Methods and analysis of TF-DNA binding data
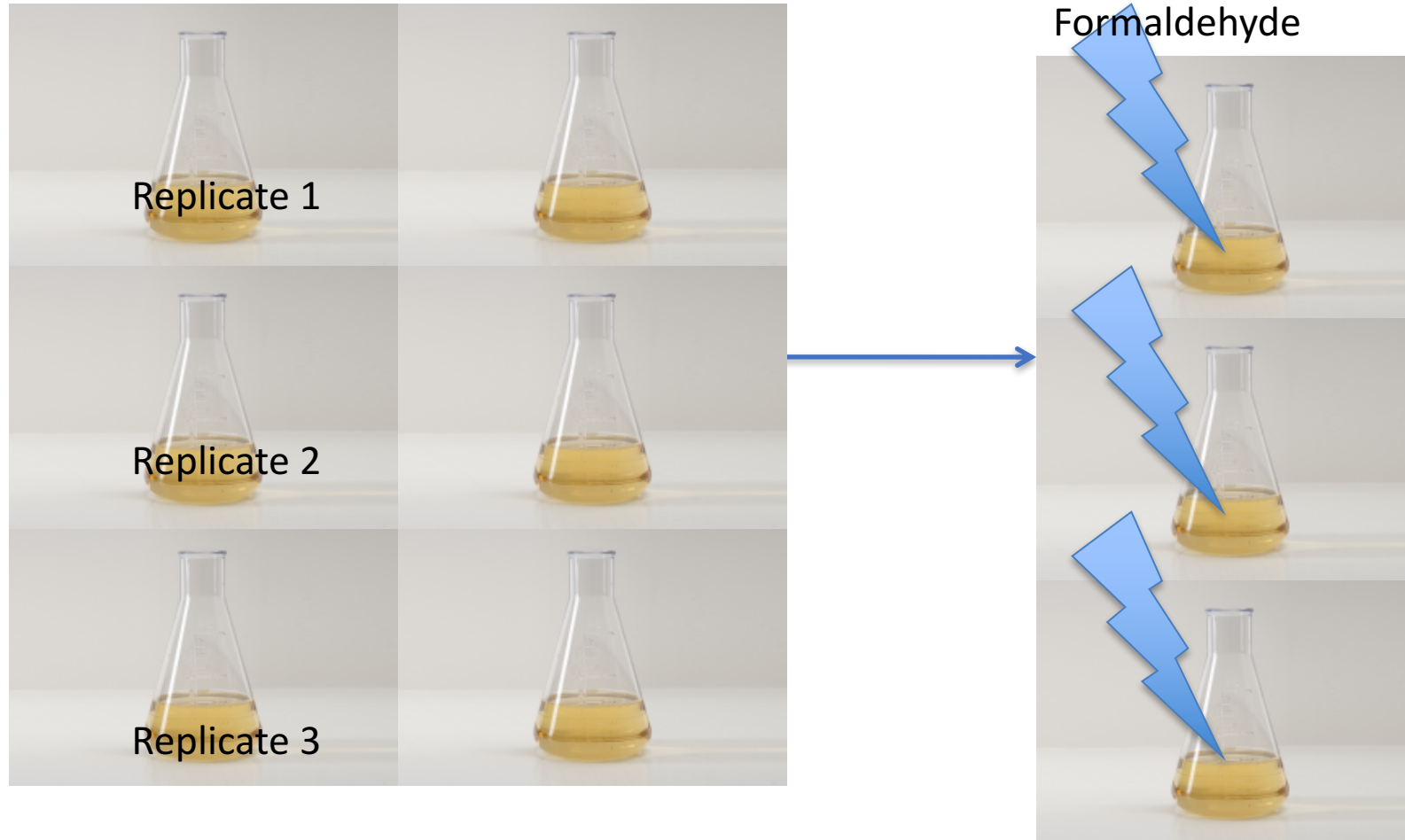
BIO311

