计算系统生物学

王 勇

中国科学院数学与系统科学研究院

http://zhangroup.aporc.org

Chinese Academy of Sciences

Transcriptional Regulatory Network Inference

Yong Wang

http://zhangroup.aporc.org

http://zhangroup.aporc.org

Chinese Academy of Sciences

Outline

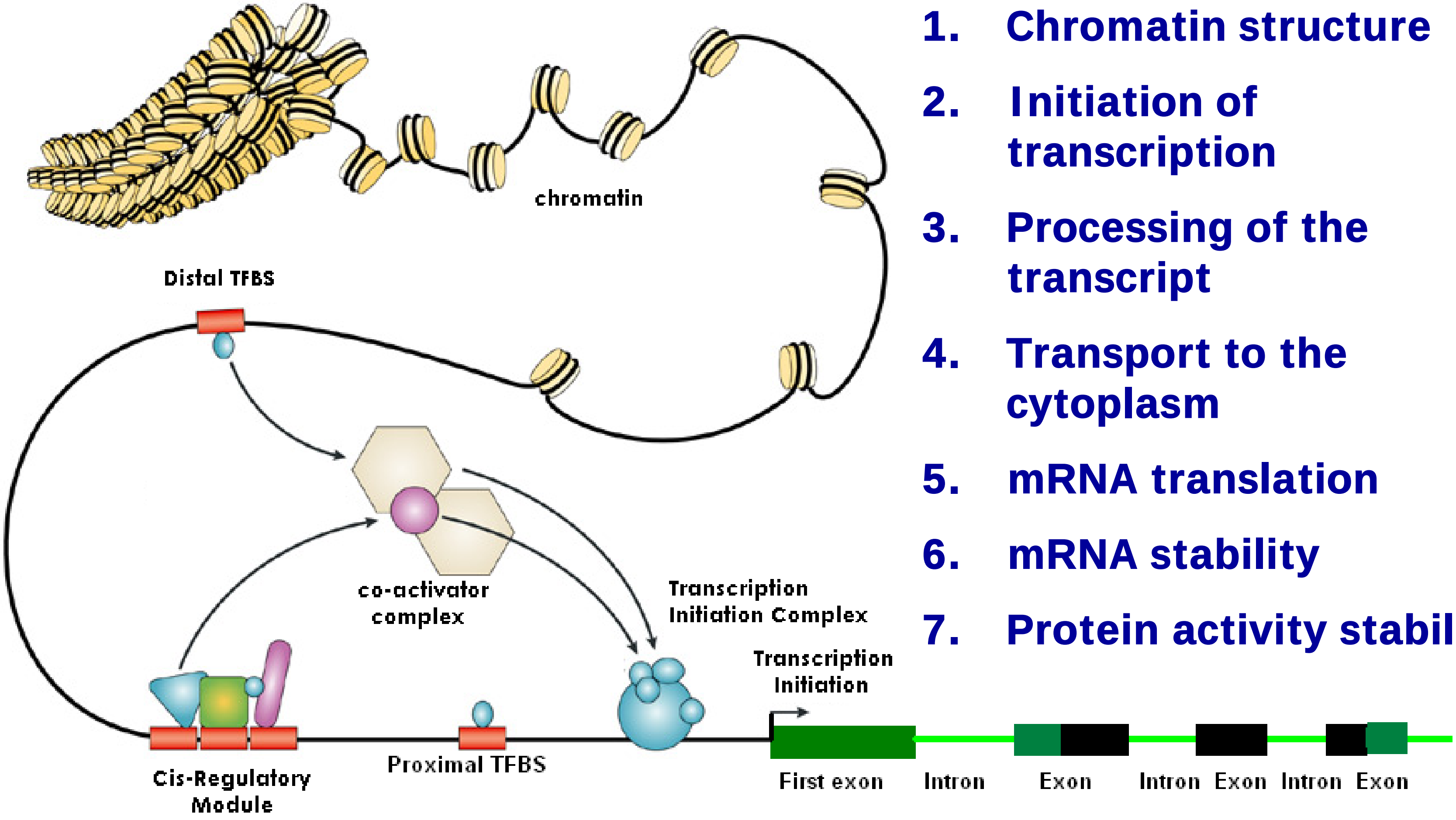
- Background: Definition of TRN inference)

- Inferring TRN from sequence’s perspective.

- Inferring TRN from gene expression’s perspective (Method: Inferelator)

- Inferring TRN from transcription complexes’ perspective (Method: TRNInfer)

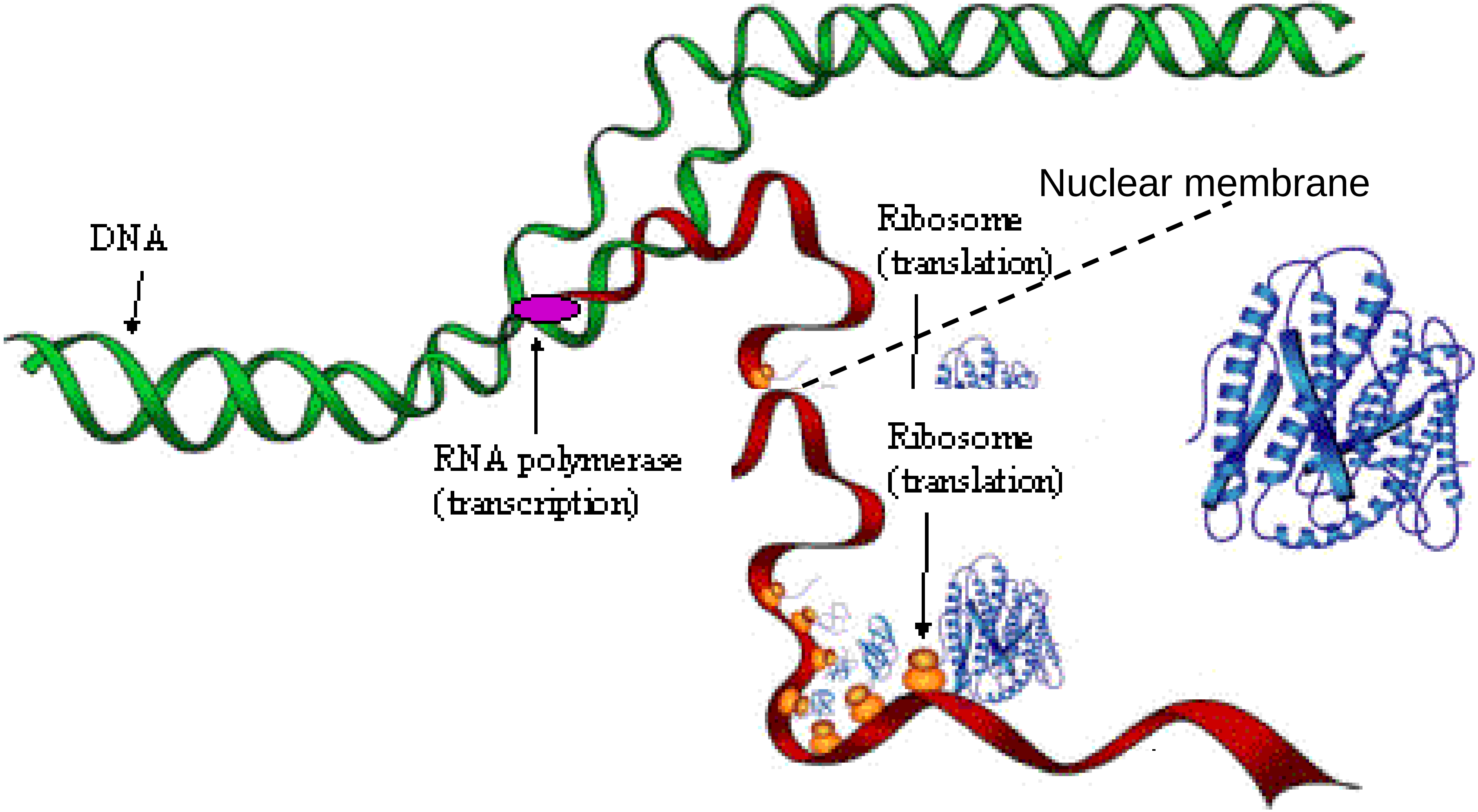
Transcription in higher eukaryotes



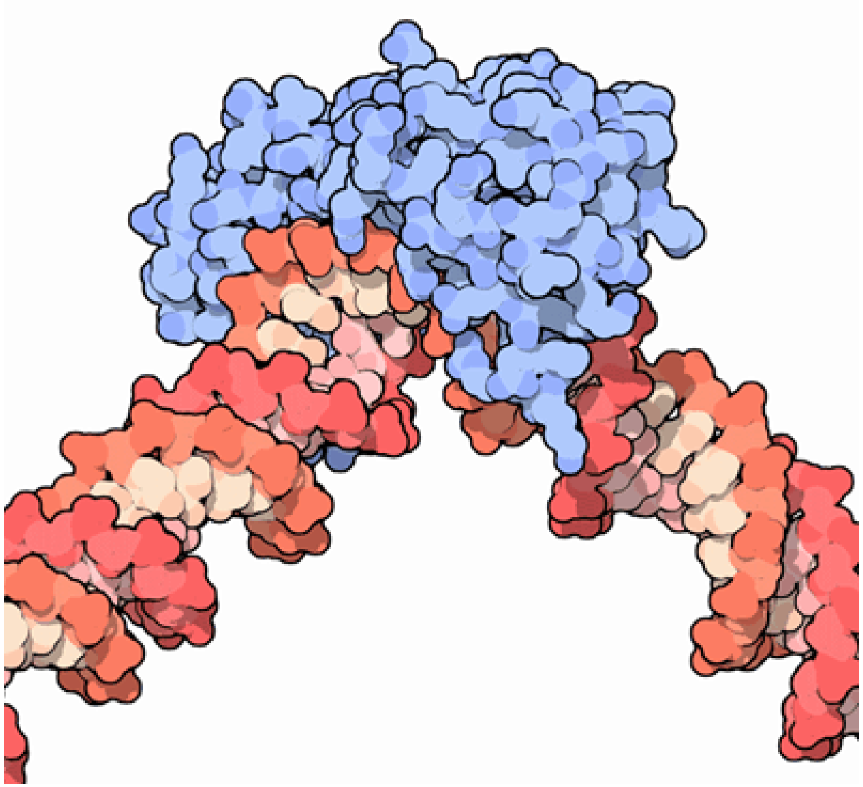
Gene Expression

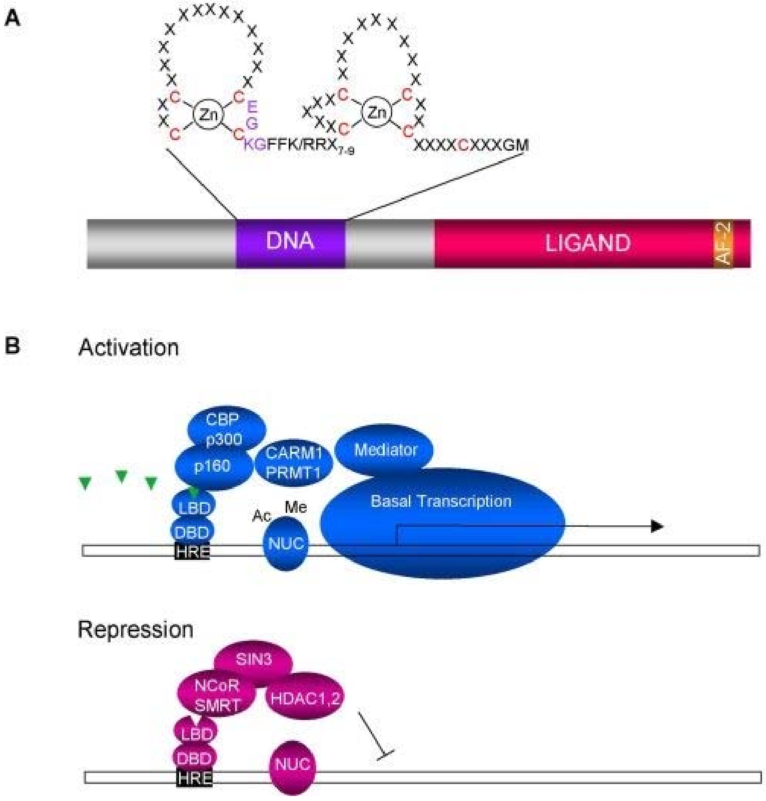
1. Chromatin structure
2. Initiation of transcription
3. Processing of the transcript
4. Transport to the cytoplasm
5. mRNA translation
6. mRNA stability
7. Protein activity stability

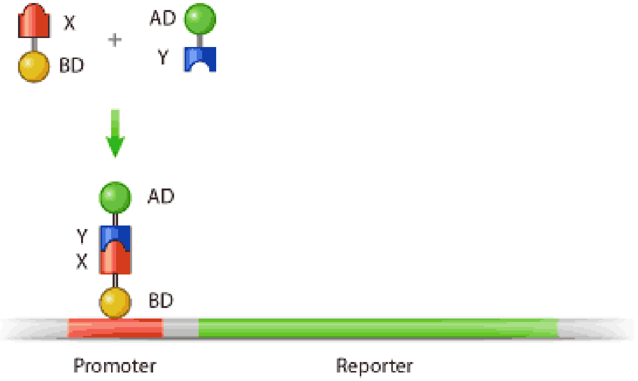
Transcriptional Regulation



Transcriptional Factor

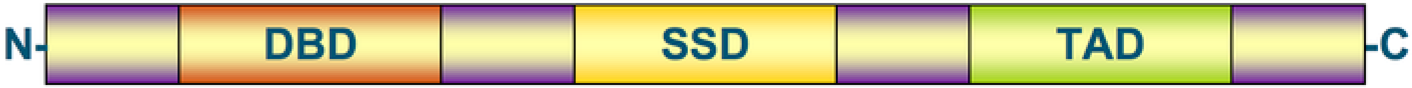






The transcription factor TATA binding protein (blue) bound to DNA (red). Image by David S. Goodsell based on the crystal structure 1cdw from the Protein Data Bank.

