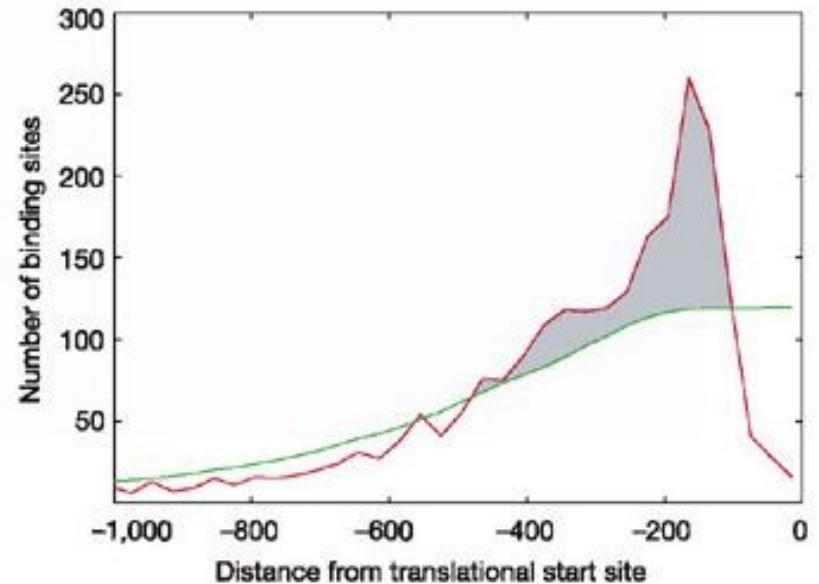
Sequence matches conserved across multiple species