02/02/2017

Outline

- Experimental designs
- Review experimental protocol
- Data analysis – from raw scans to binding peaks
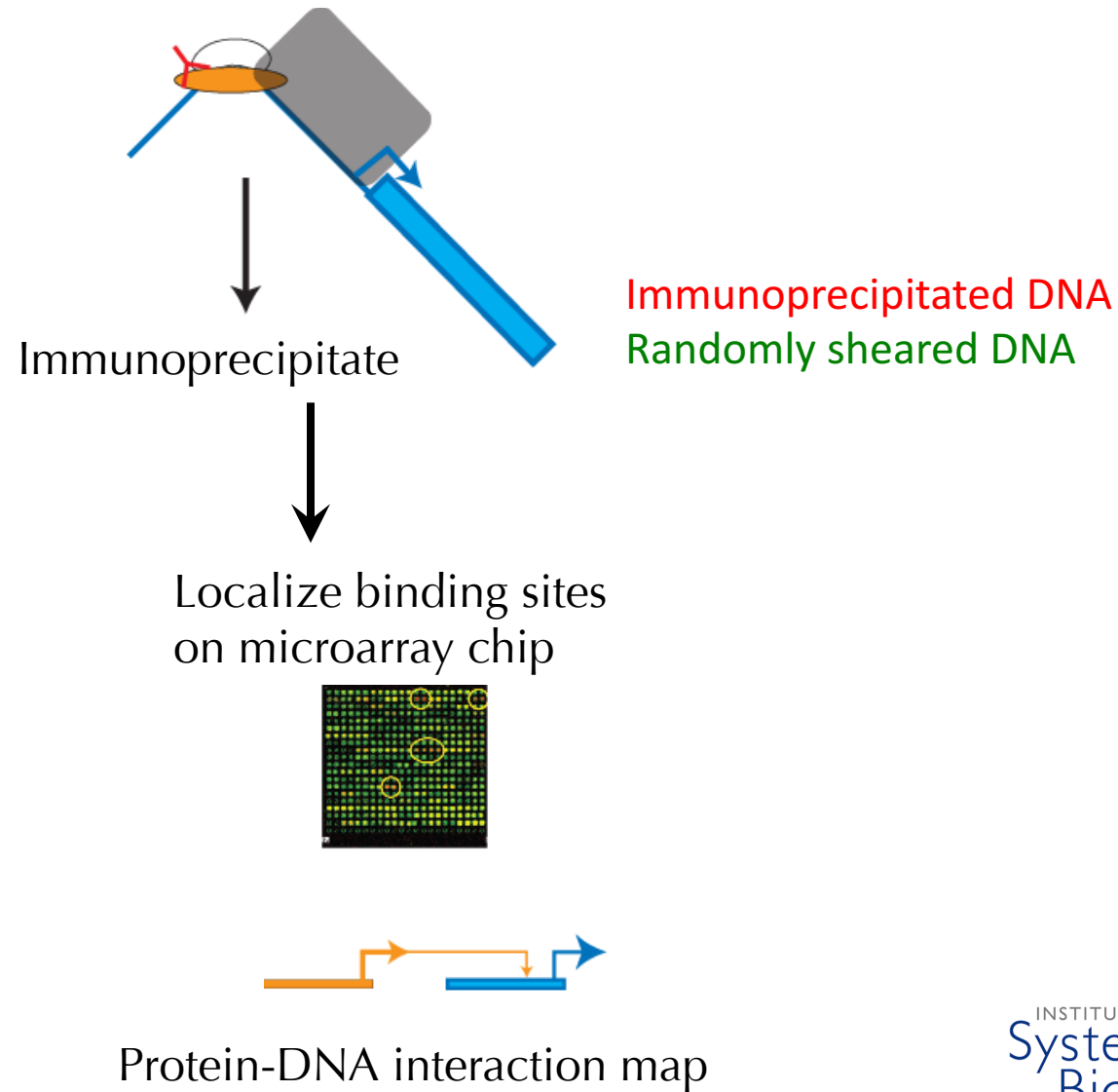- Finding binding motifs – MEME
- Integrated analysis

ChIP-chip experiments and computational analysis

1. _____does each TF bind _____?

2. _____does each TF bind _____?

3. To answer these questions, need to grow cells _____.

ChIP-chip cell harvest

Replicate 1

Replicate 2

Replicate 3

Formaldehyde

Overview: ChIP-chip to determine TF binding locations

Immunoprecipitate

Immunoprecipitated DNA
Randomly sheared DNA

Localize binding sites
on microarray chip

Protein-DNA interaction map

Facciotti, et al., PNAS (2007)

ChIP-chip / ChIP-seq experimental protocol

- What are the steps of the protocol?
1. _____with formaldehyde
2. (Wash)
3. Cell _____
4. Sonication (shear _____ into _____ _____)
5. _____(enrich for TF-DNA of interest), wash, elute,
6. Separate _____
7. _____purification
8. Prepare DNA for _____
9. Label with _____(IP DNA = one color; randomly sheared genomic DNA = other color)
10. Microarray (or seq)

# ChIP-chip data analysis: from raw data to binding lists

1. Preprocessing
   a) Quality control
   b) _____
   c) _____calculation
2. Calculate p-values for each _____for each TF (error model)
3. Combine p-values for _____(weighted average)
4. Significance filter (Fig 1b).
5. Match significant regions to _____.

How exactly was this done?

- http://younglab.wi.mit.edu/regulator_network/
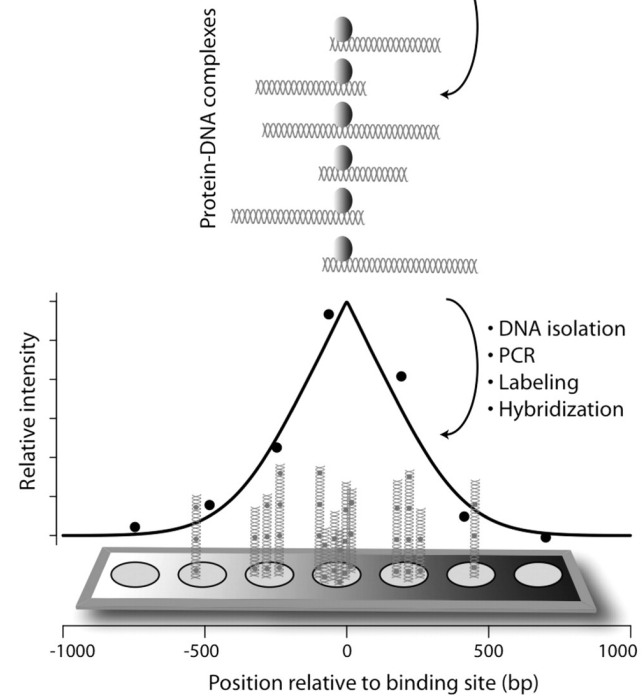- Supplementary data: www.sciencemag.org/cgi/content/full/298/5594/799/DC1