Structure



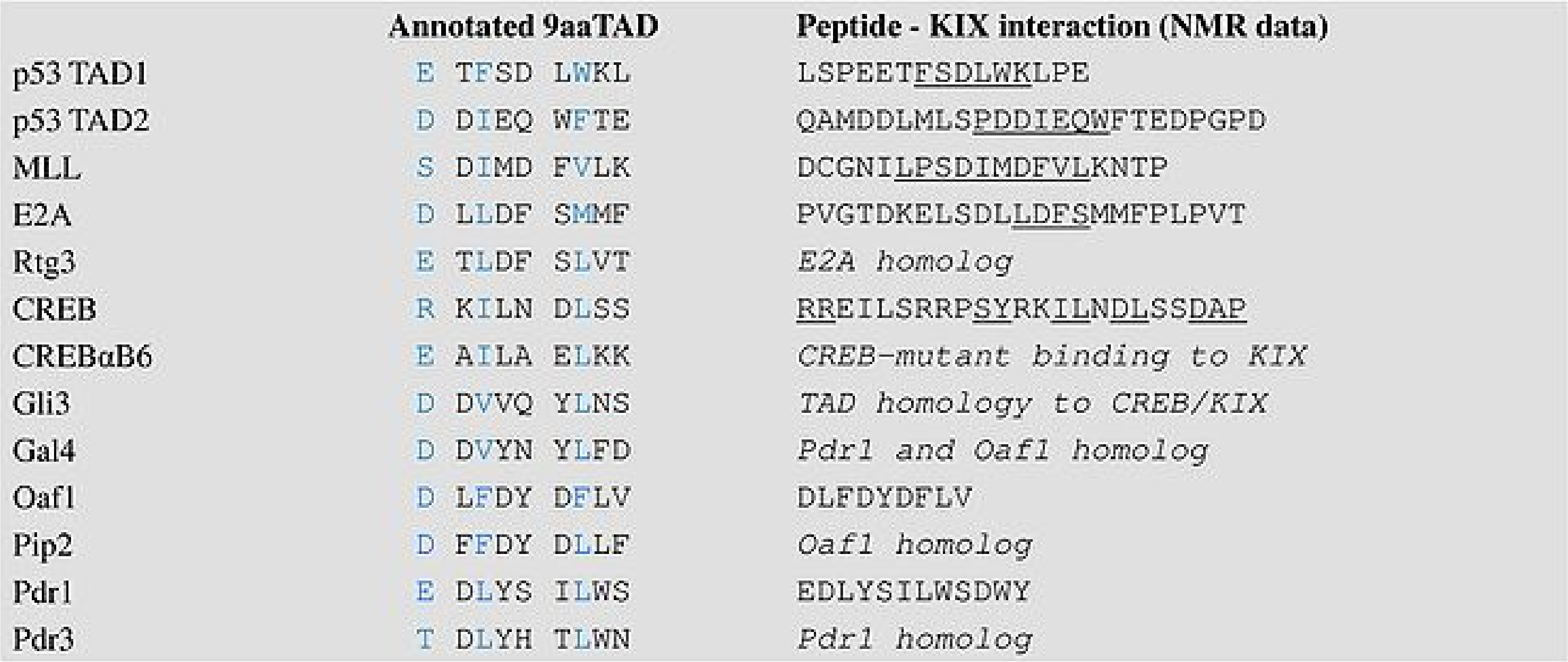
Schematic diagram of the amino acid sequence (amino terminus to the left and carboxylic acid terminus to the right) of a prototypical transcription factor that contains

(1) a DNA-binding domain (**DBD**), which attach to specific sequences of DNA (enhancer or promoter sequences) adjacent to regulated genes.

(2) signal sensing domain (**SSD**), which senses external signals and in response transmit these signals to the rest of the transcription complex, resulting in up or down regulation of gene expression. An optional **domain** (*e.g.*, a ligand binding domain).

(3) a transactivation domain (**TAD**), which contain binding sites for other proteins such as transcription coregulators. These binding sites are frequently referred to as **activation functions** (**AFs**).

Trans-activating domain



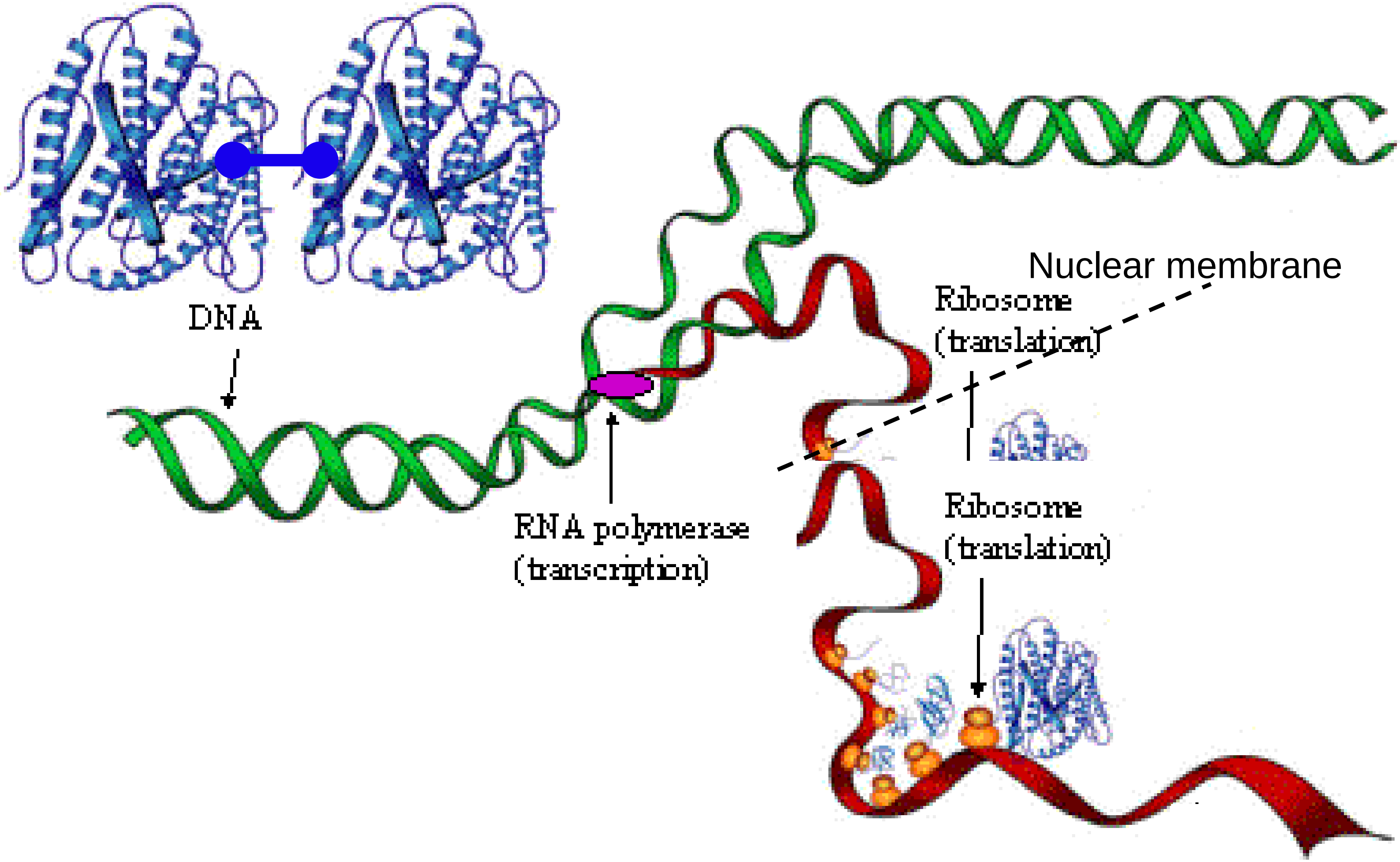
	Annotated 9aaTAD	Peptide - KIX interaction (NMR data)
p53 TAD1	E TFSD LWKL	LSPEETFSDLWKLPE
p53 TAD2	D DIEQ WFTE	QAMDDLMLSPDDIEQWFTEDPGPD
MLL	S DIMD FVLK	DCGNILPSDIMDFVLKNTP
E2A	D LLDF SMMF	PVGTDKELSDLLDFSMMFPLPVT
Rtg3	E TLDF SLVT	*E2A homolog*
CREB	R KILN DLSS	RREILSRRPSYRKILNDLSSDAP
CREBαB6	E AILA ELKK	*CREB-mutant binding to KIX*
Gli3	D DVVQ YLNS	*TAD homology to CREB/KIX*
Gal4	D DVYN YLFD	*Pdr1 and Oaf1 homolog*
Oaf1	D LFDY DFLV	DLFDYDFLV
Pip2	D FFDY DLLF	*Oaf1 homolog*
Pdr1	E DLYS ILWS	EDLYSILWSDWY
Pdr3	T DLYH TLWN	*Pdr1 homolog*

Nine-amino-acid transactivation domain (9aaTAD)

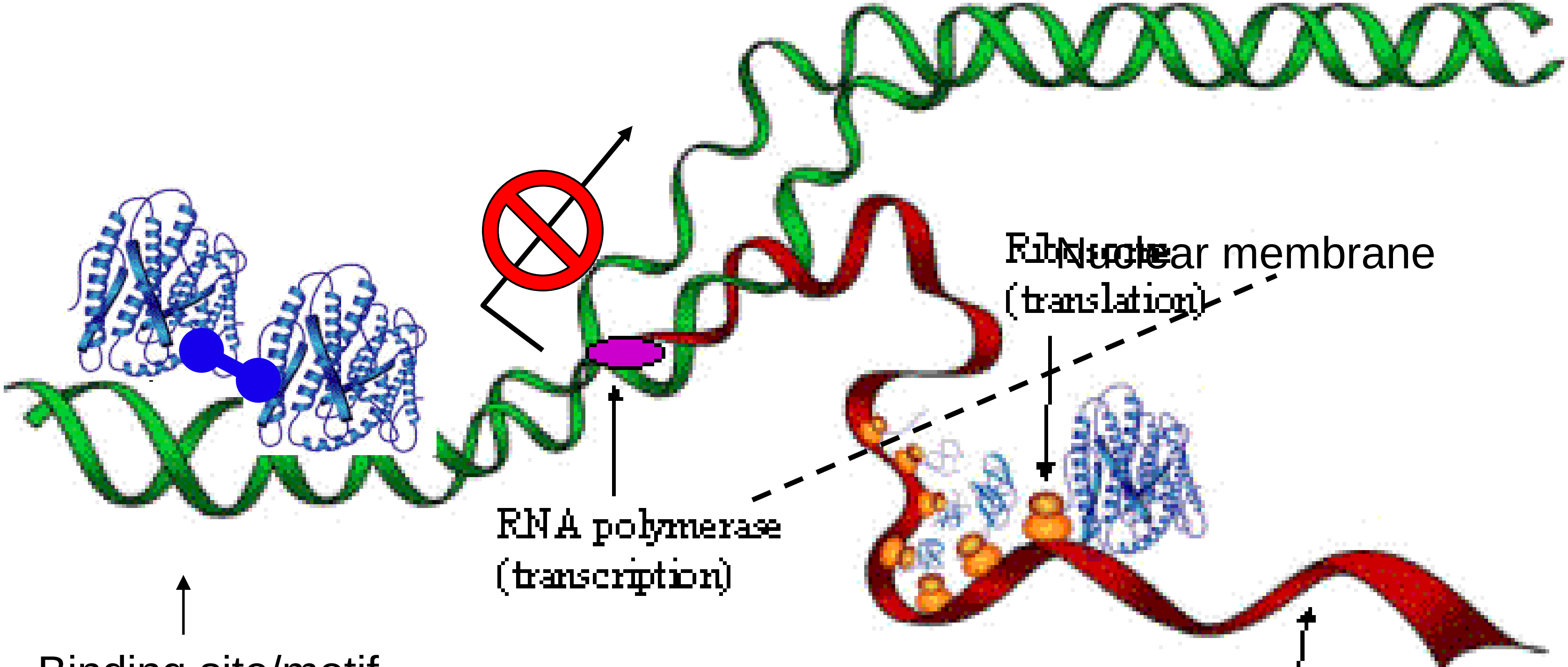
DNA-binding domain

Family	InterPro	Pfam	SCOP
basic-helix-loop-helix[43]	IPR001092	Pfam PF00010	SCOP 47460
basic-leucine zipper (bZIP)[44]	IPR004827	Pfam PF00170	SCOP 57959
C-terminal effector domain of the bipartite response regulators	IPR001789	Pfam PF00072	SCOP 46894
GCC box			SCOP 54175
helix-turn-helix[45]			
homeodomain proteins - bind to homeobox DNA sequences, which in turn encode other transcription factors. Homeodomain proteins play critical roles in the regulation of development.[46]	IPR009057	Pfam PF00046	SCOP 46689
lambda repressor-like	IPR010982		SCOP 47413
srf-like (serum response factor)	IPR002100	Pfam PF00319	SCOP 55455
paired box[47]			
winged helix	IPR013196	Pfam PF08279	SCOP 46785
zinc fingers[48]			
* multi-domain Cys_2His_2 zinc fingers[49]	IPR007087	Pfam PF00096	SCOP 57667
* Zn_2/Cys_6			SCOP 57701
* Zn_2/Cys_8 nuclear receptor zinc finger	IPR001628	Pfam PF00105	SCOP 57716

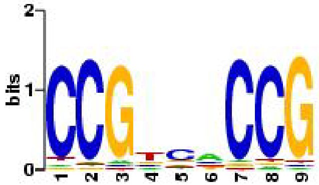
Transcriptional Regulation



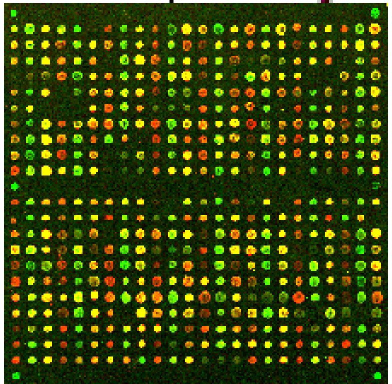
Transcriptional Regulation: output



Binding site/motif
CCG__CCG



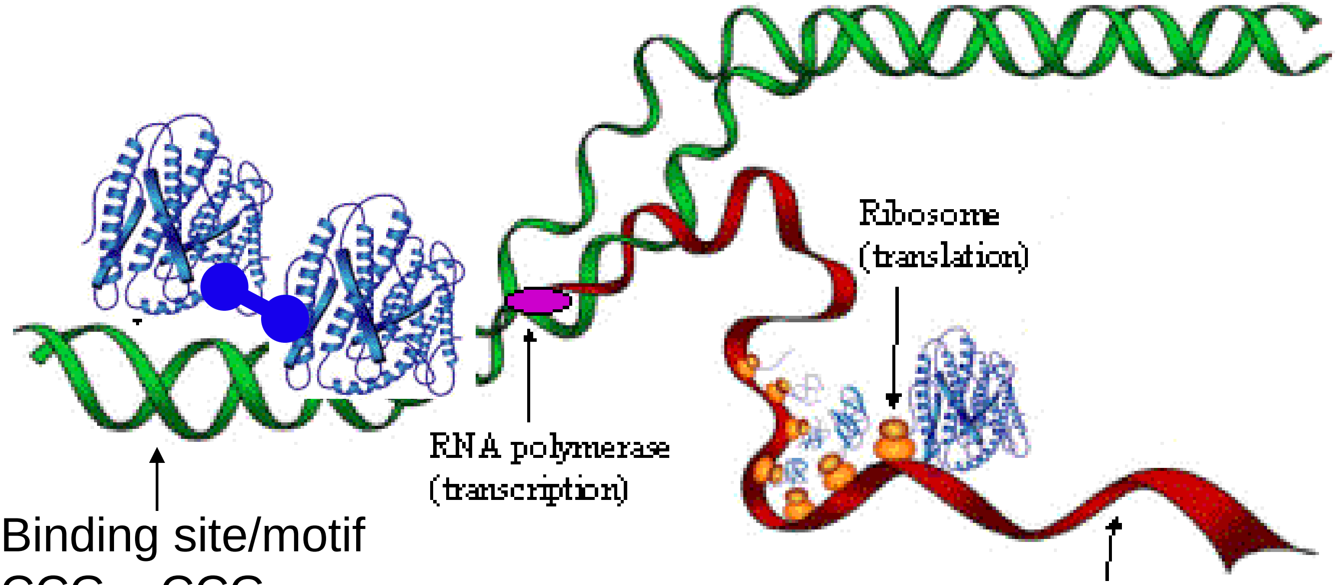
Genome-wide mRNA
transcript data (e.g.
microarrays)