Conserved sequence matches associated with bound regulators



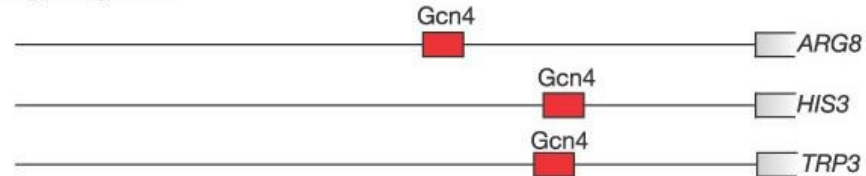
c



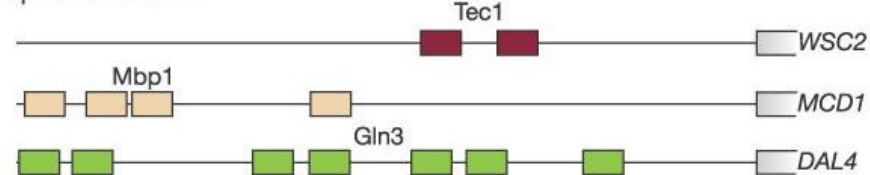
Yeast Promoter Architecture

http://web.wi.mit.edu/young/regulatory_code/

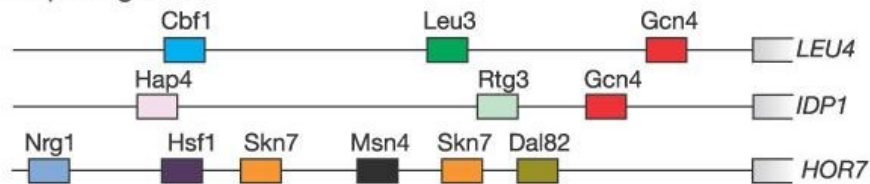
Single regulator



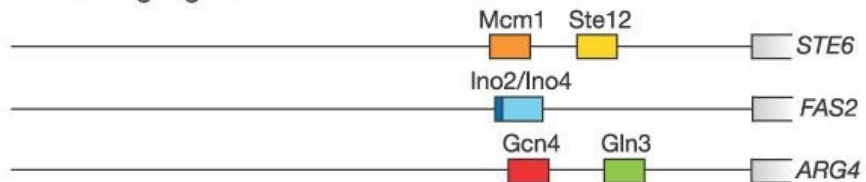
Repetitive motifs



Multiple regulators

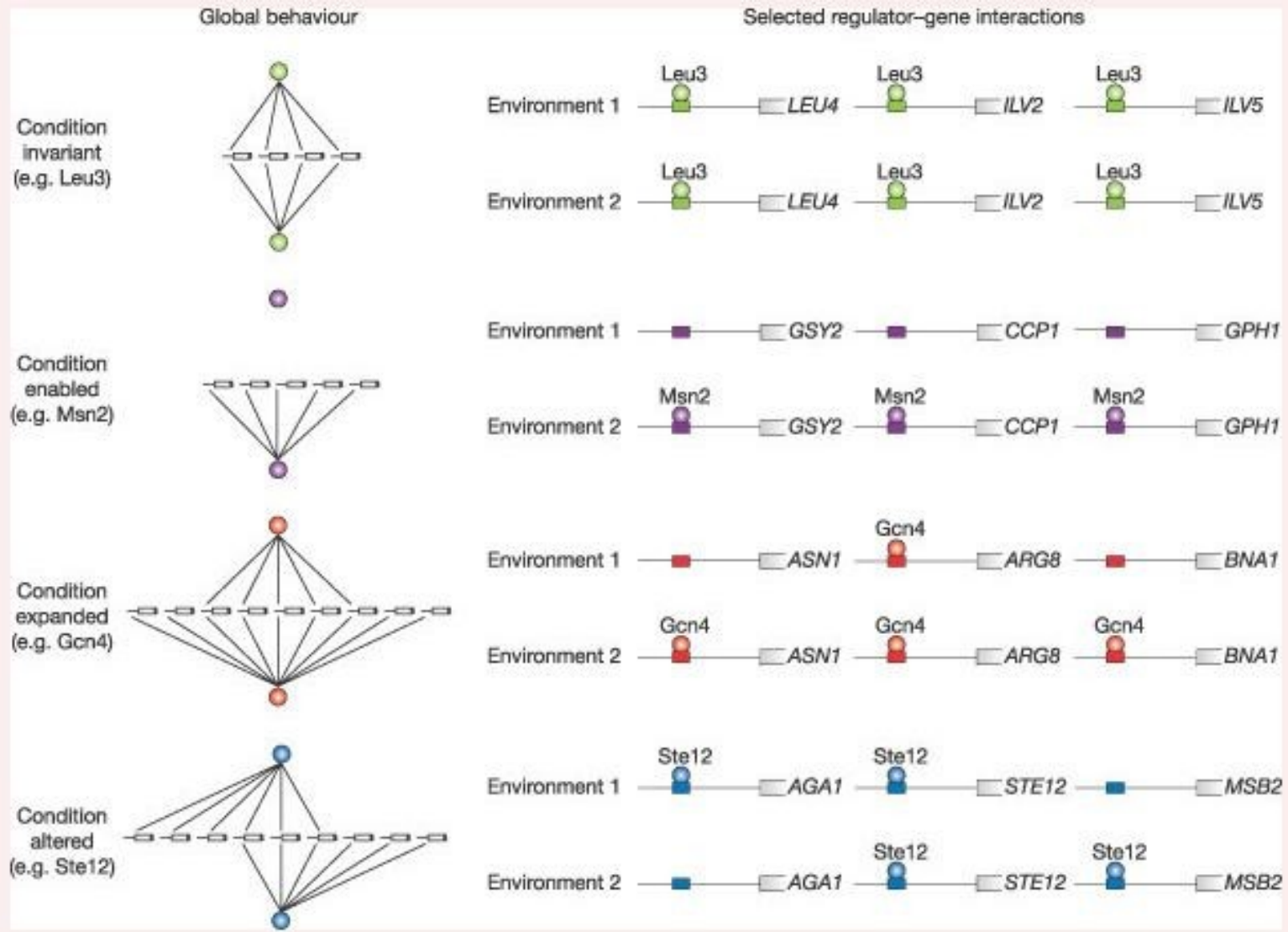


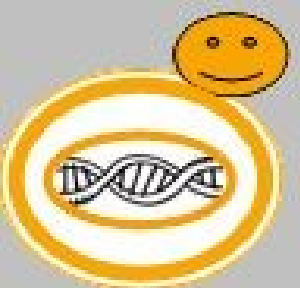
Co-occurring regulators



Environment Specific Regulators

http://web.wi.mit.edu/young/regulatory_code/





YeasTract Transcription Factor Database

<http://www.yeasttract.com/>



[Tutorial](#)

Quick search... [Search](#)

DISCOVERER NEW

Regulatory Associations:

- [Search for TFs](#)
- [Search for Genes](#)
- [Search for Associations](#)

Group genes:

- [Group by TF](#)
- [Group by GO](#)

Pattern Matching:

- [Search by DNA Motif](#)
- [Find TF Binding Site\(s\)](#)
- [Search Motifs on Motifs](#)

Utilities:

- [ORF List ↔ Gene List](#)
- [IUPAC Code Generation](#)
- [Generate Regulation Matrix](#)

Retrieve:

- [TF-Consensus List](#)
- [Upstream Sequence](#)
- [Flat files](#)

About YeasTract:

- [Contact Us](#)
- [Cite YEASTRACT](#)
- [Acknowledgments](#)
- [Credits](#)

Support & suggestions:
yeastract@kdbio.inesc-id.pt

[Home](#)



Welcome to YEASTRACT

[Contact Us](#) - [Tutorial](#) - 



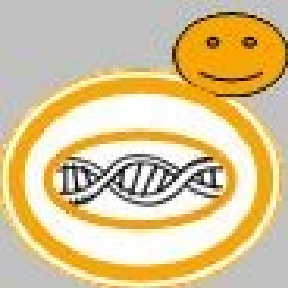
YEASTRACT (**Y**east **S**earch for **T**ranscriptional **R**egulators **A**nd **C**onsensus **T**racking) is a curated repository of more than 34469 regulatory associations between transcription factors (TF) and target genes in *Saccharomyces cerevisiae*, based on more than 1000 bibliographic references. It also includes the description of 291 specific DNA binding sites for more than a hundred characterized TFs. Further information about each Yeast gene has been extracted from the Saccharomyces Genome Database (SGD) as per the latest release, version 1.1438, from Apr 11, 2009. For each gene the associated Gene Ontology (GO) terms and their hierarchy in GO was obtained from the GO consortium as per the OBO flat file from Apr 24, 2009. Currently, YEASTRACT maintains a total of 27250 terms from GO. The nucleotide sequences of the promoter and coding regions for Yeast genes were obtained from Regulatory Sequence Analysis Tools (RSAT). All the information in YEASTRACT will be updated regularly to match the latest data from SGD, GO consortium, RSA Tools and recent literature on yeast regulatory networks.

YEASTRACT now includes [DISCOVERER](#), a set of tools that can be used to identify complex motifs found to be over-represented in the promoter regions of co-regulated genes. DISCOVERER is based on the MUSA and RISO algorithms. These algorithms take as input a list of genes and identify over-represented motifs, which can then be compared with transcription factor binding sites described in the YEASTRACT database.

Facilities are also provided to enable the exploitation of the gathered data when solving a number of biological questions, as exemplified in the Tutorial. YEASTRACT allows the identification of documented or potential transcription regulators of a given gene and of documented or potential regulons for each transcription factor. It also renders possible the comparison between DNA motifs and the transcription factor binding sites described in the literature. The system also provides a useful mechanism for grouping a list of genes (for instance a set of genes with similar expression profiles as revealed by microarray analysis) based on their regulatory associations with known transcription factors.

YEASTRACT provides a set of queries to search and retrieve important biological information from the gathered data and to predict transcription regulation networks in yeast from data emerging from gene-by-gene analysis or global approaches.