Preprocessing

- Raw spot intensities
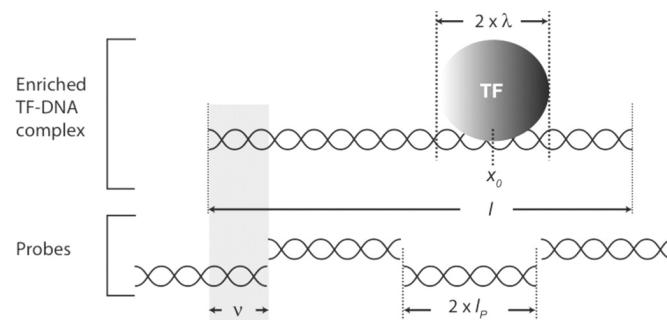- Background subtraction
- Normalizing
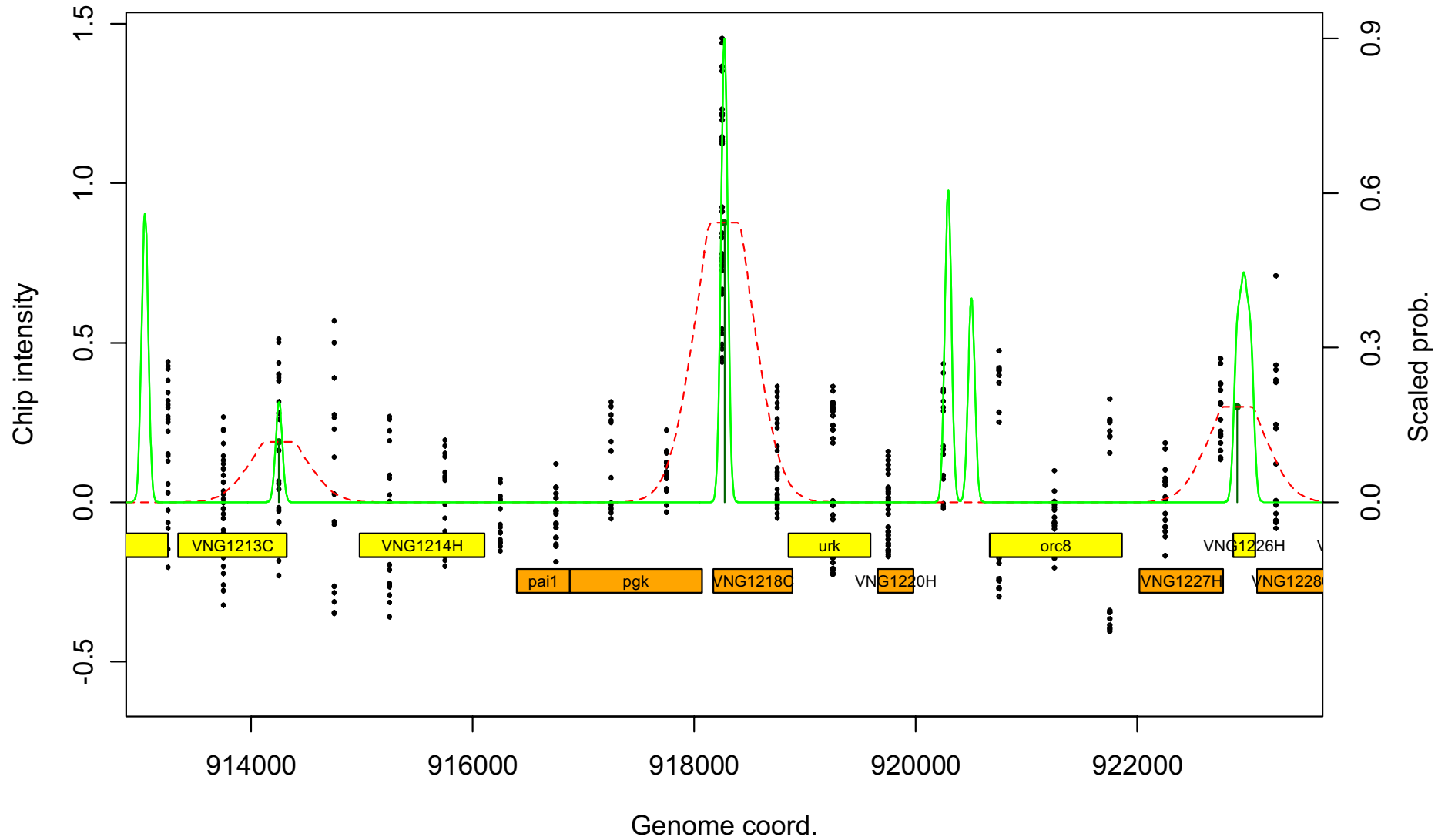- Ratio calculation
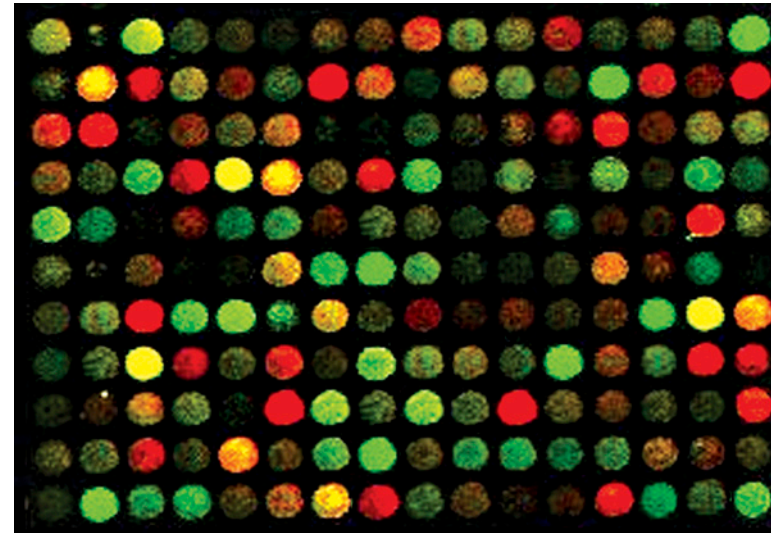
ChIP-chip data analysis: an example