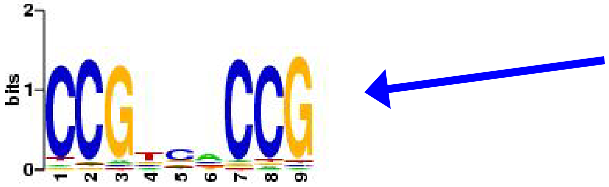
Perspective I： Cis-regulatory elements

Learning problems:

- Understand which regulators control which target genes



Binding site/motif
CCG__CCG

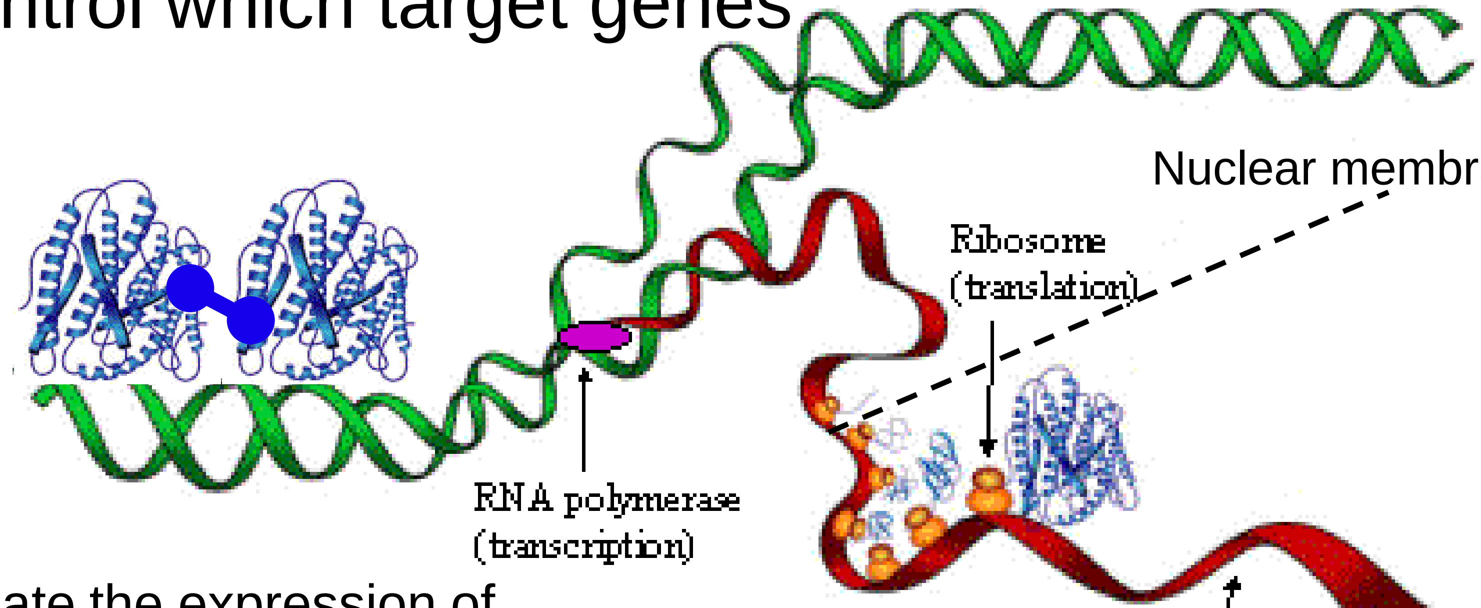


- Discover motifs representing regulatory elements
- Scan motif for potential regulatory interactions

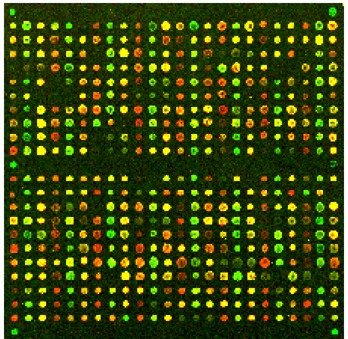
Perspective II： Target gene expression

Learning problems:

- Understand which regulators control which target genes



- Correlate the expression of transcription factor with the target gene
- Select the TF sets to explain the data

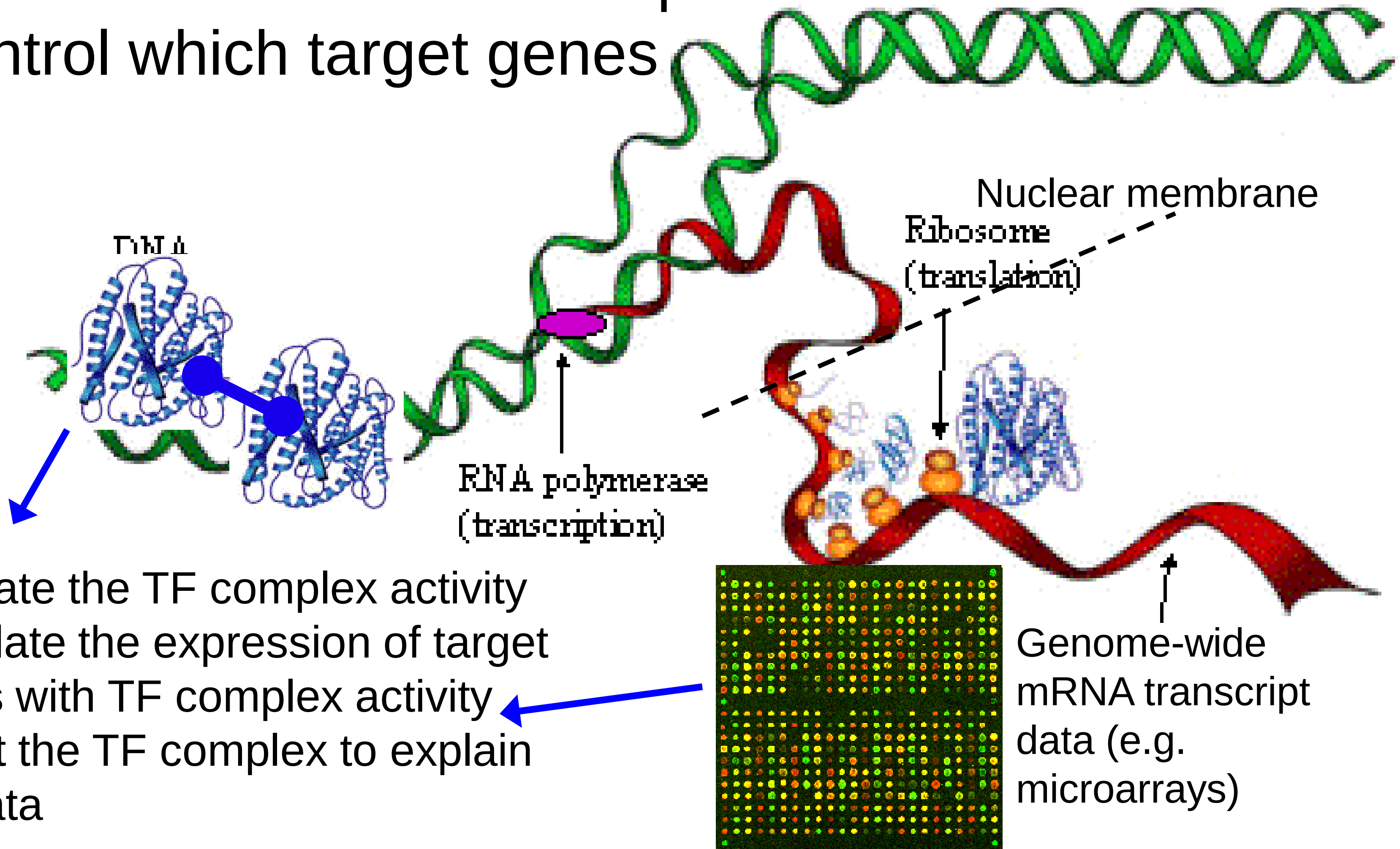


Genome-wide mRNA transcript data (e.g. microarrays)

Perspective III： Transcriptional complex

Learning problems:

- Understand which TF complex control which target genes



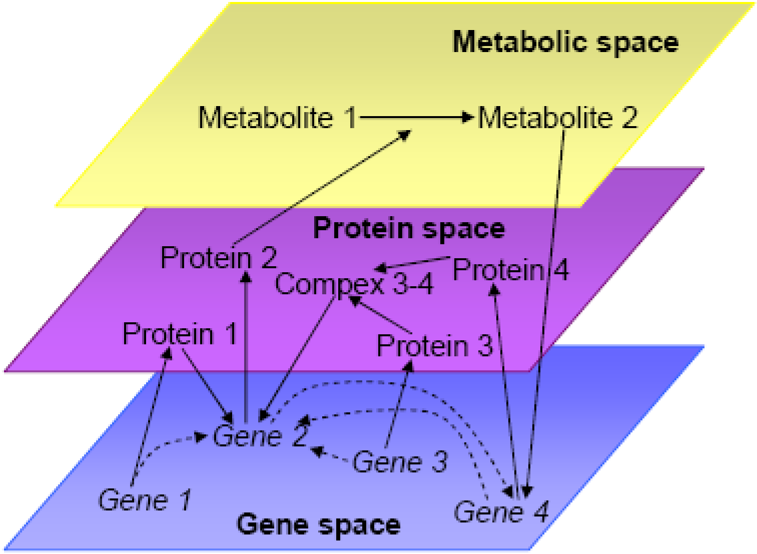
Nuclear membrane

Genome-wide mRNA transcript data (e.g. microarrays)

- Estimate the TF complex activity
- Correlate the expression of target genes with TF complex activity
- Select the TF complex to explain the data

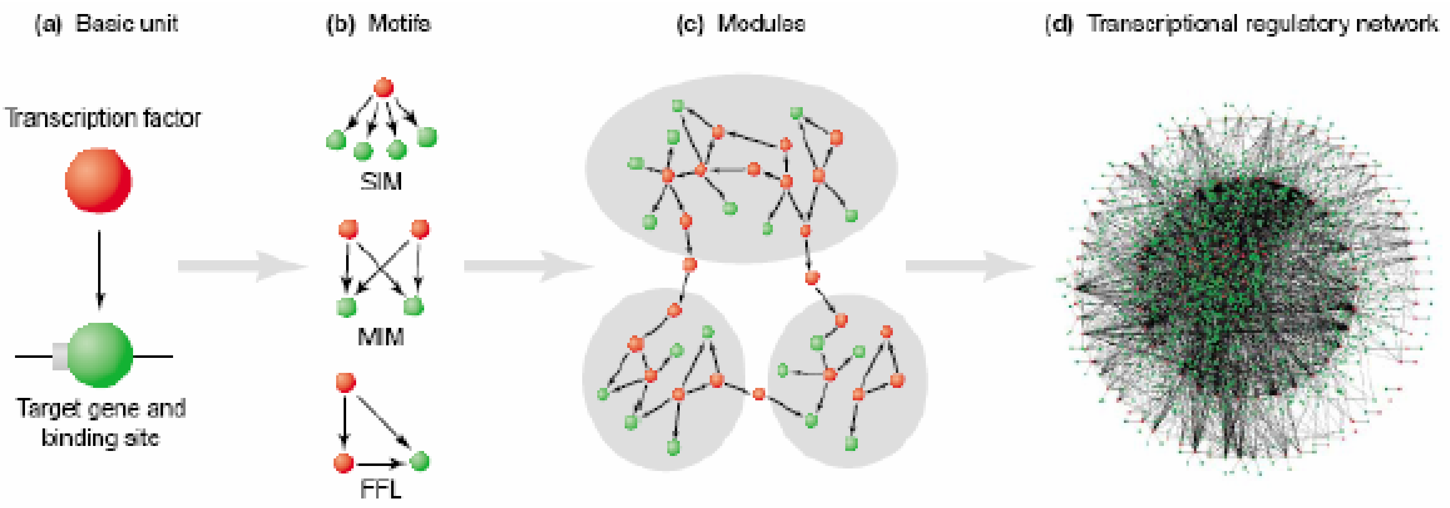
GRN and TRN ?

- Gene regulatory networks (GRN): indirect gene-gene interactions (genetic interactions)



GRN and TRN ?

- Transcription regulatory networks (TRN): direct interactions between TFs and genes (physical interactions)


(a) Basic unit
Transcription factor
Target gene and binding site
(b) Motifs
SIM
MIM
FFL
(c) Modules
(d) Transcriptional regulatory network

GRN and TRN ?

- **GRN**:

[redacted]

mRNA x(t) → mRNA x(t): indirect interactions

- **TRN**:

[redacted]

Protein a(t) → mRNA x(t): direct interaction

Outline

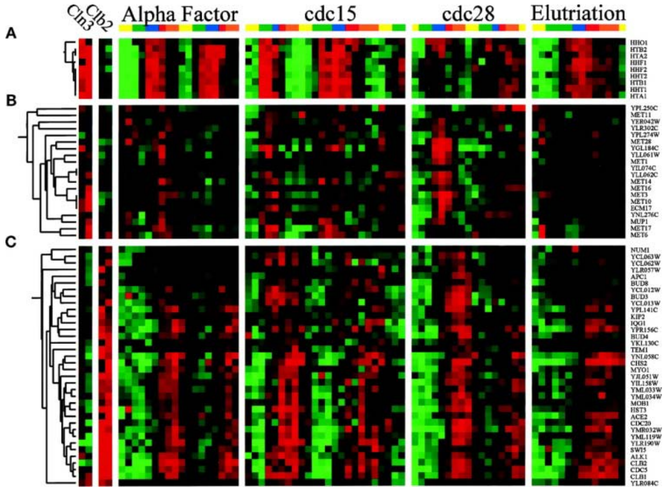
- Background: Definition of TRN inference)

- Inferring TRN from sequence’s perspective.

- Inferring TRN from gene expression’s perspective (Method: Inferelator)

- Inferring TRN from transcription complexes’ perspective (Method: TRNInfer)

- **TF binding sites discovery**
  - – Cluster genes by expression profile, annotation, … to find potentially coregulated genes
  - – Find *overrepresented* motifs in promoter sequences of *similar* genes (algorithms: MEME, Consensus, Gibbs sampler, AlignACE, …)



TFBS and PWM?

- Transcription factor binding sites (TFBSs) are usually slightly variable in their sequences.

- A positional weight matrix (PWM) specifies the probability that you will see a given base at each index position of the motif.

Pos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	18	8	5	4	1	29	7	7	7	0	1	39	1	1	6
C	8	3	3	9	33	4	21	15	14	0	0	1	43	39	18
G	13	31	34	9	8	10	11	15	19	4	44	3	0	1	6
T	7	4	4	24	4	3	7	9	6	42	1	3	2	5	16
Con	N	G	G	T	C	A	N	N	N	T	G	A	C	C	N

Calculation of PWM

1. acggcagggTGACCc
2. aGGGCAtcgTGACCc
3. cGGTCGccaGGACCt
4. tGGTCAggcTGGTCt
5. aGGTGGcccTGACCc
6. cTGTCCctcTGACCc
7. aGGCTAcgaTGACGt

⋮

41. cagggagtgTGACCc
42. gagcatgggTGACCa
43. aGGTCAtaacgattt
44. gGAACAgttTGACCc
45. cGGTGAcctTGACCc
46. gGGGCAaagTGACTg

Position frequency matrix (PFM)

(also known as *raw count matrix*)

Given N sequence fragments of fixed length, one can assemble a position frequency matrix (number of times a particular nucleotide appears at a given position). A normalized PFM, in which each column adds up to a total of one, is a matrix of probabilities for observing each nucleotide at each position.