YeastTract Transcription Factor Search

<http://www.yeasttract.com/formfindregulators.php>



 [Tutorial](#)

Quick search... [Search](#)

DISCOVERER NEW

Regulatory Associations:

- [Search for TFs](#)
- [Search for Genes](#)
- [Search for Associations](#)

Group genes:

- [Group by TF](#)
- [Group by GO](#)

Pattern Matching:

- [Search by DNA Motif](#)
- [Find TF Binding Site\(s\)](#)
- [Search Motifs on Motifs](#)

Utilities:

- [ORF List ⇌ Gene List](#)
- [IUPAC Code Generation](#)
- [Generate Regulation Matrix](#)

Retrieve:

- [TF-Consensus List](#)
- [Upstream Sequence](#)
- [Flat files](#)

About Yeasttract:

- [Contact Us](#)
- [Cite YEASTRACT](#)
- [Acknowledgments](#)
- [Credits](#)

[Home](#) > [Search Transcription Factors](#)

[Contact Us](#) - [Tutorial](#) - 



Search Transcription Factors

By regulated genes

Search Type	Regulated Gene(s)
<input checked="" type="radio"/> Documented <input type="radio"/> Potential <input type="checkbox"/> Image <input type="radio"/> Both	POL30 MCD1 MSH6 CLN2 TOS4 RAD27 RFA1 CSI2

[Search](#)

[Clear](#)



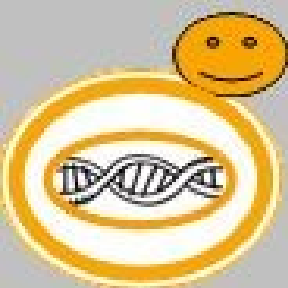
By keyword

Keyword

[Search](#)

[Clear](#)





YeastTract Transcription Factor Search Results

<http://www.yeasttract.com/formfindregulators.php>

[Tutorial](#)

Quick search... [Search](#)

DISCOVERER NEW

Regulatory Associations:

- Search for TFs
- Search for Genes
- Search for Associations

Group genes:

- Group by TF
- Group by GO

Pattern Matching:

- Search by DNA Motif
- Find TF Binding Site(s)
- Search Motifs on Motifs

Utilities:

- [ORF List ⇌ Gene List](#)
- [IUPAC Code Generation](#)
- [Generate Regulation Matrix](#)

Retrieve:

- [TF-Consensus List](#)
- [Upstream Sequence](#)
- [Flat files](#)

About Yeasttract:

- [Contact Us](#)
- [Cite YEASTRACT](#)
- [Acknowledgments](#)
- [Credits](#)

[Home](#) > [Search Transcription Factors](#) > [Result](#)

[Contact Us](#) - [Tutorial](#) -



Search Transcription Factors

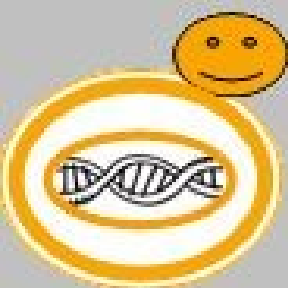


ORF/Gene Name	Documented Transcription Factor - Reference
YBR088c/POL30 Back to top ▲	Mbp1p - Reference + Sfp1p - Reference Swi4p - Reference
YDL003w/MCD1 Back to top ▲	Cin5p - Reference + Mbp1p - Reference Sfp1p - Reference Swi4p - Reference
YDR097c/MSH6 Back to top ▲	Dot6p - Reference + Mbp1p - Reference Sfp1p - Reference Swi6p - Reference
YPL256c/CLN2 Back to top ▲	Gcn4p - Reference + Gcr1p - Reference Mbp1p - Reference Mcm1p - Reference Rme1p - Reference

Jump to ▼

- [YBR088c - POL30](#)
- [YDL003w - MCD1](#)
- [YDR097c - MSH6](#)
- [YPL256c - CLN2](#)
- [YLR183c - TOS4](#)
- [YKLI13c - RAD27](#)
- [YAR007c - RFA1](#)
- [YOL007c - CSI2](#)
- [YOL090w - MSH2](#)
- [YGR152c - RSR1](#)
- [YML027w - YOX1](#)
- [YGR151c](#)
- [YLL022c - HIF1](#)
- [YILO66c - RNR3](#)
- [YOL017w - ESC8](#)
- [YKLO45w - PRI2](#)
- [YDL164c - CDC9](#)
- [YKRO13w - PRY2](#)
- [YNL312w - RFA2](#)
- [YIL140w - AXL2](#)
- [YERO70w - RNR1](#)





YeastTract UpStream Sequences

<http://www.yeasttract.com/formseqretrieval.php>

 **Tutorial**

Quick search...

DISCOVERER NEW

Regulatory Associations:

- [Search for TFs](#)
- [Search for Genes](#)
- [Search for Associations](#)

Group genes:

- [Group by TF](#)
- [Group by GO](#)

Pattern Matching:

- [Search by DNA Motif](#)
- [Find TF Binding Site\(s\)](#)
- [Search Motifs on Motifs](#)

Utilities:

- [ORF List ⇌ Gene List](#)
- [IUPAC Code Generation](#)
- [Generate Regulation Matrix](#)

Retrieve:


- [TF-Consensus List](#)
- [Upstream Sequence](#)
- [Flat files](#)

About Yeastract:

- [Contact Us](#)
- [Cite YEASTRACT](#)
- [Acknowledgments](#)
- [Credits](#)

Support & suggestions:

[Home](#) > [Retrieve Upstream Sequence](#)

[Contact Us](#) - [Tutorial](#) - 



Retrieve Upstream Sequence

All *Saccharomyces Cerevisiae* genes

or for a Gene List


POL30
MCD1
MSH6
CLN2
TOS4
RAD27
RFA1
CSI2

Output:

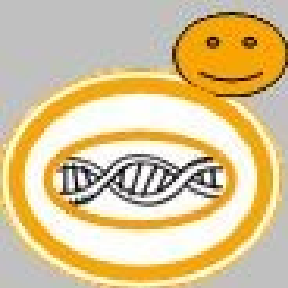
Display

E-mail

From: To:







YeastTract UpStream Sequences Retrieval

<http://www.yeasttract.com/formseqretrieval.php>



[Tutorial](#)

Quick search... [Search](#)

DISCOVERER NEW

Regulatory Associations:

- [Search for TFs](#)
- [Search for Genes](#)
- [Search for Associations](#)

Group genes:

- [Group by TF](#)
- [Group by GO](#)

Pattern Matching:

- [Search by DNA Motif](#)
- [Find TF Binding Site\(s\)](#)
- [Search Motifs on Motifs](#)

Utilities:

- [ORF List ↔ Gene List](#)
- [IUPAC Code Generation](#)
- [Generate Regulation Matrix](#)

Retrieve:

- [TF-Consensus List](#)
- [Upstream Sequence](#)
- [Flat files](#)

About Yeasttract:

- [Contact Us](#)
- [Cite YEASTRACT](#)
- [Acknowledgments](#)
- [Credits](#)

Support & suggestions:
yeasttract@kdbio.inesc-id.pt

[Home](#) > [Retrieve Upstream Sequence](#) > [Result](#)

[Contact Us](#) - [Tutorial](#) -

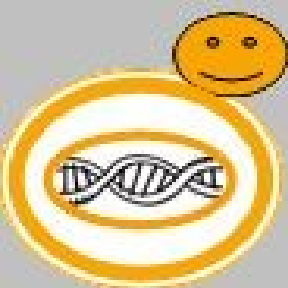


Upstream Sequence




```
>RF11 upstream sequence, from -1000 to -1, size 1000
GTGAATGAAAACGAAGAGGATCACCAGGGTATGCTATAAACTTCCCACAAAACGTCAC
CTGGGGACAAAACGTACCCTTGATCTCTCAATGCCTGTGATAAGAAATAGGCTGCATT
TTCCTGCCTGCTGACTGCTTAAATAACAAATCACTGATATCATCGCGGGATAAGCTGT
CTGAATCACTCCGTATGCTCTTAAAATCATAGAGGTATAGCAATGCAATGAACAAAG
AAGATTCAATTAGCTGGTCATTATTGCTAGATTTATCAAGCCGTTGCAATTTTCAGCTG
TGGTTTATCATGATTCACCGATTTTCTTAAATTTGCAATGTTCAATCTTTTGA
ACTTCTCTGATATTCAAATGACTTGTTCAAAACGATCGTTGACAATTTAGATAGTTGG
CGCCCCGCTTAATGTTATTCCCTCCATAATCTCATCTCCCTCTGTCTTATTAATTTCA
CATCGGCAAGATTGTTCAAATGCAGCCACAGCACATCTTCTAACATAAGCCTCAAAGGTA
CCGACAAAATAACATTTTCTGCTGCTGCAGTAGGCAAAAGTTCAGATAACACACCACAAA
TACCCCATTTCTTAGAAGTTGATGATCTATGCTCAAATACTAGCGGTGGCATAAAAA
CCTCTCGCCCTTTTGGCGTTCTTTTGGATATATGTTAATAGCTATGCGCGTACCACCTTC
TTCAGCCATATACGAAAGATCTATGCTAATTTGGGTGGATTTCTTATCGCATCACTTTCT
TATTCCTAACTTCTTCTCGCTGCCTAAATAAATATCTCTGCTTGGAAAAATGTCACGGG
TAAAAATTAGACGAGGGCAGCAAAGACGCGTGAACATAAAGGTTTACTTGAACAAA
TGTATCACTTTAAGTTCTACACGTAACAAAACAACTATTCACTCTTCTTATACGGTAA
AGGCGAAACCAGCAAGAACAGATATACTTACAAGAG
>POL30 upstream sequence, from -1000 to -1, size 1000
GAGAAGTAAAGAGGGGTAGACAATGATGGTAATCTTATAAACCGGCTACAATGAAGTT
GTAGCAGCAAGGAAGATGATATTTAATACGGTTCAGGTGAAATGAAATAGCCGCCATA
ACGGCATGCTCAAGTTGTAAGTCAGGACTCTAGCTTCTACTGATATCTCTTAAAGGA
CTGCTGTTCTGTGCACCCCTTCCCTTTGTTATCATAGCGCAGCAAGAGTACTAACTA
ATTAACCTAGAACATTAACATATATAAACTAGCGCTATGCGCGCAACTAAAAGACAAA
GCAACCAAGGAACCAAGAAGAGGACCACCAGGAGAAAGAAGGATCCTAACGCCCTAA
GAGCGGTTGTGAGCTTATATGTTCTTGTCAATGAAACAGAGACATGTCCGTTCCGA
GAATCCTGACGTAACCTTTGGCCAAAGTAGCAGAATATTGGGTGAGAGGTGGAAGCCCT
AATGCTGAAGAAAGCAACCCCTATGAATCTAAGGCTCAAGCAGACAAGAAGAGATACGA
ATCTGAAAAGGAATTGTACAATGCTACACGCTGCTGATCTAACAACTCAATGTTCTCTAC
CTCCCATACAACGTAACCTTTGAAACTTCAGAAAGAAAGTTGAGTAGGTTGCATATGTT
ATTATATTACTATTATCATTTTACTTATATATCTCTCTTTCTTTTATGATTTCTC
TTCTTTTTCGCACAACCTTATGCTGATTCTTCAATATCTAATTTTAGCATTTTCTTC
TCCATCCGACGCCAAAACGCGAAACGCGTAACCTTTTTTTTTTTGGATTTCAACTGATA
GTTTTGCTACTTTGCTTCTCTGTTACATAAAATATATATAAGAAACACTTTTGTCTTA
GCCTTCTTTCTTTCCACTTGCACCTTCACTTTTCGCGCTCTTTTCACTCACAGCAAC
AAGCAGCAAGCACTAAGTACGCAGTCAAAGAGAGAAAA
>MCD1 upstream sequence, from -1000 to -1, size 1000
TACAGAAAATTTTATTATTTATATTTCAAAAACAAGAAAACAAAAGAGAAAACATTAACA
TGTAATTTTCACTAATAGTCCTAGAAAAGTAATCTTTTACAACGTTCTTTAAAT
```