ChIP-chip data analysis: Pick peaks

In your projects

- You will receive data that has already been pre-processed (ratios, p-values, and genes assigned)

- Why will you need to analyze this further in the context of gene expression?

Biological questions

- Are the genes we think are regulated based on gene expression data also bound by the TF?
- Do TF binding events result in a change in gene expression?
- If genes do not overlap between the two datasets, what could this mean?
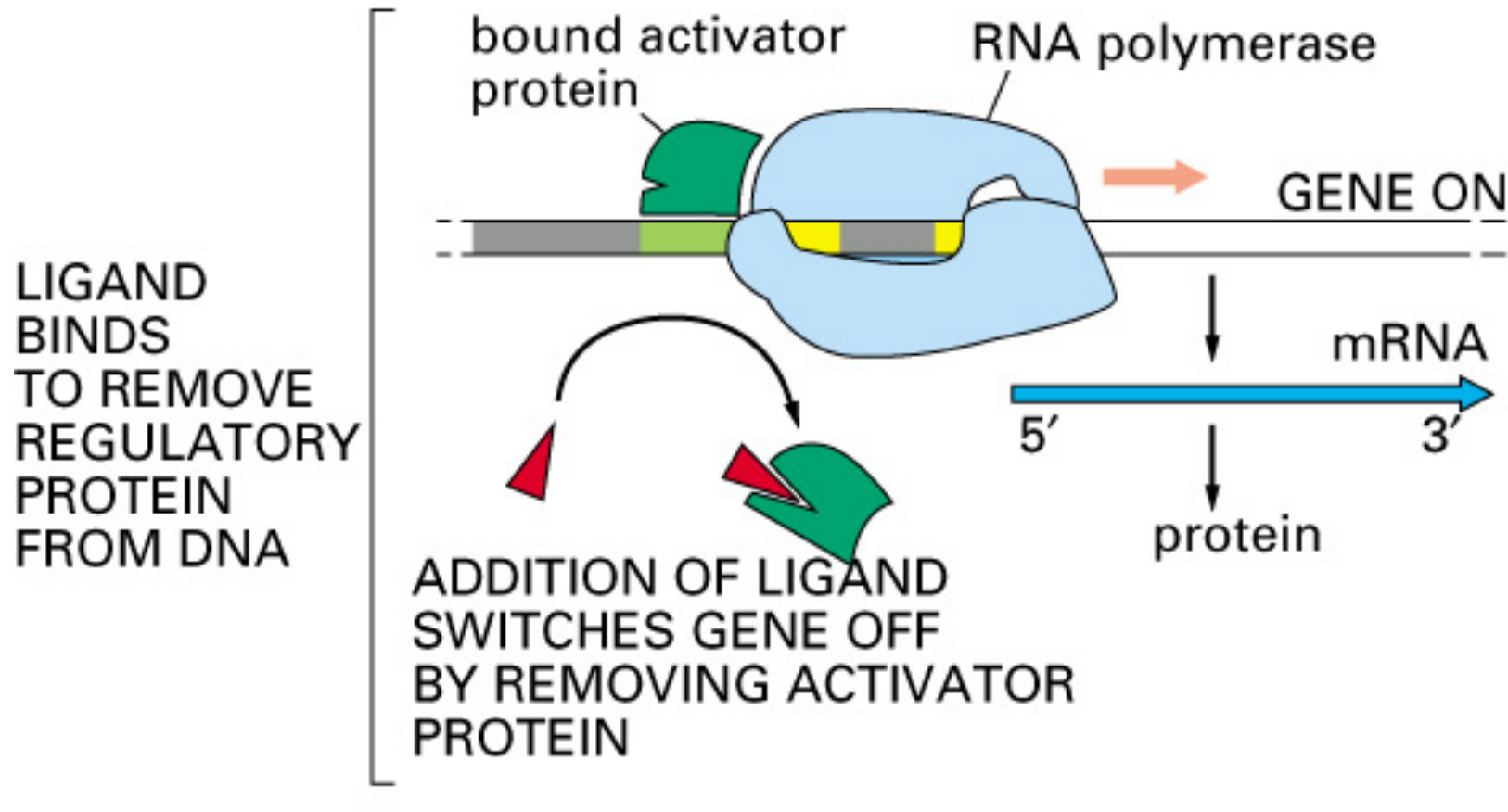
Integrated analysis: GE + ChIP-chip data
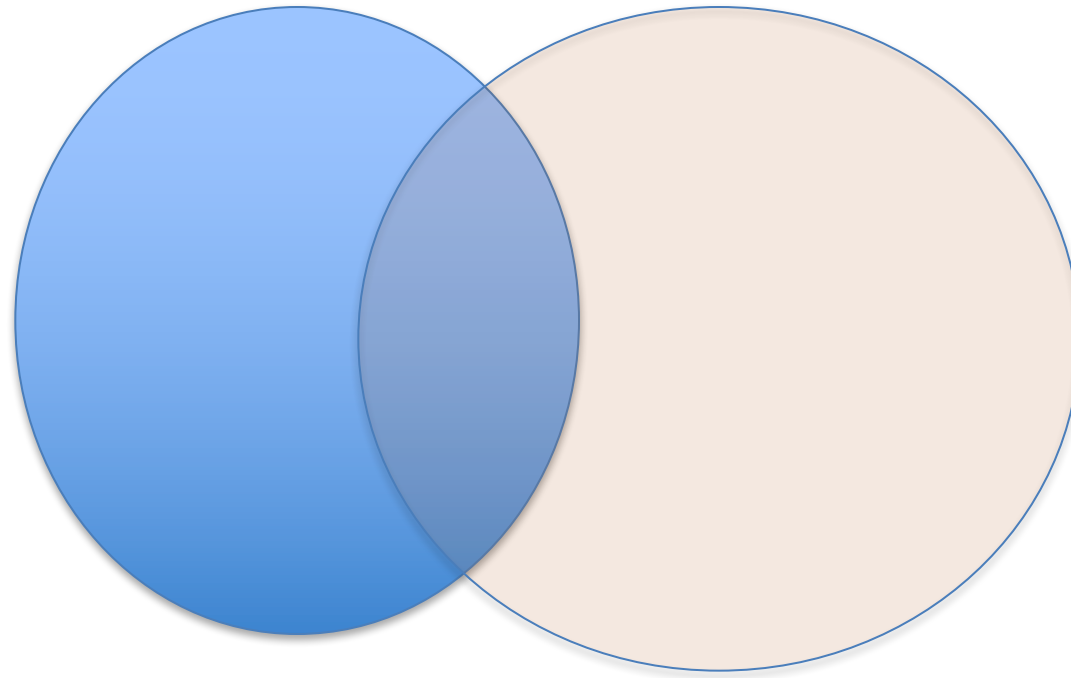
What additional information would ChIP-chip / -seq under different environmental conditions tell us?
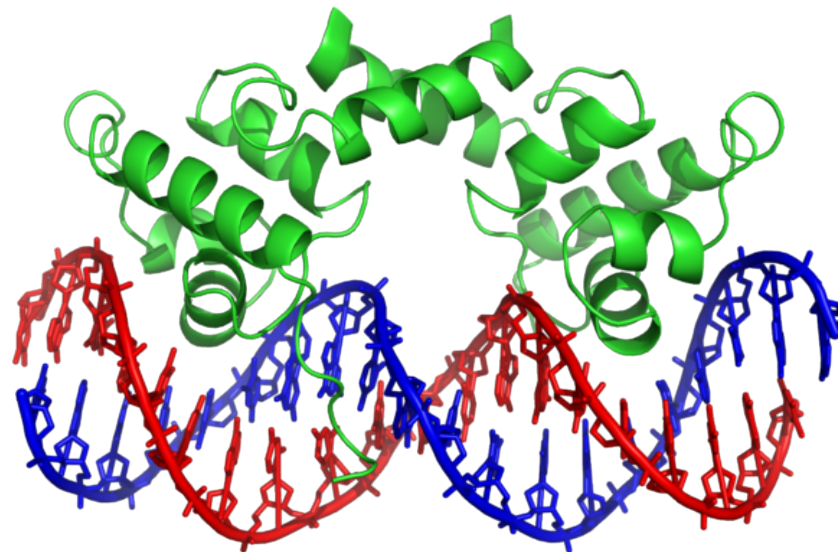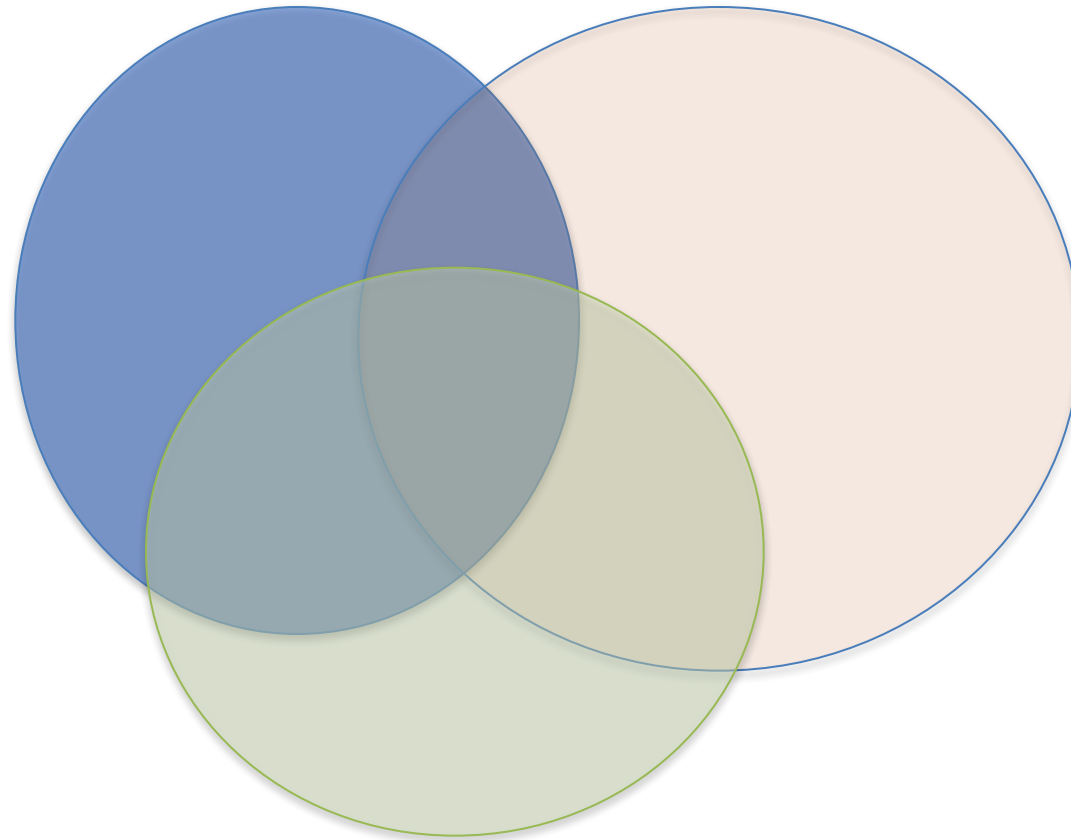
Important consideration: ChIP-chip under varying conditions