Position weight matrix (PWM)

(also known as *position-specific scoring matrix*)

PFM should be converted to log-scale for efficient computational analysis. To eliminate null values before log-conversion, and to correct for small samples of binding sites, a sampling correction, known as *pseudocounts,* is added to each cell of the PFM.

Position Weight Matrix

Converting a PFM into a PWM

A	18	8	5	4	1	29	7	7	7	0	1	39	1	1	6
C	8	3	3	9	33	4	21	15	14	0	0	1	43	39	18
G	13	31	34	9	8	10	11	15	19	4	44	3	0	1	6
T	7	4	4	24	4	3	7	9	6	42	1	3	2	5	16

For each matrix element do:

$$w(b,i) = \log_2 \frac{p(b,i)}{p(b)} = \log_2 \frac{\dfrac{f_{b,i} + \dfrac{\sqrt{N}}{4}}{N + \sqrt{N}}}{p(b)}$$

A	0.58	-0.44	-0.98	-1.21	-2.29	1.22	-0.60	-0.60	-0.60	-2.96	-2.29	1.62	-2.29	-2.29	-0.72
C	-0.44	-1.49	-1.49	-0.30	1.39	-1.21	0.78	0.34	0.25	-2.96	-2.96	-2.29	1.76	1.62	0.46
G	0.16	1.31	1.44	-0.30	-0.44	-0.17	-0.06	0.34	0.65	-1.21	1.79	-1.49	-2.96	-2.29	-0.64
T	-0.60	-1.21	-1.21	0.96	-1.21	-1.49	-0.60	-0.30	-0.78	1.73	-2.29	-1.49	-1.84	-0.98	0.23

$f_{b,i}$ – raw count (PFM matrix element) of nucleotide ***b*** in column ***i***

N – number of sequences used to create PFM (= column sum)

$\frac{\sqrt{N}}{4}$ and $\sqrt{N}$ - pseudocounts (correction for small sample size)

p(b) - background frequency of nucleotide ***b, this one usually defaults to 0.25***

Hertz GZ, Stormo GD. Bioinformatics (*1999*)

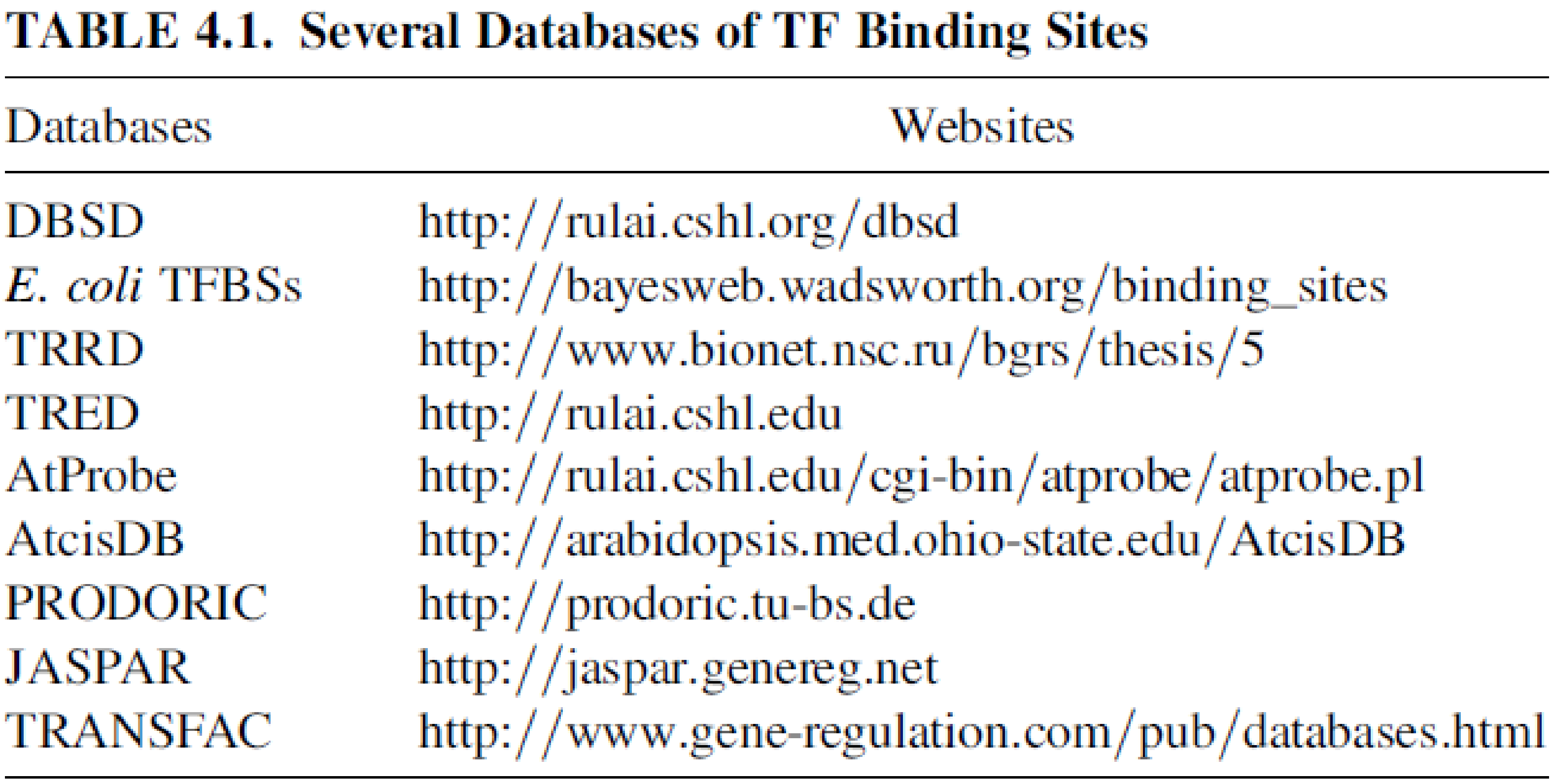


TABLE 4.1. Several Databases of TF Binding Sites

Databases	Websites
DBSD	http://rulai.cshl.org/dbsd
E. coli TFBSs	http://bayesweb.wadsworth.org/binding_sites
TRRD	http://www.bionet.nsc.ru/bgrs/thesis/5
TRED	http://rulai.cshl.edu
AtProbe	http://rulai.cshl.edu/cgi-bin/atprobe/atprobe.pl
AtcisDB	http://arabidopsis.med.ohio-state.edu/AtcisDB
PRODORIC	http://prodoric.tu-bs.de
JASPAR	http://jaspar.genereg.net
TRANSFAC	http://www.gene-regulation.com/pub/databases.html

Scoring putative transcriptional regulation by scanning the promoter with PWM

GGGTCAGCATGGCCA

A	0.58	-0.44	-0.98	-1.21	-2.29	1.22	-0.60	-0.60	-0.60	-2.96	-2.29	1.62	-2.29	-2.29	-0.72
C	-0.44	-1.49	-1.49	-0.30	1.39	-1.21	0.78	0.34	0.25	-2.96	-2.96	-2.29	1.76	1.62	0.46
G	0.16	1.31	1.44	-0.30	-0.44	-0.17	-0.06	0.34	0.65	-1.21	1.79	-1.49	-2.96	-2.29	-0.64
T	-0.60	-1.21	-1.21	0.96	-1.21	-1.49	-0.60	-0.30	-0.78	1.73	-2.29	-1.49	-1.84	-0.98	0.23

Absolute score of the site $S = \sum_{i=1}^{m} w(b,i)$ **=11.57**

															Row Sum
Max	0.58	1.31	1.44	0.96	1.39	1.22	0.78	0.34	0.65	1.73	1.79	1.62	1.76	1.62	17.20
Min	-0.60	-1.49	-1.49	-1.21	-2.29	-1.49	-0.60	-0.60	-0.78	-2.96	-2.96	-2.29	-2.96	-2.29	-24.02

$$relative\_score = \frac{Absolute\_score - Minimum\_score}{Maximum\_score - Minimum\_score}$$

$$= \frac{11.57 - (-24.02)}{17.20 - (-24.02)} = 0.86$$

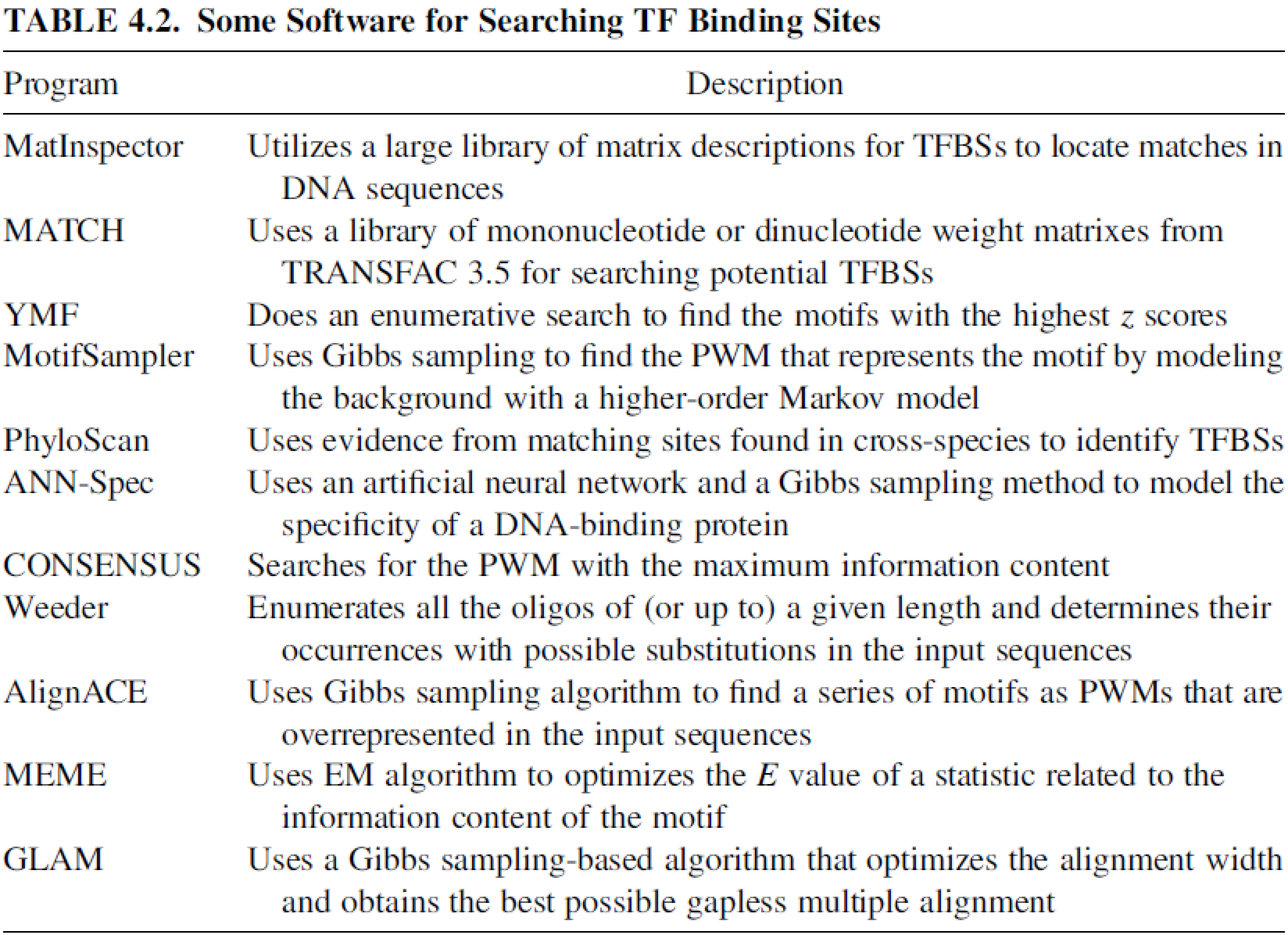


TABLE 4.2. Some Software for Searching TF Binding Sites

Program	Description
MatInspector	Utilizes a large library of matrix descriptions for TFBSs to locate matches in DNA sequences
MATCH	Uses a library of mononucleotide or dinucleotide weight matrixes from TRANSFAC 3.5 for searching potential TFBSs
YMF	Does an enumerative search to find the motifs with the highest *z* scores
MotifSampler	Uses Gibbs sampling to find the PWM that represents the motif by modeling the background with a higher-order Markov model
PhyloScan	Uses evidence from matching sites found in cross-species to identify TFBSs
ANN-Spec	Uses an artificial neural network and a Gibbs sampling method to model the specificity of a DNA-binding protein
CONSENSUS	Searches for the PWM with the maximum information content
Weeder	Enumerates all the oligos of (or up to) a given length and determines their occurrences with possible substitutions in the input sequences
AlignACE	Uses Gibbs sampling algorithm to find a series of motifs as PWMs that are overrepresented in the input sequences
MEME	Uses EM algorithm to optimizes the *E* value of a statistic related to the information content of the motif
GLAM	Uses a Gibbs sampling-based algorithm that optimizes the alignment width and obtains the best possible gapless multiple alignment