YeasTract Motif Discoverer

<http://www.yeastract.com/discoverer/>



[Tutorial](#)

Quick search... [Search](#)

DISCOVERER NEW

Regulatory Associations:

- [Search for TFs](#)
- [Search for Genes](#)
- [Search for Associations](#)

Group genes:

- [Group by TF](#)
- [Group by GO](#)

Pattern Matching:

- [Search by DNA Motif](#)
- [Find TF Binding Site\(s\)](#)
- [Search Motifs on Motifs](#)

Utilities:

- [ORF List ⇌ Gene List](#)
- [IUPAC Code Generation](#)
- [Generate Regulation Matrix](#)

Retrieve:

- [TF-Consensus List](#)
- [Upstream Sequence](#)
- [Flat files](#)

About YeasTract:

- [Contact Us](#)
- [Cite YEASTRACT](#)
- [Acknowledgments](#)
- [Credits](#)

[Home](#) > [Discoverer](#)

[Contact Us](#) - [Tutorial](#)



Welcome to Discoverer at YeasTract



[Introduction](#)

[MUSA](#)

[RISO](#)

[Tutorial](#)

[Credits](#)

[Publications](#)

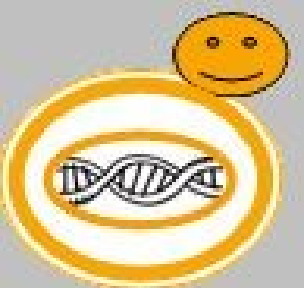
[How to cite](#)

DISCOVERER provides tools for the purpose of motif extraction: the identification of *de novo* binding site consensus sequences from a given set of non-coding DNA sequences (such as the promoter regions of a gene). DISCOVERER contains two distinct structured motif discovery algorithms: MUSA and RISO.

Contrary to many modern motif finders, [MUSA](#) (Motif finding using an UnSupervised Approach) does not require the user to specify parameters (such as box lengths and distances between boxes) in order to extract motifs. The algorithm can therefore either be used autonomously to search for motifs, or to estimate the search parameters to be used in other motif discovery tools. Requiring as input a list of genes, MUSA returns the list of structured or simple motifs found, ordered by their p-value, and the proportion of sequences containing each motif (the quorum).

[RISO](#), another complex motif extraction tool, searches for complex motifs with certain characteristics specified by the user, through the assignment of a set of parameters such as the number and sizes of the boxes that form the structured motif, the distances between them and the minimum quorum expected. It is also possible to specify a number of substitutions for each box. Like MUSA, RISO also requires the list of genes whose promoters are to be searched for motifs, and returns the list of motifs found and their corresponding quorums.





YeasTract Motif Discoverer Results

<http://www.yeasttract.com/discoverer/>





[Tutorial](#)
 [Search](#)
DISCOVERER NEW
Regulatory Associations:
[Search for TFs](#)
[Search for Genes](#)
[Search for Associations](#)
Group genes:
[Group by TF](#)
[Group by GO](#)
Pattern Matching:
[Search by DNA Motif](#)
[Find TF Binding Site\(s\)](#)
[Search Motifs on Motifs](#)
Utilities:
[ORF List ↔ Gene List](#)
[IUPAC Code Generation](#)
[Generate Regulation Matrix](#)
Retrieve:
[TF-Consensus List](#)
[Upstream Sequence](#)
[Flat files](#)
About YeasTract:
[Contact Us](#)
[Cite YEASTRACT](#)
[Acknowledgments](#)
[Credits](#)
Support & suggestions:
yeastract@kdbio.inesc-id.pt

Home > Discoverer > Family Output [Contact Us](#) - [Tutorial](#)

IBB
cebq bSrg













Motif Finder Family Output

 Click to download output file

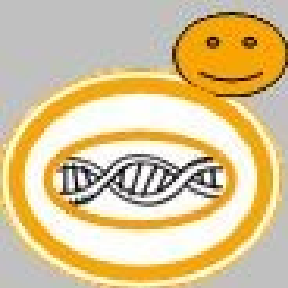
Parameters

Select the column comparison metric:

Select to trim the PWM: Threshold:

Select	Family	PWM	P-value	List of the motifs in the family and its PWM
<input checked="" type="radio"/>	0		8.18697e-15	
<input type="radio"/>	1		8.18697e-15	
<input type="radio"/>	2		6.43356e-06	
<input type="radio"/>	3		6.43356e-06	
<input type="radio"/>	4		7.76378e-06	
<input type="radio"/>	5		8.37493e-05	








Yeast Tract Transcription Factor Associations

<http://www.yeasttract.com/formregassociations.php>

Home > Search Regulatory Associations
Contact Us - Tutorial - ?