(B)

POSITIVE REGULATION
bound activator protein promotes transcription

bound activator protein

RNA polymerase

GENE ON

LIGAND BINDS TO REMOVE REGULATORY PROTEIN FROM DNA

mRNA

5′   3′

protein

ADDITION OF LIGAND SWITCHES GENE OFF BY REMOVING ACTIVATOR PROTEIN

Integrated analysis: ChIP-chip + GE + cis-regulatory binding sequences

Combinatorial complexity

- Heterodimerization
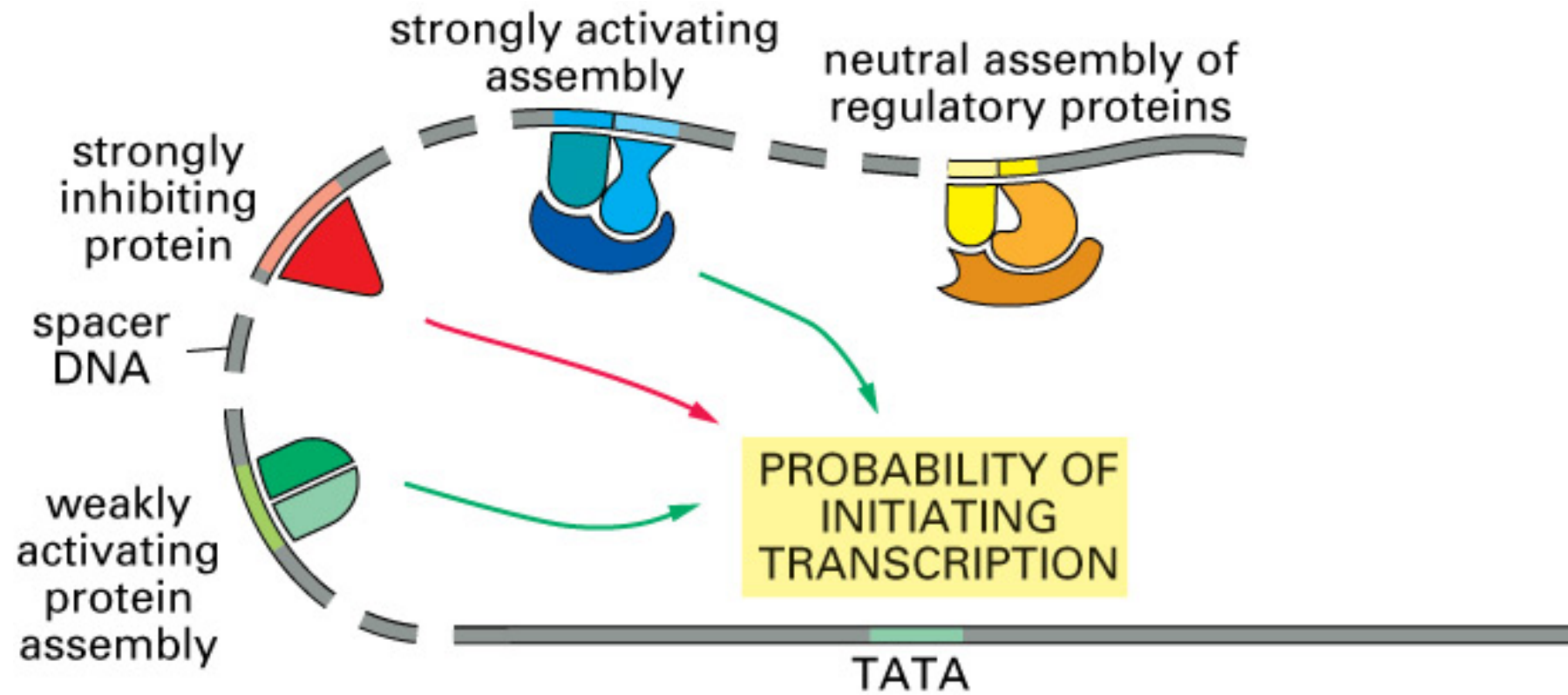- Binding of different transcription factors to same regulatory region