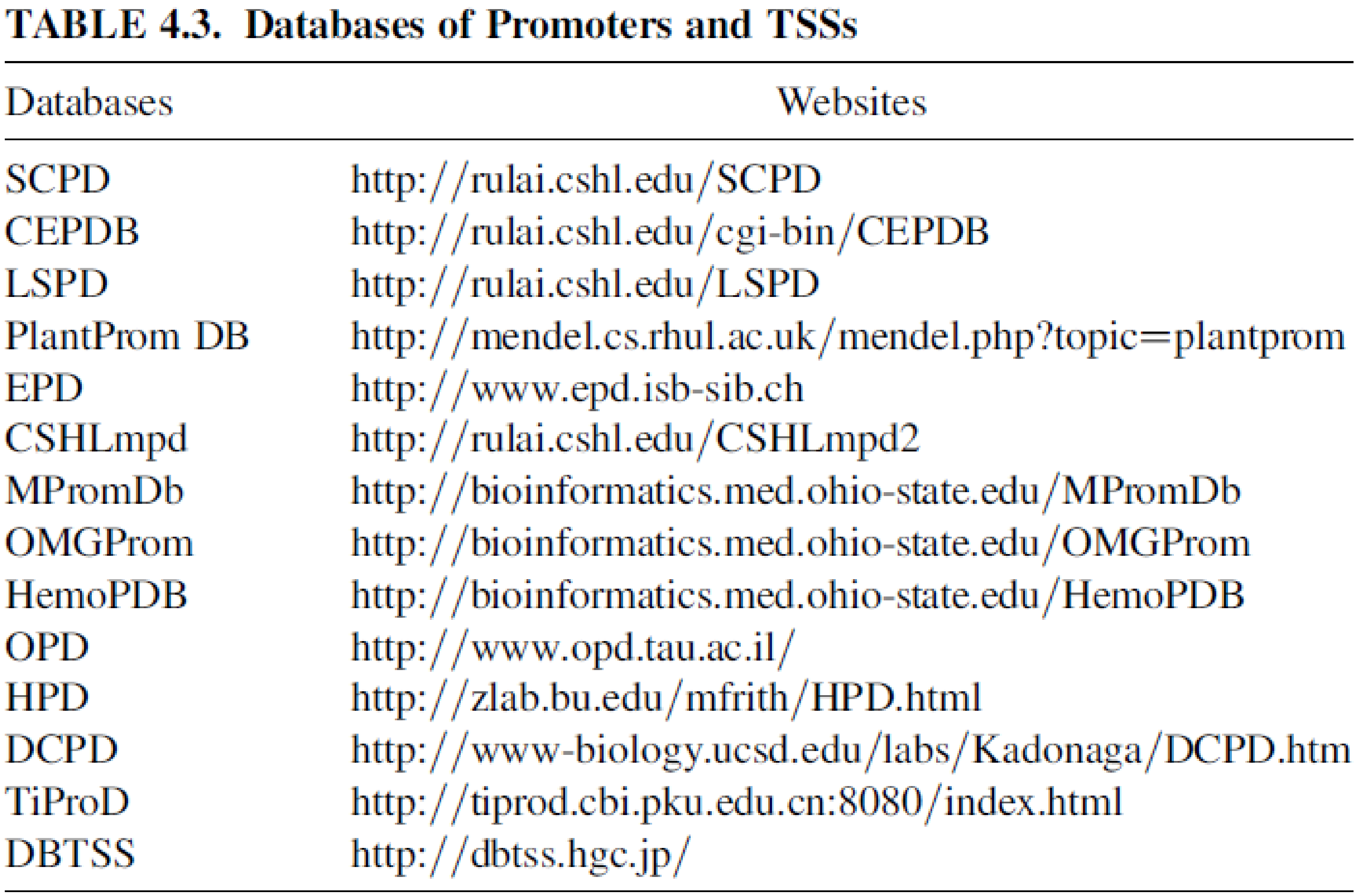


TABLE 4.3. Databases of Promoters and TSSs

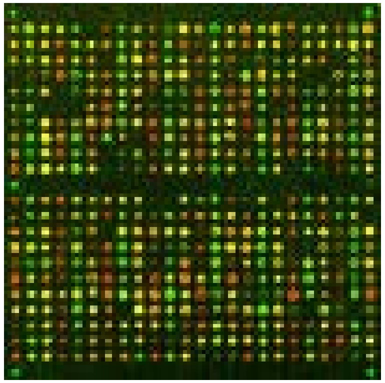
Databases	Websites
SCPD	http://rulai.cshl.edu/SCPD
CEPDB	http://rulai.cshl.edu/cgi-bin/CEPDB
LSPD	http://rulai.cshl.edu/LSPD
PlantProm DB	http://mendel.cs.rhul.ac.uk/mendel.php?topic=plantprom
EPD	http://www.epd.isb-sib.ch
CSHLmpd	http://rulai.cshl.edu/CSHLmpd2
MPromDb	http://bioinformatics.med.ohio-state.edu/MPromDb
OMGProm	http://bioinformatics.med.ohio-state.edu/OMGProm
HemoPDB	http://bioinformatics.med.ohio-state.edu/HemoPDB
OPD	http://www.opd.tau.ac.il/
HPD	http://zlab.bu.edu/mfrith/HPD.html
DCPD	http://www-biology.ucsd.edu/labs/Kadonaga/DCPD.htm
TiProD	http://tiprod.cbi.pku.edu.cn:8080/index.html
DBTSS	http://dbtss.hgc.jp/

Outline

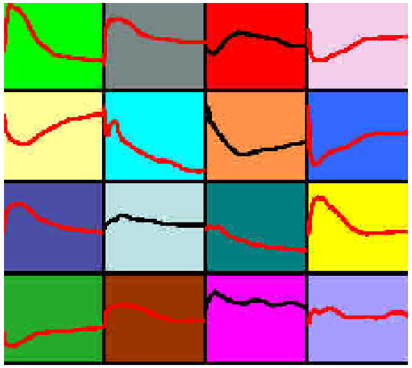
- Background: Definition of TRN inference)
- Inferring TRN from sequence’s perspective.
- Inferring TRN from gene expression’s perspective (Method: Inferelator)
- Inferring TRN from transcription complexes’ perspective (Method: TRNInfer)

Inferring transcriptional networks

From microarray data alone

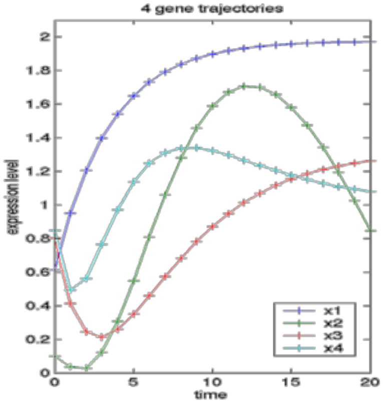


Gene expression data **X**

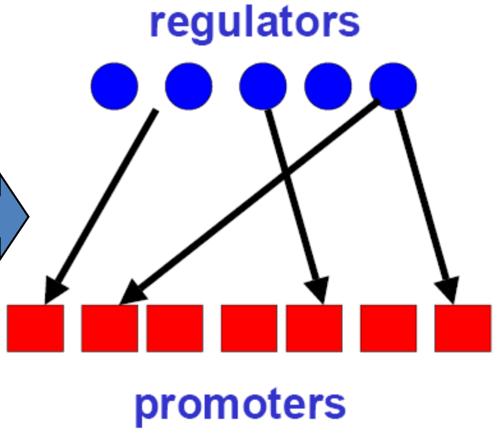


TF expression

Correlation



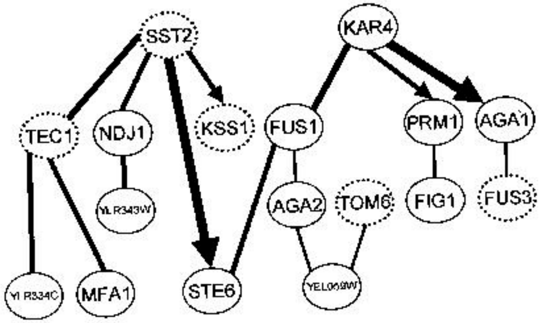
Target gene expression



TRN J

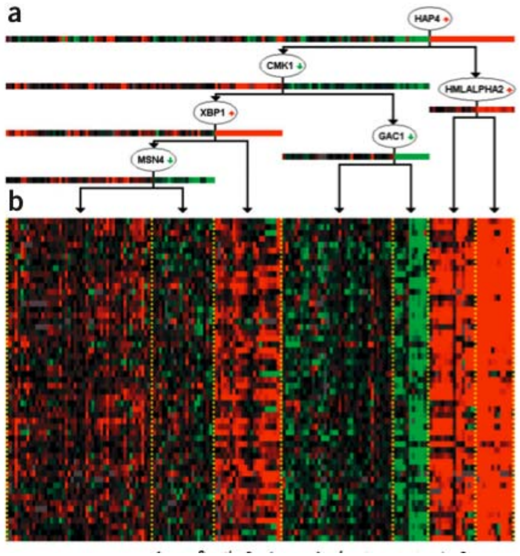
Structure learning

- – Learn structure of “regulatory network”, “regulatory modules”, etc.
- – Fit interpretable model to training data
- – Many *computational* and *statistical* challenges; often used for qualitative hypotheses rather than prediction



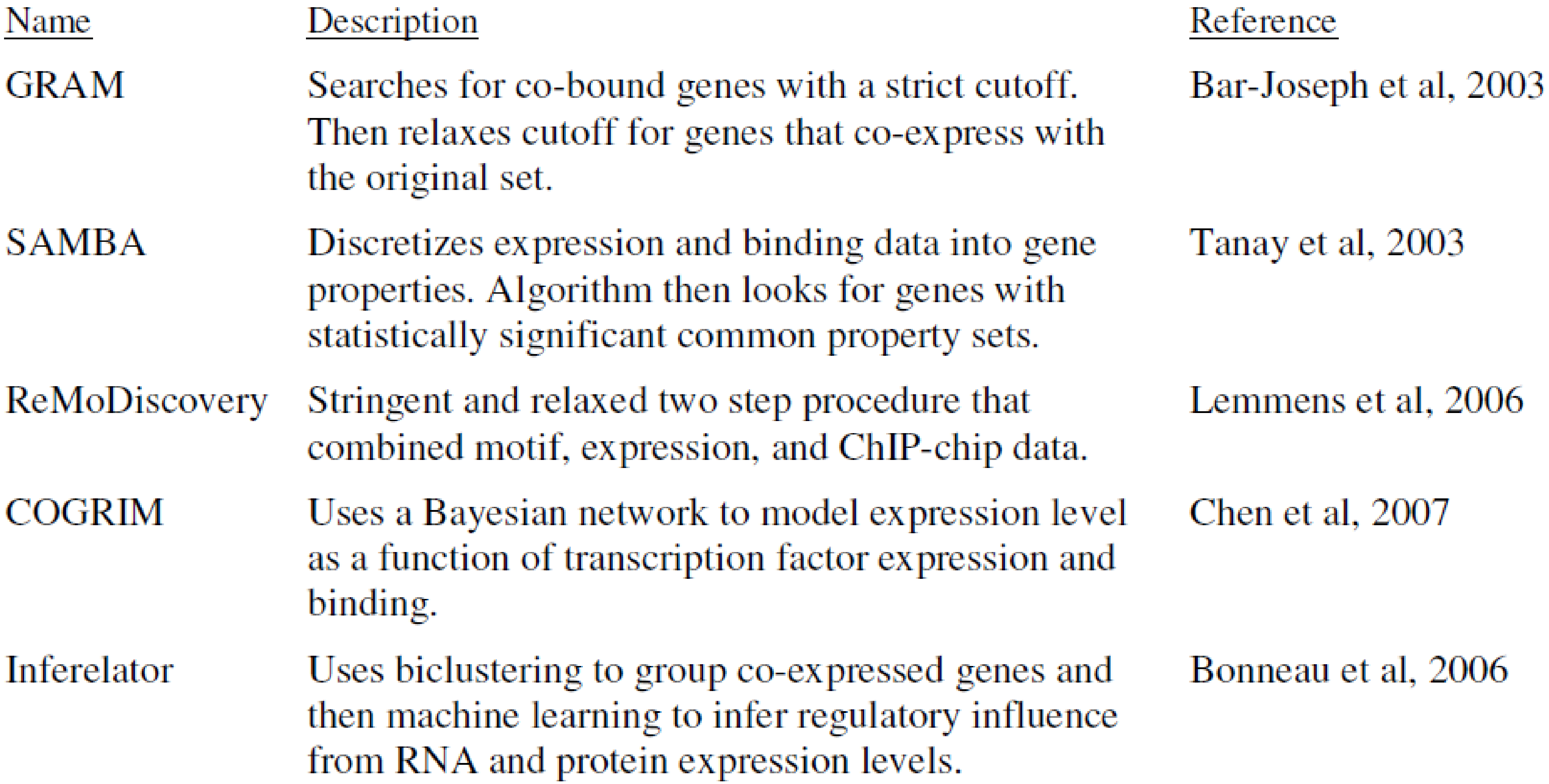
(Pe’er et al. 2001)

Interpretable



(Segal et al, 2003, 2004)

A list of relevant computational methods



Name	Description	Reference
GRAM	Searches for co-bound genes with a strict cutoff. Then relaxes cutoff for genes that co-express with the original set.	Bar-Joseph et al, 2003
SAMBA	Discretizes expression and binding data into gene properties. Algorithm then looks for genes with statistically significant common property sets.	Tanay et al, 2003
ReMoDiscovery	Stringent and relaxed two step procedure that combined motif, expression, and ChIP-chip data.	Lemmens et al, 2006
COGRIM	Uses a Bayesian network to model expression level as a function of transcription factor expression and binding.	Chen et al, 2007
Inferelator	Uses biclustering to group co-expressed genes and then machine learning to infer regulatory influence from RNA and protein expression levels.	Bonneau et al, 2006

Differential Equation Models

- Attempt to reconstruct the dynamical system that produced the gene expression data
  - Reduce dimensionality of the data
  - Approximate dynamics
    - Modeled using ordinary differential equations
  - Restrict model complexity
- Example system : The Inferelator

Dimensionality Reduction

- Regulators (genes and environment)
  - – Limited to transcription factors
  - – Factors with correlated profiles are merged
- Genes
  - – Clustered based on putative coregulation
  - – Used cMonkey to form biclusters across genes and conditions [Bonneau, 2006]
    - Correlated expression
    - Shared regulatory sequence motifs

(Bonneau, et al, Genome Biology, 2006)

Model Details

- Expression of *y (gene or bicluster mean)* is influenced by the expression of N regulators:

X = (x1, x2, …, xN)

$$\tau \frac{dy}{dt} = -y + g(\beta \bullet Z)$$

$$Z = (z_1[X], z_2[X] \ldots z_P[X])$$

(Bonneau, et al, Genome Biology, 2006)

Model Details

$$\tau \frac{dy}{dt} = -y + g(\beta \bullet Z)$$

$$Z = (z_1[X], z_2[X] \ldots z_P[X])$$

Choice of Squashing Function

压缩函数（*Squashing Function*）

$$g(\beta \bullet Z) = \frac{1}{1 + e^{-\beta \bullet Z}}$$

$$g(\beta Z) = \begin{cases} \beta Z: & \text{if } \min(\mathrm{y}) < \beta Z < \max(\mathrm{y}) \\ \max(y): & \text{if } \beta Z > \max(\mathrm{y}) \\ \min(y): & \text{if} \beta Z < \min(\mathrm{y}) \end{cases}$$

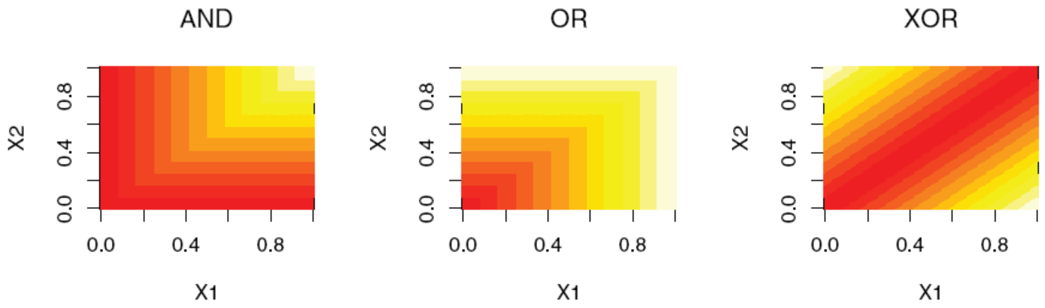
(Bonneau, et al, Genome Biology, 2006)

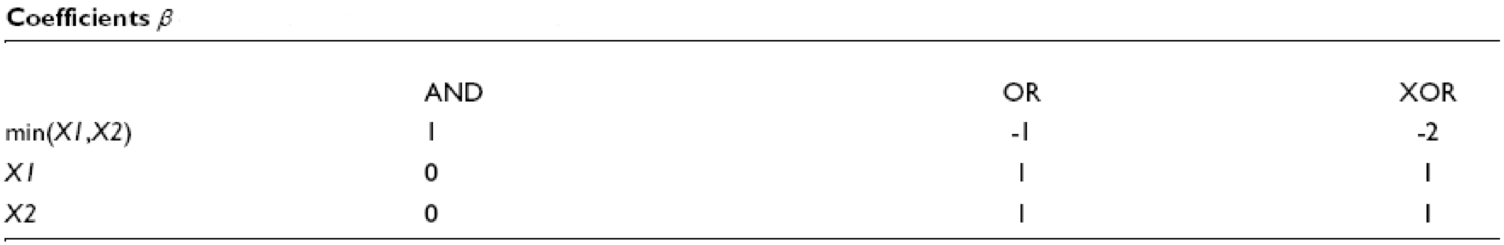
Model Details

$$\tau \frac{dy}{dt} = -y + g(\beta \cdot Z)$$

$$Z = (z_1[X], z_2[X] \dots z_P[X])$$

Choice of Z: $\beta Z = \beta_1 x_1 + \beta_2 x_2 + \beta_3 \min(x_1, x_2)$





Coefficients β

	AND	OR	XOR
min(X1,X2)	1	-1	-2
X1	0	1	1
X2	0	1	1

(Bonneau, et al, Genome Biology, 2006)