Search Regulatory Associations



Tutorial

Quick search... [Search](#)

DISCOVERER NEW

Regulatory Associations:

- [Search for TFs](#)
- [Search for Genes](#)
- [Search for Associations](#)

Group genes:

- [Group by TF](#)
- [Group by GO](#)

Pattern Matching:

- [Search by DNA Motif](#)
- [Find TF Binding Site\(s\)](#)
- [Search Motifs on Motifs](#)

Utilities:

- [ORF List ⇌ Gene List](#)
- [IUPAC Code Generation](#)
- [Generate Regulation Matrix](#)

Retrieve:

- [TF-Consensus List](#)
- [Upstream Sequence](#)
- [Flat files](#)

About Yeasttract:

- [Contact Us](#)
- [Cite YEASTRACT](#)
- [Acknowledgments](#)
- [Credits](#)

Transcription Factors

ORFs/Genes

YGR151C

HIF1

RNR3

ESC8

PRI2

CDC9

PRY2

RFA2

AXL2

RNR1

While searching regulations, consider:

Input TF list against input Gene list

All TFs in the database against input Gene list

Input TF list against all Genes in the database

Search regulations between:

(Each | Every) Transcription Factor to Any Gene

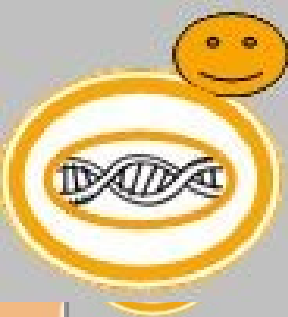
Select Documented Regulations under:

Direct Evidence Codes

Indirect Evidence Codes

Undefined Evidence Codes





YeasTract Transcription Factor Associations

<http://www.yeasttract.com/formregassociations.php>

Home > Search Regulatory Associations > Result

Contact Us - Tutorial -



Regulatory Associations



Quick search... [Search](#)

DISCOVERER NEW

Regulatory Associations:

- [Search for TFs](#)
- [Search for Genes](#)
- [Search for Associations](#)

Group genes:

- [Group by TF](#)
- [Group by GO](#)

Pattern Matching:

- [Search by DNA Motif](#)
- [Find TF Binding Site\(s\)](#)
- [Search Motifs on Motifs](#)

Utilities:

- [ORF List ⇌ Gene List](#)
- [IUPAC Code Generation](#)

[Generate Regulation Matrix](#)

Retrieve:

- [TF-Consensus List](#)
- [Upstream Sequence](#)
- [Flat files](#)

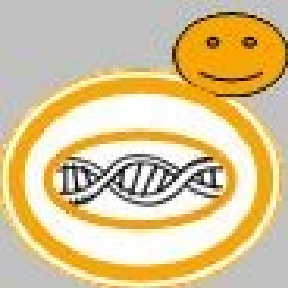
About Yeastract:

- [Contact Us](#)
- [Cite YEASTRACT](#)
- [Acknowledgments](#)
- [Credits](#)

Support & suggestions:

yeastract@kdbio.inesc-id.pt

Transcription Factors	Documented Regulated	Potential Regulated	Jump to
	Genes - Ref	Genes - Promoter	
<p>Abf1p DNA binding protein with possible chromatin-reorganizing activity involved in transcriptional activation, gene silencing, and DNA replication and repair</p>	<p>RFA2 - Reference</p>	<p>HIF1 - Promoter POL30 - Promoter PRI2 - Promoter PRY2 - Promoter RAD27 - Promoter RFA1 - Promoter RFA2 - Promoter RNR1 - Promoter RNR3 - Promoter YOX1 - Promoter</p>	<p>Abf1p Acalp Ace2p Adrip Aft1p Aft2p Arg8Op Arg81p Aro8Op Arr1p Ash1p Azf1p Bas1p Bye1p Cad1p Cat8p Cb1p Cdc14p Chatp Cin5p Crz1p Cst6p Cup2p Cup9p Dal8Op Dal81p Dal82p Dig1p Die2p</p>
<p>Ace2p Transcription factor that activates expression of early G1-specific genes, localizes to daughter cell nuclei after cytokinesis and delays G1 progression in daughters, localization is regulated by phosphorylation; potential Cdc28p substrate</p>	<p>No rows were found!</p>	<p>AXL2 - Promoter CDC9 - Promoter RAD27 - Promoter RFA1 - Promoter RFA2 - Promoter RNR1 - Promoter YGR151c - Promoter</p>	
<p>Adr1p Carbon source-responsive zinc-finger transcription factor, required for transcription of the glucose-repressed gene ADH2, of peroxisomal protein genes, and of genes required for ethanol, glycerol, and fatty acid utilization</p>	<p>RNR3 - Reference</p>	<p>CDC9 - Promoter CLN2 - Promoter CSI2 - Promoter ESC8 - Promoter MCD1 - Promoter MSH2 - Promoter MSH6 - Promoter PRI2 - Promoter RAD27 - Promoter RNR1 - Promoter TOS4 - Promoter YOX1 - Promoter</p>	
<p>Aft1p Transcription factor involved in iron utilization and homeostasis; binds the consensus site PyPuCACCCPu and activates the expression of target genes in response to changes in iron availability</p>	<p>No rows were found!</p>	<p>POL30 - Promoter PRI2 - Promoter RAD27 - Promoter RNR1 - Promoter</p>	
<p>Aft2p Iron-regulated transcriptional activator; activates genes involved in intracellular iron use and required for iron homeostasis and resistance to oxidative stress; similar to Aft1p</p>	<p>No rows were found!</p>	<p>POL30 - Promoter PRI2 - Promoter PRY2 - Promoter RAD27 - Promoter RNR1 - Promoter</p>	



YeastTract Group Genes by Go Terms

<http://www.yeasttract.com/formregassociations.php>

 [Tutorial](#)

Quick search... [Search](#)

DISCOVERER NEW

Regulatory Associations:

- [Search for TFs](#)
- [Search for Genes](#)
- [Search for Associations](#)

Group genes:

- [Group by TF](#)
- [Group by GO](#)

Pattern Matching:

- [Search by DNA Motif](#)
- [Find TF Binding Site\(s\)](#)
- [Search Motifs on Motifs](#)

Utilities:

- [ORF List ⇌ Gene List](#)
- [IUPAC Code Generation](#)
- [Generate Regulation Matrix](#)

Retrieve:

- [TF-Consensus List](#)
- [Upstream Sequence](#)
- [Flat files](#)

About Yeasttract:



- [Contact Us](#)
- [Cite YEASTRACT](#)
- [Acknowledgments](#)
- [Credits](#)