strongly activating assembly

neutral assembly of regulatory proteins

strongly inhibiting protein

spacer DNA

weakly activating protein assembly

PROBABILITY OF INITIATING TRANSCRIPTION

TATA

Figure 7–57. Molecular Biology of the Cell, 4th Edition.

3.19

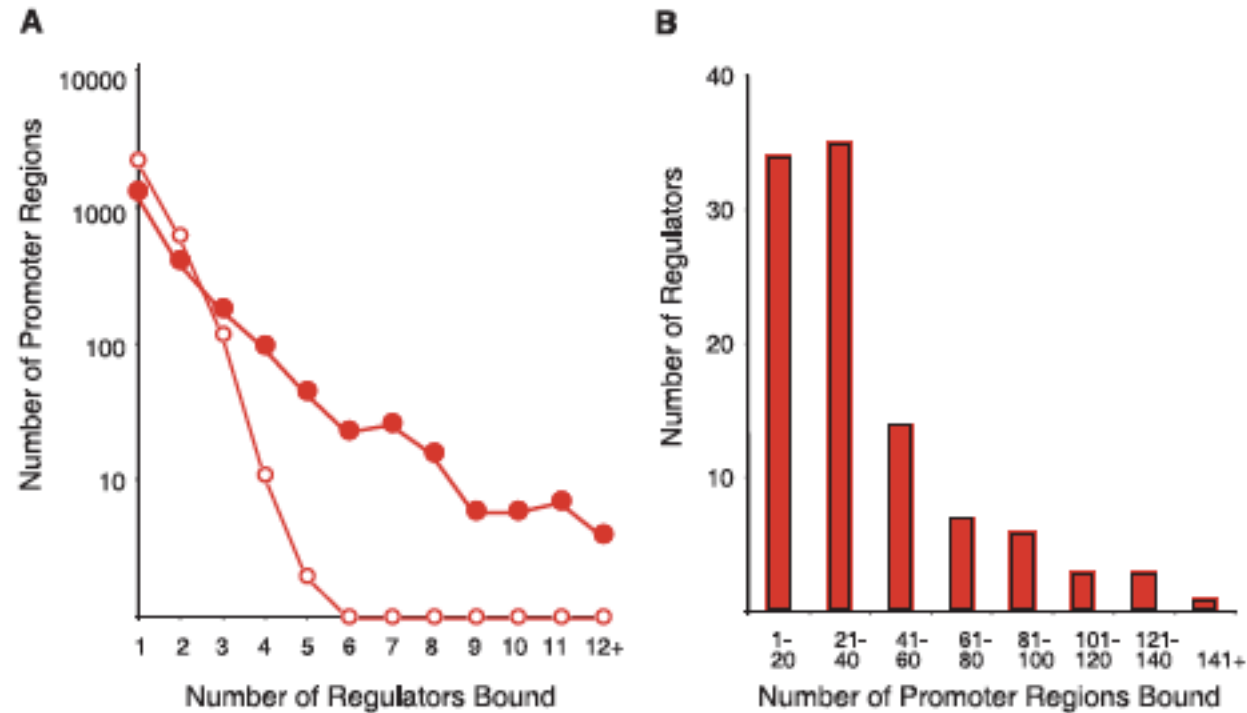# Quantifying combinatorial relationships

Figure 2, Lee et al., 2002.
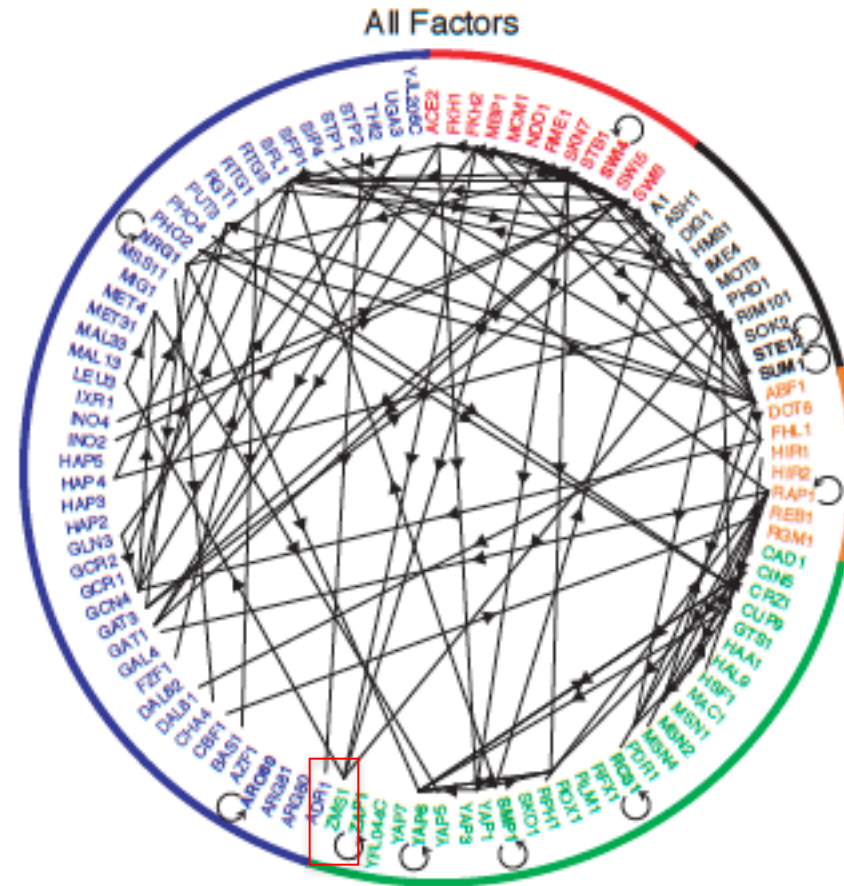
Gene regulatory network