Model Details

$$\tau \frac{dy}{dt} = -y + g(\beta \bullet Z)$$

Steady state

$$y = g(\beta \bullet Z_{SS})$$

Time course

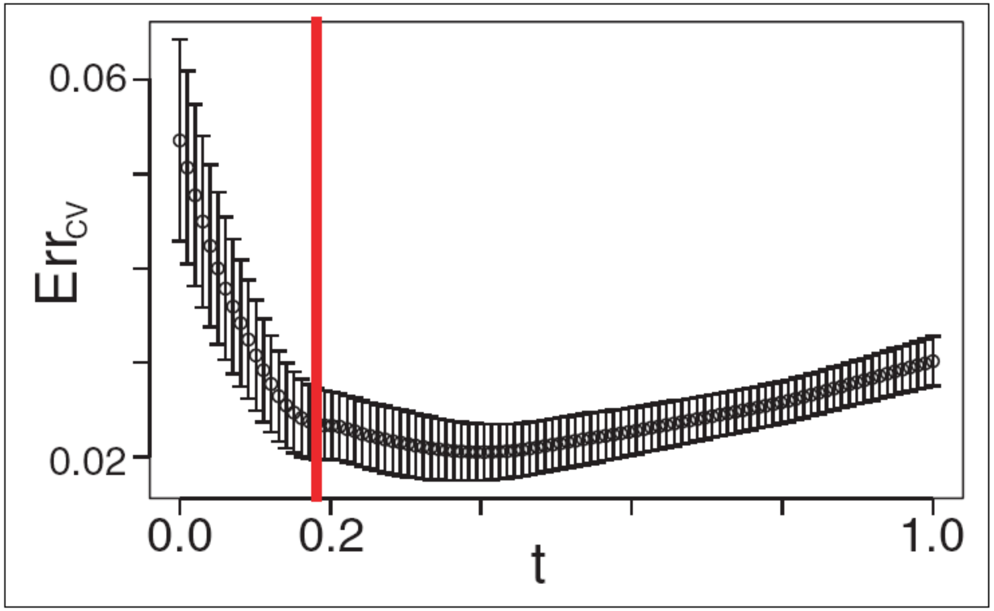
$$\tau \frac{y_{m+1} - y_m}{\Delta t_m} + y_m = g\left(\sum_{j=1}^{P} \beta_j z_{mj}\right) \quad for \quad m = 1, 2, \ldots, T-1$$

(Bonneau, et al, Genome Biology, 2006)

Model Learning with LASSO

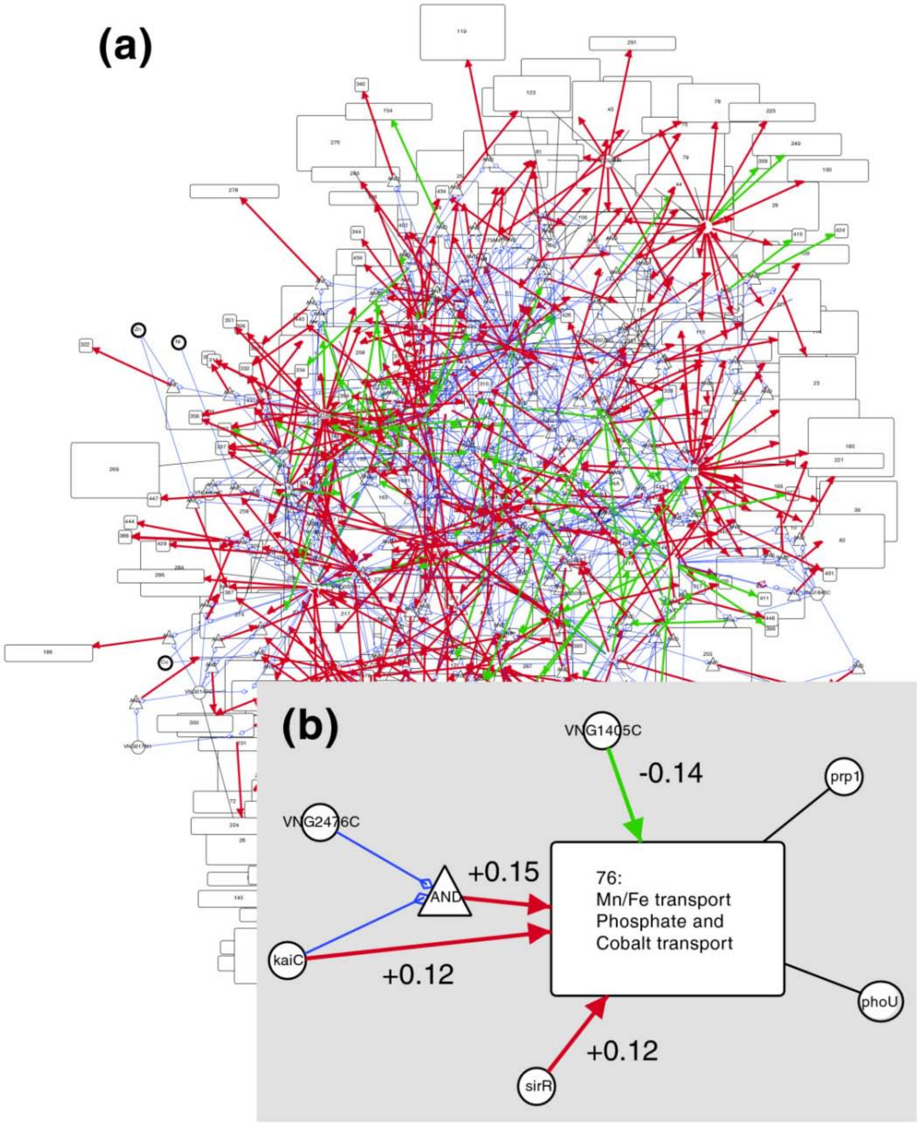
- LASSO, a.k.a. L1 shrinkage

$$\left(\hat{\alpha}, \hat{\beta}\right) = \underset{\alpha,\beta}{\arg\min}\left\{\sum_{i=1}^{N}\left(y_i - \alpha - \sum_{j=1}^{p}\beta_j z_{ij}\right)^2\right\} \quad \text{S.T.} \quad \sum_{j=1}^{p}\left|\beta_j\right| \le t\left|\beta_{ols}\right|$$



(Bonneau, et al, Genome Biology, 2006)

Results



The inferred regulatory network of *Halobacterium NRC-1*

Regulators are indicated as circles

Target gene biclusters are indicated by rectangles

Outline

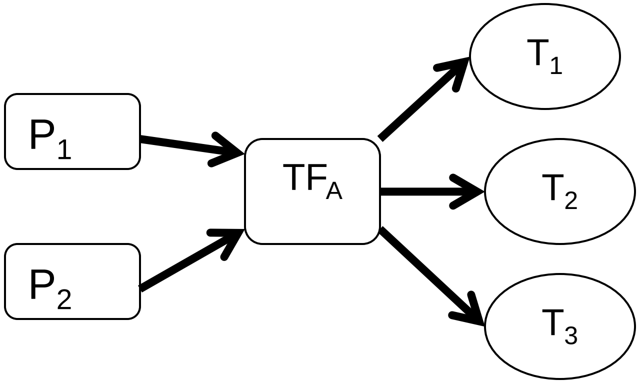
- Background: Definition of TRN inference)
- Inferring TRN from sequence’s perspective.
- Inferring TRN from gene expression’s perspective (Method: Inferelator)
- Inferring TRN from transcription complexes’ perspective (Method: TRNInfer)

Motivation

- TF activity level cannot be measured directly by microarray due to post-translational modifications

- Most existing algorithms has an implicit assumption that TFAs are proportional to their mRNA levels (like the previous example)

- TF generally regulates a gene with many collaborators **(Transcription complex)**



Protein

Target Gene

P_1

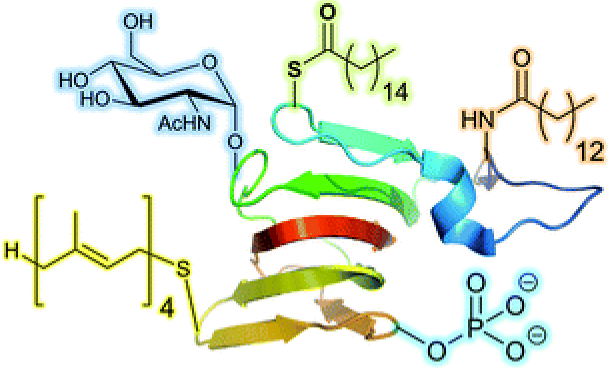
P_2

TF_A

T_1

T_2

T_3



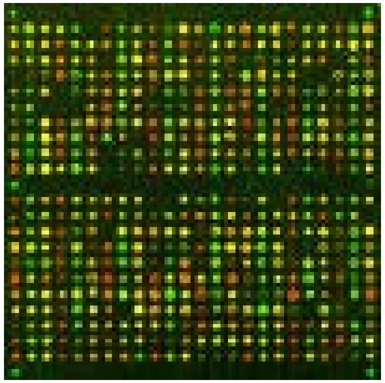
TF Activity

- Use TF-TG relation benefit the regulatory network identification

- TF expression level is not a good measure of the TF activity. The activated protein level of a TF, rather than its expression level, is what controls gene expression.

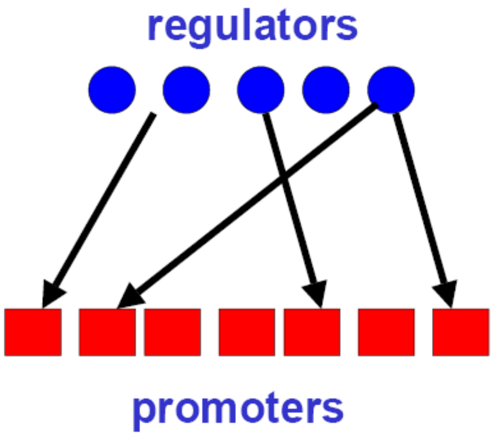
- The activity of a transcription factor is regulated according to the cell’s need, largely through signal transduction. It may not be directly observed, but can be reflected by the genes it regulates.

Inferring transcriptional networks



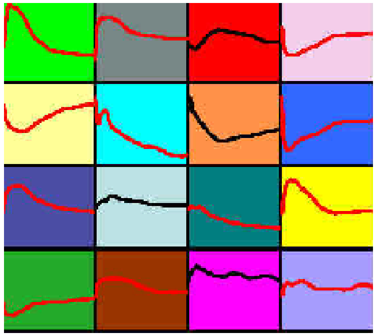
Gene expression data $\mathbf{X}$

⟺



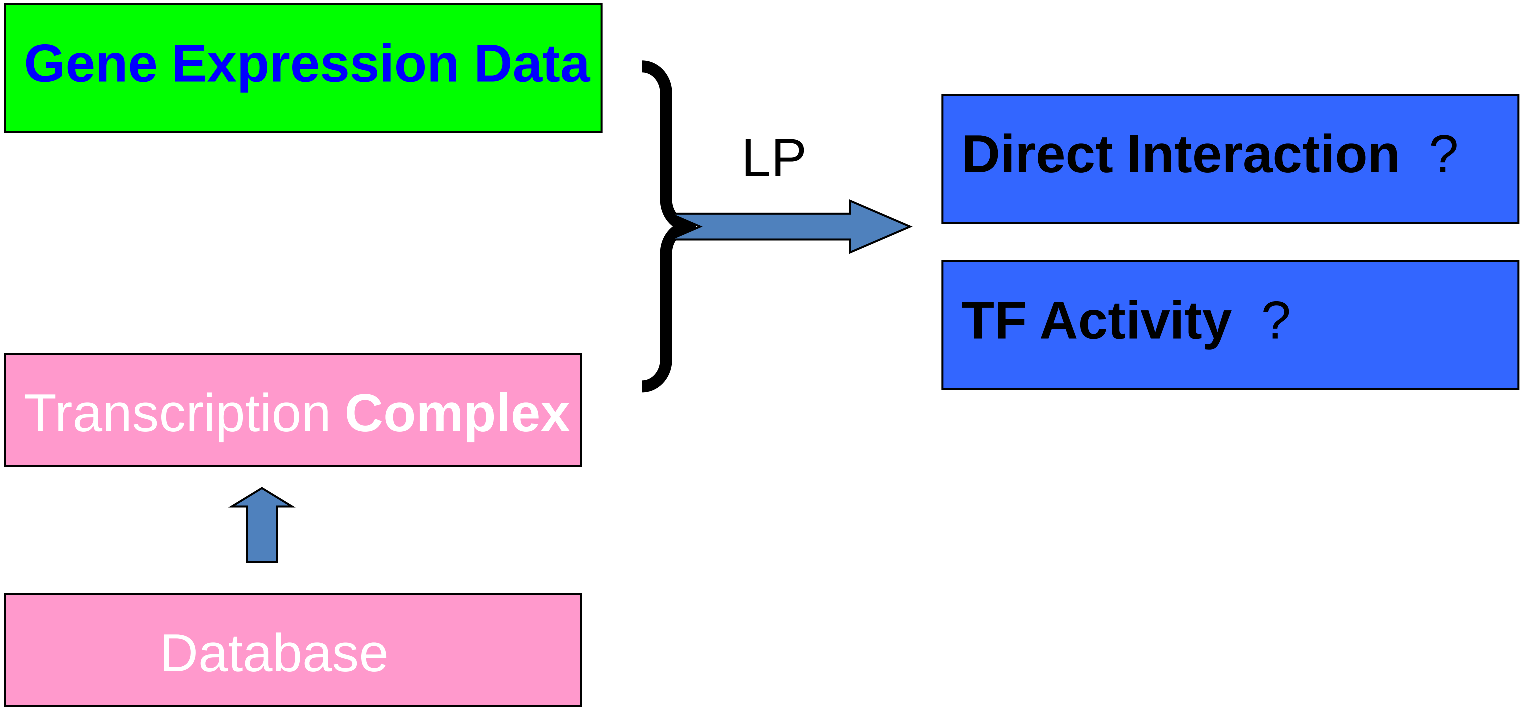
TRN $\mathbf{J}$

+



TF activity level $\mathbf{A}$

Framework for TRNinfer



Wang et al. Bioinformatics, 2007

Transcription Regulatory network

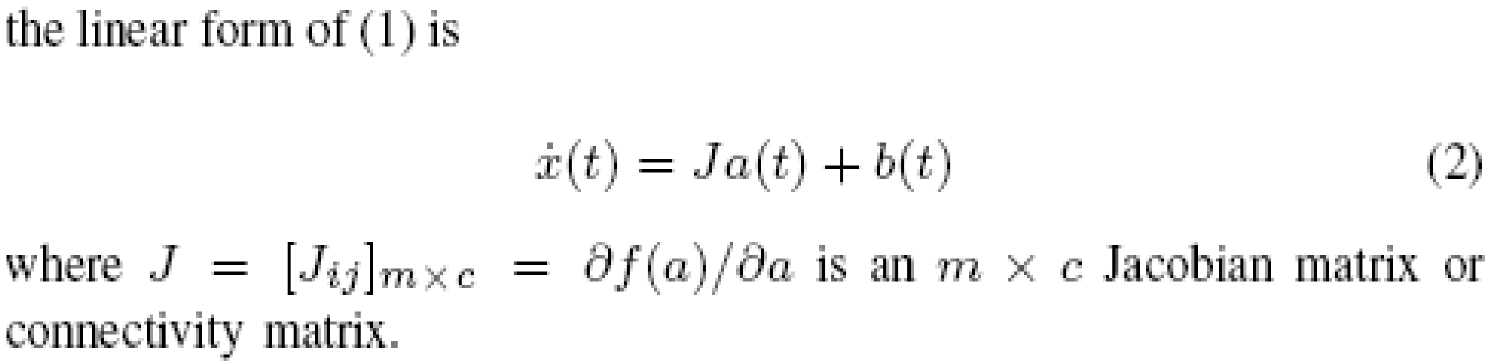
- **The general form**

The transcription processes can be represented by differential equations with gene expression and TFAs:

$$\dot{x}(t) = f(a(t)) - Kx(t) \qquad (1)$$

where $x(t) = (x_1(t), \cdots, x_m(t))^T$ is gene expression level (RNA), $a(t) = (a_1(t), \cdots, a_c(t))$ denotes TF activity level (Protein).