[Home](#) > [Group Genes by GO](#)

[Contact Us](#) - [Tutorial](#) - 

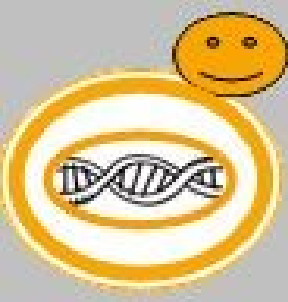


Group Genes by GO

Gene List	YGR151C HIF1 RNR3 ESC8 PRI2 CDC9 PRY2 RFA2 AXL2 RNR1
	Select Ontology <input type="text" value="Biological process"/>
Select Level	<input type="text" value="2"/>
<input type="button" value="Submit"/>	<input type="button" value="Clear"/>  



YeastTract Group Genes by Go Terms



[Home](#) > [Group Genes by GO](#) > [Result](#)

[Contact Us](#) - [Tutorial](#) -



GO Associations for Genes Using Biological Process at level 2



Quick search... [Search](#)

DISCOVERER NEW

Regulatory Associations:

- [Search for TFs](#)
- [Search for Genes](#)
- [Search for Associations](#)

Group genes:

- [Group by TF](#)
- [Group by GO](#)

Pattern Matching:

- [Search by DNA Motif](#)
- [Find TF Binding Site\(s\)](#)
- [Search Motifs on Motifs](#)

Utilities:

- [ORF List ⇌ Gene List](#)
- [IUPAC Code Generation](#)

[Generate Regulation Matrix](#)

Retrieve:

- [TF-Consensus List](#)
- [Upstream Sequence](#)
- [Flat files](#)

About Yeastract:

- [Contact Us](#)
- [Cite YEASTRACT](#)
- [Acknowledgments](#)
- [Credits](#)

Go terms	Percentage	N. Genes	Genes/ORF
cellular process	90.0 %	18	Rsr1 Axl2 Yox1 Cdc9 Pri2 Rfa2 Tos4 Rad27 Cln2 Esc8 Hif1 Msh6 Pry2 Mcd1 Pol30 Msh2 Rnr1 Rnr3 +
metabolic process	65.0 %	13	Axl2 Yox1 Cdc9 Pri2 Rfa2 Rad27 Hif1 Pry2 Pol30 Msh6 Rnr1 Rnr3 Tos4 +
biological regulation	30.0 %	6	Cln2 Esc8 Yox1 Pol30 Rnr1 Rsr1 +
developmental process	20.0 %	4	Axl2 Mcd1 Msh6 Rad27 +
response to stimulus	20.0 %	4	Rnr1 Tos4 Cln2 Rnr3 +
reproductive process	10.0 %	2	Msh6 Rad27 +
localization	5.0 %	1	Rfa2 +
establishment of localization	5.0 %	1	Rfa2 +
multi-organism process	5.0 %	1	Cln2 +
reproduction	5.0 %	1	Axl2 +
anatomical structure formation	5.0 %	1	Hif1 +

[Go Back](#)

Services:

[Swissregulon](#)

[Phylogibbs](#)

[MARA](#)

[TCS](#)

Publications

Links

Contact

Swissregulon is a database of genome-wide annotations of regulatory sites. The predictions are based on Bayesian probabilistic analysis of a combination of input information including:

- Experimentally determined binding sites reported in the literature.
- Known sequence-specificities of transcription factors.
- ChIP-chip and ChIP-seq data.
- Alignments of orthologous non-coding regions.

Predictions were made using the PhyloGibbs, MotEvo, IRUS and MARA algorithms developed in our group, depending on the data available for each organism. Annotations can be viewed in a Gbrowse genome browser and can also be downloaded in flat file format.

Please cite this site as following:

SwissRegulon: a database of genome-wide annotations of regulatory sites
Mikhail Pachkov, Ionas Erb, Nacho Molina and Erik van Nimwegen
Nucleic Acids Research, 2007, Vol. 35, Database issue D127-D131

Swiss Regulon Organisms

<http://www.swissregulon.unibas.ch/cgi-bin/regulon>

Services:

[Swissregulon](#)

[Phylogibbs](#)

[MARA](#)

[TCS](#)

Publications

Links

Contact

All | [Eukaryotes](#) | [Prokaryotes](#)

- Homo Sapiens
- Mus Musculus
- Saccharomyces cerevisiae
- Agrobacterium tumefaciens
- Bacillus subtilis
- Brucella suis
- Burkholderia
- Chlamydomonas reinhardtii
- Corynebacterium glutamicum
- Ehrlichia canis
- Escherichia coli
- Mycobacterium tuberculosis
- Neisseria meningitidis
- Prochlorococcus marinus
- Pseudomonas syringae
- Ralstonia eutropha
- Rickettsia typhi wilmington
- Staphylococcus aureus
- Streptococcus pneumoniae
- Vibrio cholerae

Swiss Regulon Yeast TF Binding Sites Near Pol30

<http://www.swissregulon.unibas.ch/cgi-bin/gbrowse/yeast/>

Showing 1.504 kbp from chrII, positions 424,620 to 426,123

Instructions: Search using a sequence name, gene name, locus, or other landmark. The wildcard character * is allowed. To center on a location, click Scroll/Zoom buttons to change magnification and position.

Examples: chrI, chrII:80,000..120,000, Mit, NPY1, NAB2, YGL123*.