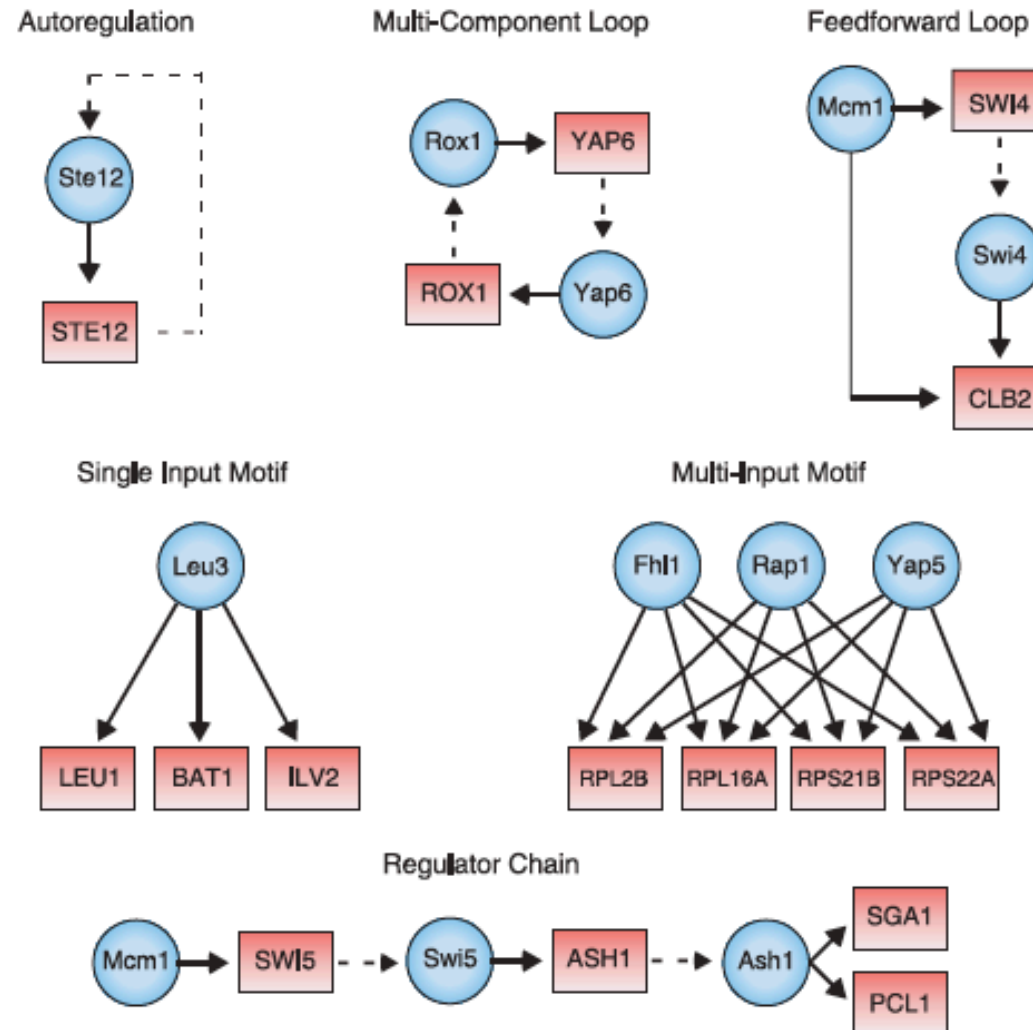Figure 5, Lee et al., 2002

MOTIFS

Fig. 3. Examples of network motifs in the yeast regulatory network. Regulators are represented by

Lee et al., 2002, Science 298:799

- Main conclusions of the study?
  - Recurring _____give clues to _____
  - TFs regulate other _____, providing a link between different cellular processes.
  - Combining global _____data with _____binding data allows building a network with good biological significance with no prior knowledge.

- Many other studies have been done in other organisms (humans, mice, bacteria, flies, etc etc) that have similar conclusions.

Evaluating the ChIP-chip / -seq method

- Drawbacks (also called "caveats") of the approach? What future studies does this open up?

- Why is this method so important?