- **The linear form**



the linear form of (1) is

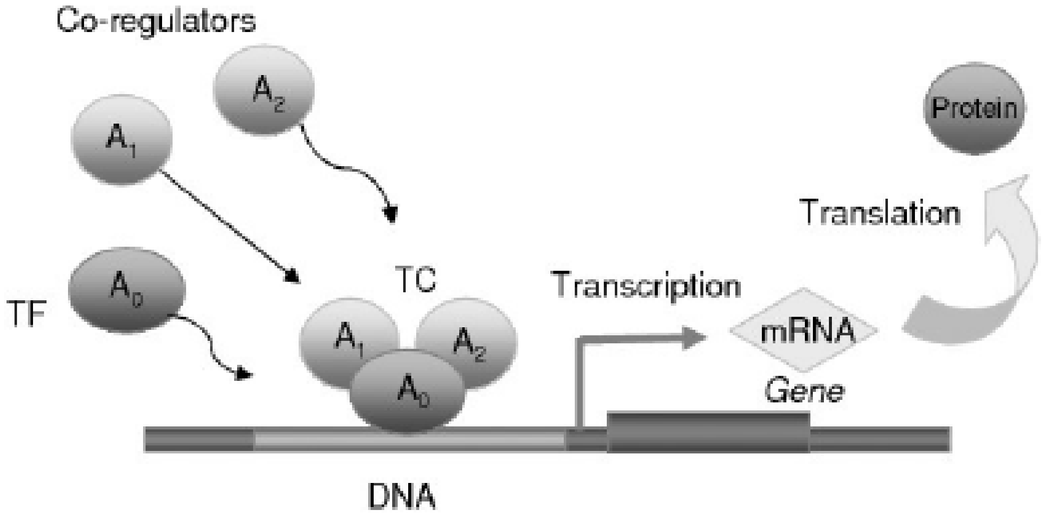
$$\dot{x}(t) = Ja(t) + b(t) \qquad (2)$$

where $J = [J_{ij}]_{m\times c} = \partial f(a)/\partial a$ is an $m \times c$ Jacobian matrix or connectivity matrix.

Approximating TF activity

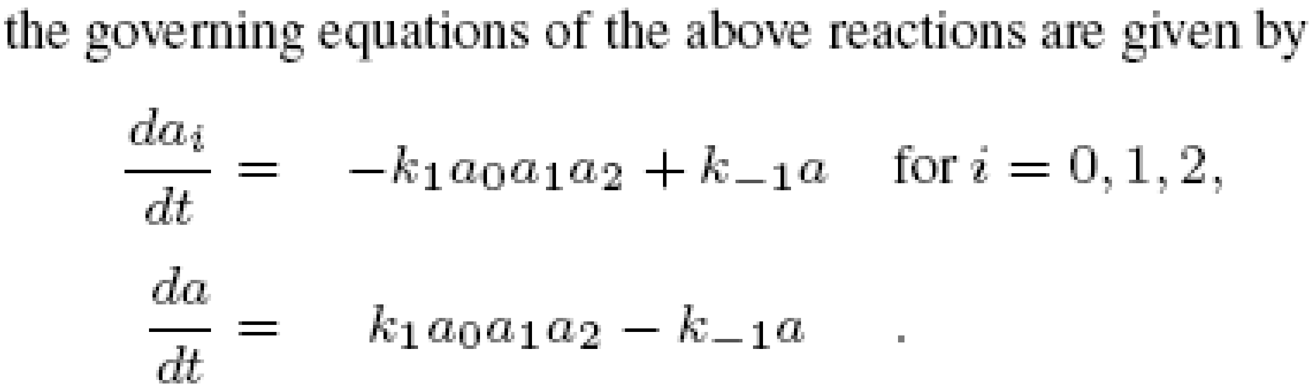
- TFs and many cooperative proteins regulate a gene by a transcription complex (TC).
- TF activity depends on TC.
- A TC is formed by a series of biochemical reactions:

$$A_0 + A_1 + A_2 \rightleftharpoons^{k_1}_{k_{-1}} A$$



Approximating TF activity

- According to the law of mass action,



the governing equations of the above reactions are given by

$$\frac{da_i}{dt} = -k_1 a_0 a_1 a_2 + k_{-1} a \quad \text{for } i = 0, 1, 2,$$

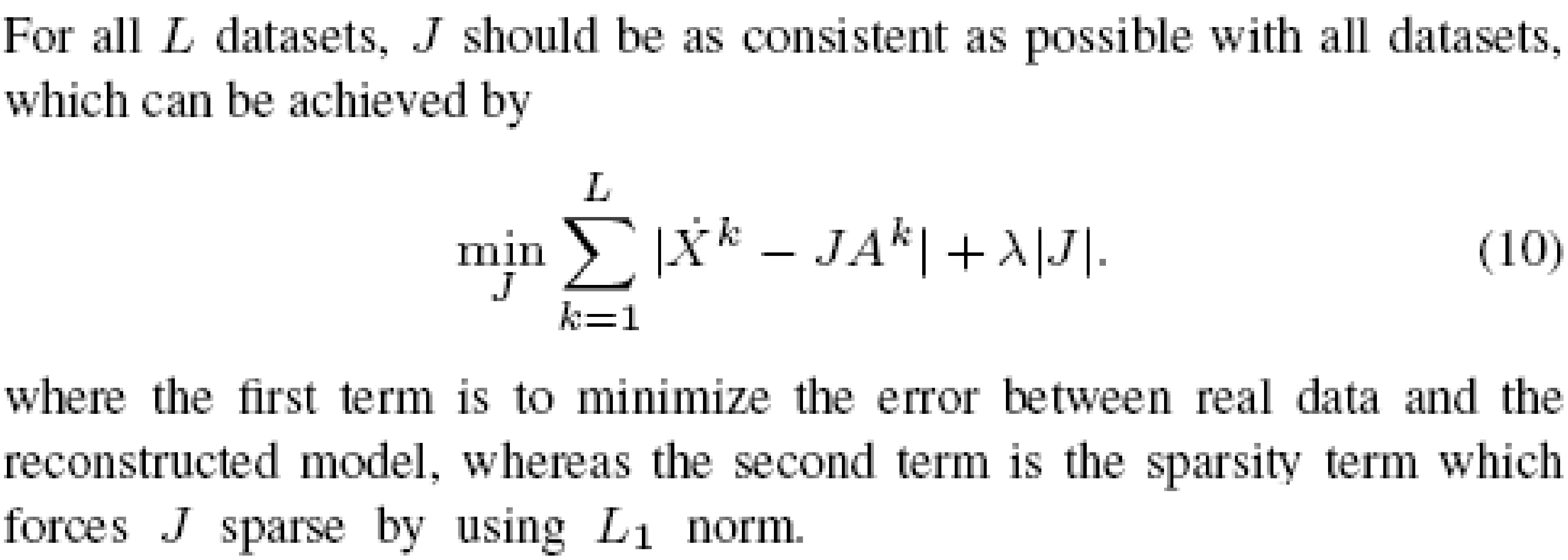
$$\frac{da}{dt} = k_1 a_0 a_1 a_2 - k_{-1} a \quad .$$

- TF activity can be given

$$a = k_0 a_0 a_1 a_2 \approx k_1 x_0 x_1 x_2$$

a : TF activity **x : gene expression**

LP model



For all L datasets, J should be as consistent as possible with all datasets, which can be achieved by

$$\min_{J} \sum_{k=1}^{L} |\dot{X}^k - JA^k| + \lambda |J|. \tag{10}$$

where the first term is to minimize the error between real data and the reconstructed model, whereas the second term is the sparsity term which forces J sparse by using L_1 norm.

Experimental results

- In the budding yeast *S. cerevisiae*, ChIP-chip experiments have been utilized to elucidate the binding interactions between 6270 genes and 113 preselected TFs.

- By checking yeast protein complexes in MIPS, we found 26 TFs in transcriptional protein complexes.

- Among these 26 TFs, some are related to yeast cell cycle and some are related to polyphosphate metabolism in S. cerevisiae

Yeast cell cycle data

- There are 11 TFs that are known to be related to cell-cycle regulation, among which 5 TFs are in 4 different TCs.
- Except these 5 TFs, we selected 8 genes that are closely related to cell cycle based on the information in YEASTRACT (http://www.yeastract.com/index.php).
- According to the gene expression data from Spellman et al. [26], we generated 4 datasets with the number of time points as 18, 17, 24, and 14 respectively.

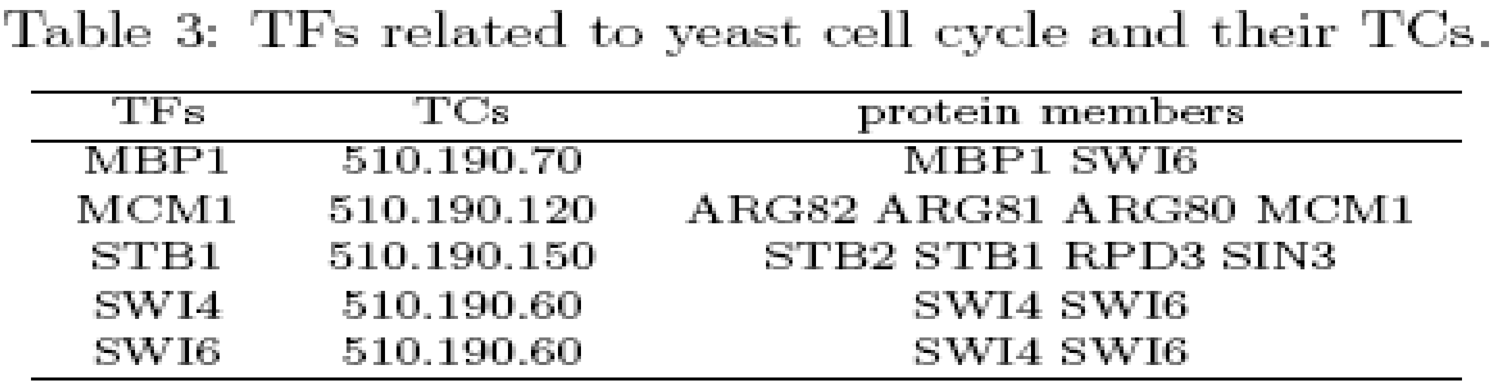
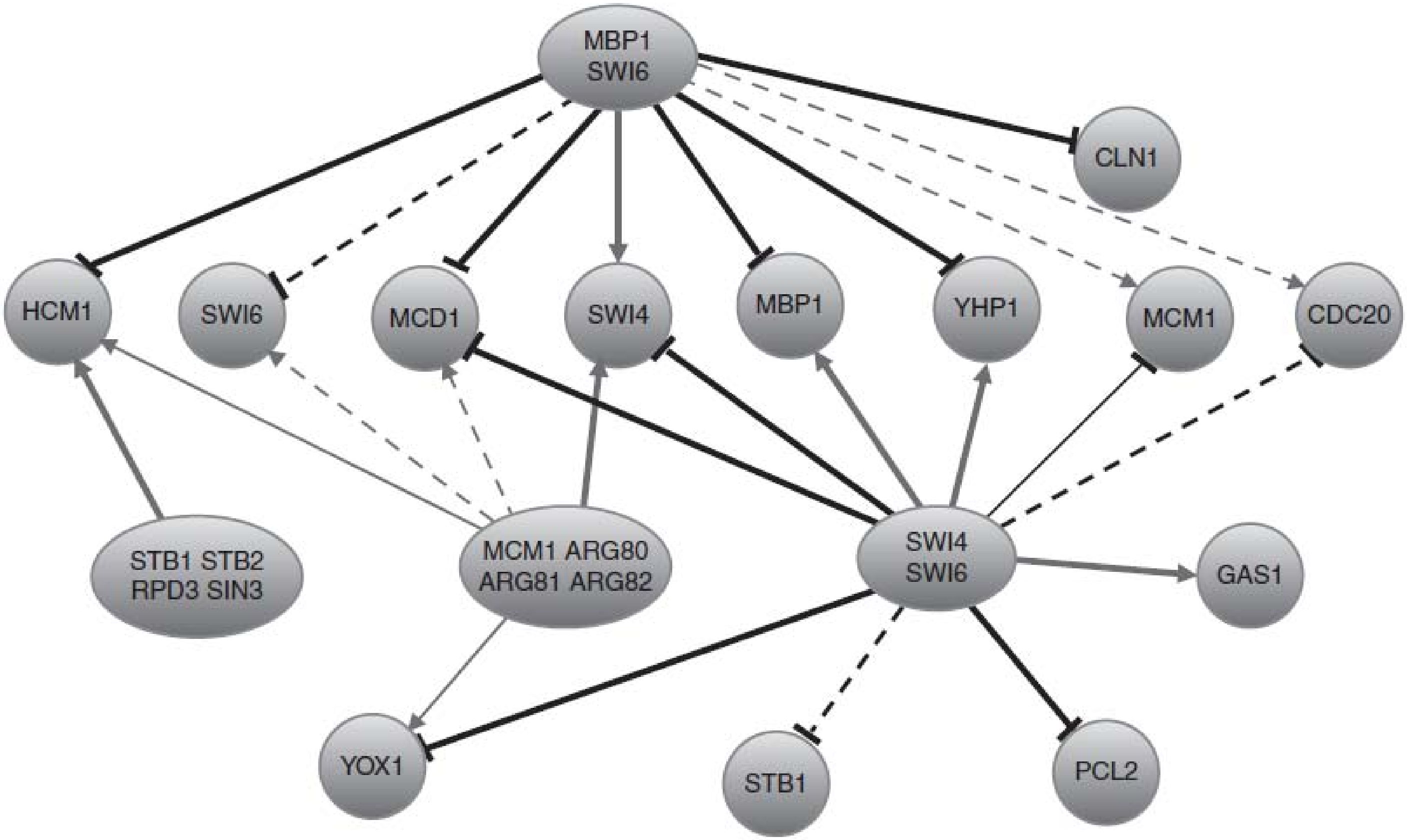


Table 3: TFs related to yeast cell cycle and their TCs.