[\[Hide banner\]](#) [\[Bookmark this\]](#) [\[Link to Image\]](#) [\[High-res Image\]](#) [\[Help\]](#)

Search

Landmark or Region:

chrII:424620..426123

Data Source

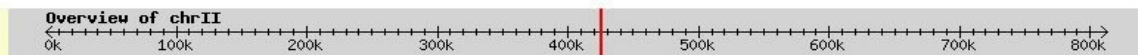
Saccharomyces cerevisiae (Apr 14 2006)

Reports & Analysis:

Annotate Restriction Sites

Scroll/Zoom: Show 1.504 kbp Flip

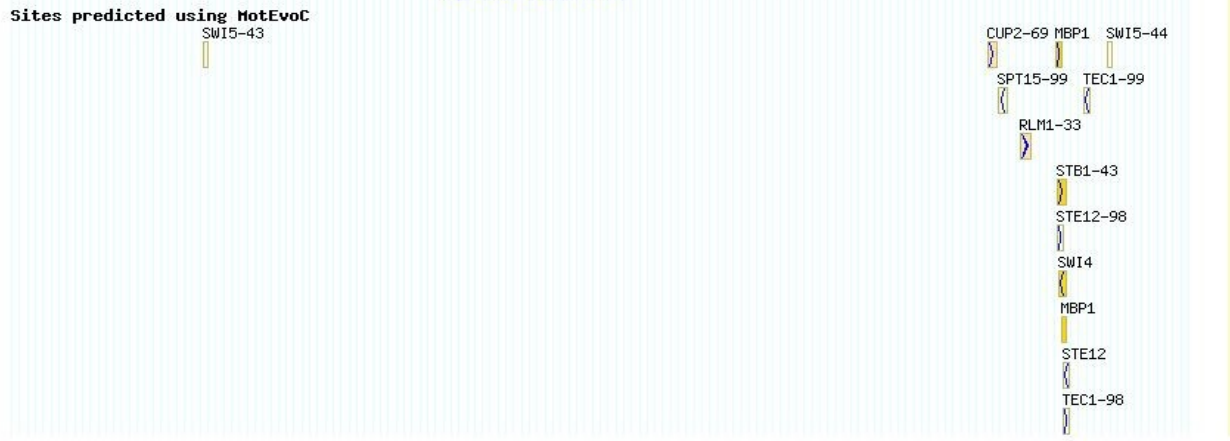
Overview



Details



Named gene
 YBR087W
 RFC5 : Subunit of heteropentameric Replication factor C (RF-C), which is a DNA binding protein and ATPase that acts as a clamp loader of DNA polymerase delta.
 YBR088C
 POL30 : Proliferating cell nuclear antigen (PCNA), functions as the sliding clamp for DNA polymerase delta.
 YBR089W
 Hypothetical protein



Swiss Regulon Yeast Transcription Factor Ste12

<http://www.swissregulon.unibas.ch/cgi-bin/gbrowse/yeast/>

Saccharomyces cerevisiae (Apr 14 2006)

[back to main page](#)

```

Name: STE12_DIG1-68
Class: TF_binding_site
Type: TF_binding_site
Source: MotEvoC
Position: chrII:425967..425961 (- strand)
Length: 7
Score: 0.071272
Alias: STE12_DIG1
Note: CGTTTTG
factor_name: STE12_DIG1
weight_matrix: 0.00 32.62 0.96 250.22 42.97 0.00 241.37 1.98 271.22 0.00 12.78 0.00 252.90 1.67 27.03 0.00 278.87 0.00 0.00
                0.00 0.98 277.79 0.00 0.97 153.01 30.02 93.87 22.75
    
```

```

>STE12_DIG1-68 class=TF_binding_site position=chrII:425967..425961 (- strand)
CCAAAAC
    
```

Corresponding Weight Matrix



```

//
NA STE12_DIG1
PO      A          C          G          T
01      0.00      32.62      0.96      250.22
02      42.97      0.00      241.37      1.98
03      271.22      0.00      12.78      0.00
04      252.90      1.67      27.03      0.00
05      278.87      0.00      0.00      0.00
06      0.98      277.79      0.00      0.97
07      153.01      30.02      93.87      22.75
.
    
```

Other locations of STE12_DIG1 sites:

Upstream gene	Upstream region coordinates	Site coordinates	Probability	Strand
---	---	chrVII:878320..878314	0.984212	-1
---	---	chrII:408661..408655	0.972057	-1
---	---	chrIII:229021..229027	0.972057	1
---	---	chrXIV:110765..110759	0.970263	-1
---	---	chrX:585873..585867	0.969271	-1
---	---	chrIII:13116..13110	0.966688	-1
---	---	chrII:192199..192193	0.966341	-1
---	---	chrXII:561858..561864	0.966132	1
---	---	chrXI:114864..114870	0.961639	1
---	---	chrVI:82256..82262	0.958973	1
---	---	chrXII:1041610..1041604	0.953367	-1
---	---	chrX:445331..445337	0.948736	1
---	---	chrVII:437032..437038	0.94671	1
---	---	chrXII:790167..790161	0.946707	-1
---	---	chrXIII:411475..411481	0.944509	1
---	---	chrV:188204..188210	0.944509	1
---	---	chrIII:13188..13194	0.944509	1
---	---	chrVIII:273742..273748	0.942583	1
---	---	chrIV:543992..543986	0.941509	-1
---	---	chrXIV:276112..276118	0.939875	1
---	---	chrXIV:703355..703349	0.935385	-1
---	---	chrXVI:462194..462188	0.933203	-1
---	---	chrVIII:273703..273709	0.932334	1
---	---	chrX:126581..126587	0.931886	1
---	---	chrVII:516467..516473	0.930058	1
---	---	chrV:167341..167335	0.930058	-1
---	---	chrIV:976108..976102	0.92753	-1
---	---	chrVII:437228..437234	0.922055	1
---	---	chrXIII:664436..664442	0.916896	1
---	---	chrXVI:781892..781898	0.916798	1
---	---	chrXIII:181232..181226	0.914531	-1
---	---	chrIV:50735..50729	0.914278	-1
---	---	chrXIII:659599..659605	0.908352	1

Datafiles for download:

- PhyloGibbs predictions made using orthologs only
- PhyloGibbs predictions based on CHIP-on-chip data
- MotEvo predictions based on CHIP-on-chip data
- Weight matrices