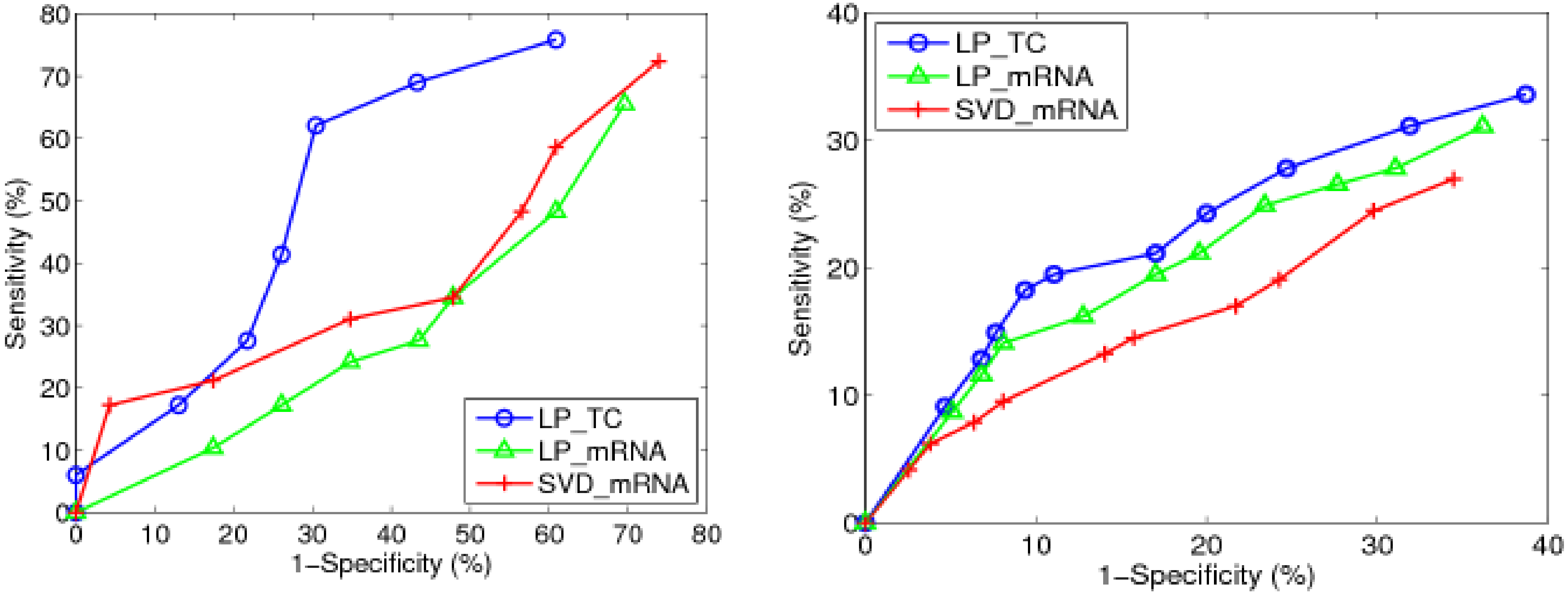
TFs	TCs	protein members
MBP1	510.190.70	MBP1 SWI6
MCM1	510.190.120	ARG82 ARG81 ARG80 MCM1
STB1	510.190.150	STB2 STB1 RPD3 SIN3
SWI4	510.190.60	SWI4 SWI6
SWI6	510.190.60	SWI4 SWI6

Yeast cell cycle data



The inferred yeast cell cycle transcriptional regulatory network. The red arrows in the figure indicate repression while the blue arrows indicate activation.

The comparison results of LP method based on transcription complexes (LP TC), LP method based on only mRNA levels of TFs (LP mRNA) and SVD method based on mRNA levels of TFs (SVD mRNA). (a) on yeast cell cycle data set; (b) on yeast polyphosphate metabolism data set.



Yeast cell cycle data

- We can check the periodicity of the activity levels of the TFs (or TCs) because it is believed that the activities of TFs related to cell cycle tend to be periodic. This fact can be confirmed by Fisher's g-test.

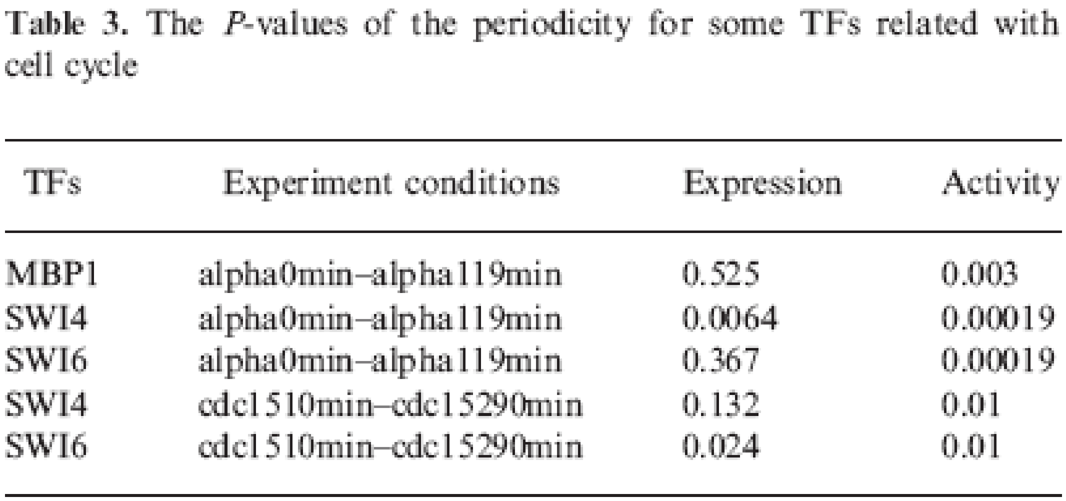


Table 3. The *P*-values of the periodicity for some TFs related with cell cycle

TFs	Experiment conditions	Expression	Activity
MBP1	alpha0min–alpha119min	0.525	0.003
SWI4	alpha0min–alpha119min	0.0064	0.00019
SWI6	alpha0min–alpha119min	0.367	0.00019
SWI4	cdc1510min–cdc15290min	0.132	0.01
SWI6	cdc1510min–cdc15290min	0.024	0.01

Experimental results
---Polyphosphate metabolism data

- Among the TFs related to polyphosphate metabolism verified by the ChIP experiments [8], there are 14 TFs in 9 different TCs.
- Gene expression data: Ogawa N, DeRisi J, Brown PO (2000).
- Among the genes in this dataset, some genes of those with change of 2 fold up or down in at least two time points of the expression levels are believed to be closely related to polyphosphate metabolism.
- In such a way, totally 64 genes (including 14 TFs) form a test data

Polyphosphate metabolism data

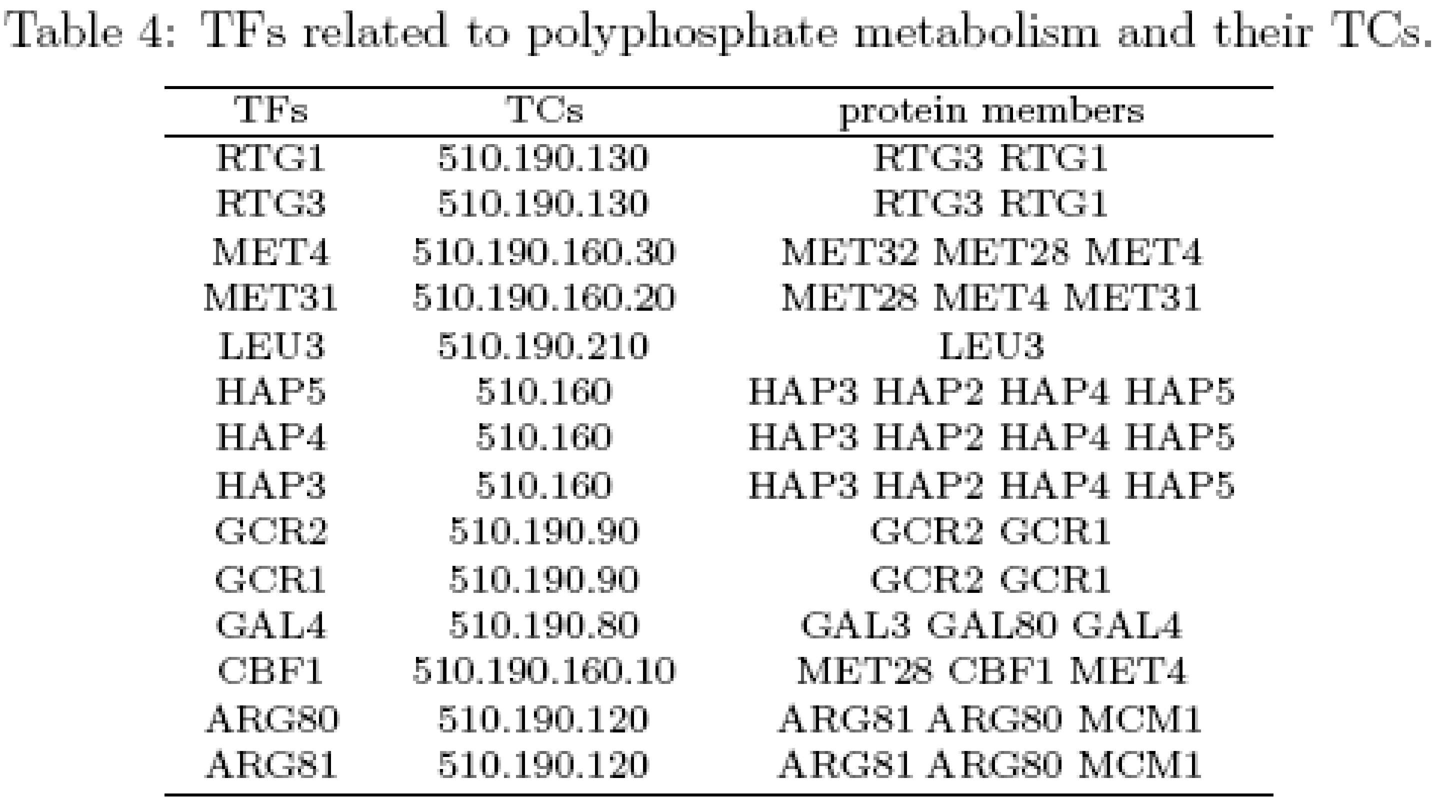
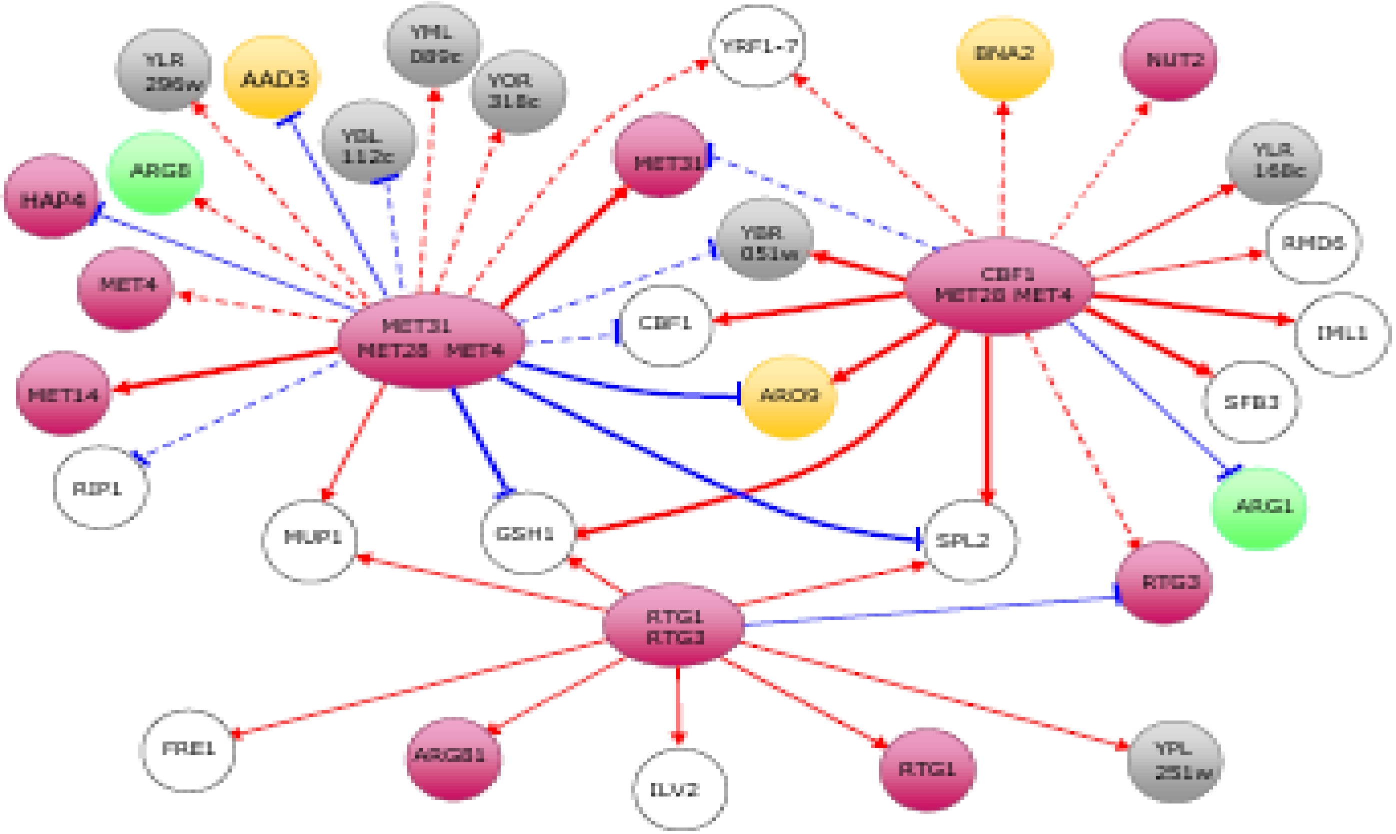


Table 4: TFs related to polyphosphate metabolism and their TCs.

TFs	TCs	protein members
RTG1	510.190.130	RTG3 RTG1
RTG3	510.190.130	RTG3 RTG1
MET4	510.190.160.30	MET32 MET28 MET4
MET31	510.190.160.20	MET28 MET4 MET31
LEU3	510.190.210	LEU3
HAP5	510.160	HAP3 HAP2 HAP4 HAP5
HAP4	510.160	HAP3 HAP2 HAP4 HAP5
HAP3	510.160	HAP3 HAP2 HAP4 HAP5
GCR2	510.190.90	GCR2 GCR1
GCR1	510.190.90	GCR2 GCR1
GAL4	510.190.80	GAL3 GAL80 GAL4
CBF1	510.190.160.10	MET28 CBF1 MET4
ARG80	510.190.120	ARG81 ARG80 MCM1
ARG81	510.190.120	ARG81 ARG80 MCM1

Polyphosphate metabolism data



Transcriptional regulatory network for polyphosphate metabolism. The red arrows in the figure indicate repression while the blue arrows indicate activation.

Take-home messages

- Looking at the same transcriptional regulatory interactions from different perspectives.
- For inferring a TRN, one must first determine which genes or proteins are TFs.
- Furthermore, it is also very difficult to measure the protein concentration levels of TFs and determine their regulatory effects on gene transcription.
- The interactions or cooperations between multiple TFs and their coregulators is a big challenge
- We develop TRNinfer for inferring transcriptional networks by using transcription complexes. http://zhangroup.aporc.org/ResourceBioinformatics