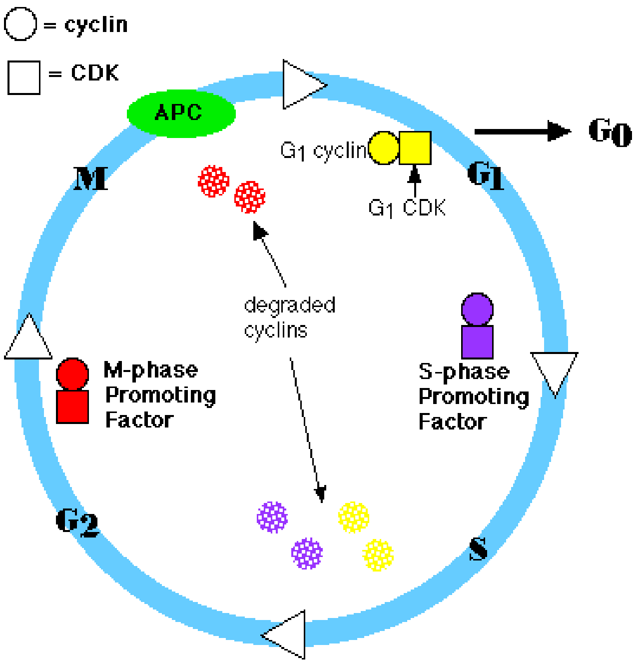
15-780: Graduate Artificial Intelligence

Computational biology: Systems biology and the cell cycle

Movie1.gvp

The Cell Cycle

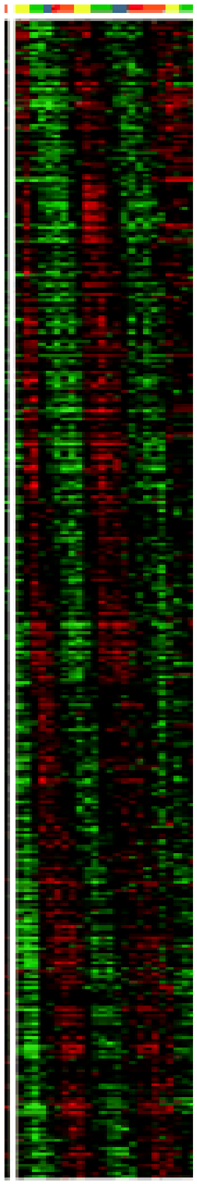
- The process in which cells divide.
- Plays key role in development and cancer.



Time series expression data

Systems Biology Group

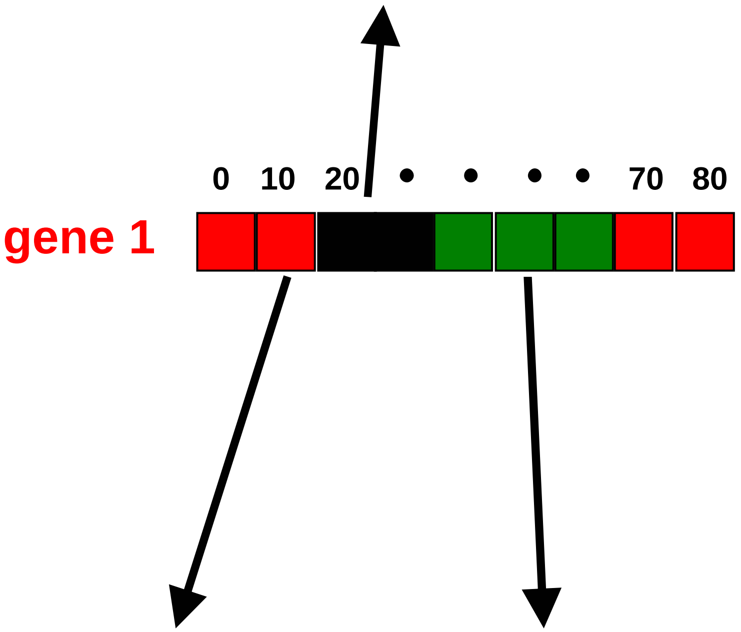
Experiments (over time)



Expression = level of gene (protein) in this experiment

genes

baseline expression



Higher expression compared to baseline

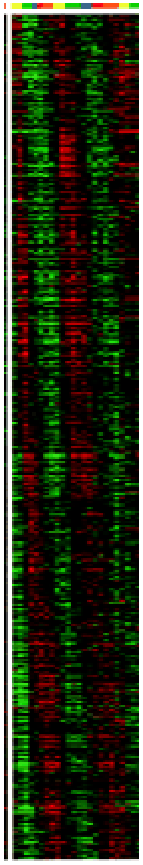
Lower expression compared to baseline

Spellman *et al Mol. Biol. Cell* 1998

Time line

Systems Biology Group

- 1997, 1998 – budding yeast cell cycle expression
- 2000 – plants
- 1999, 2000 - human
- 2001 – mouse

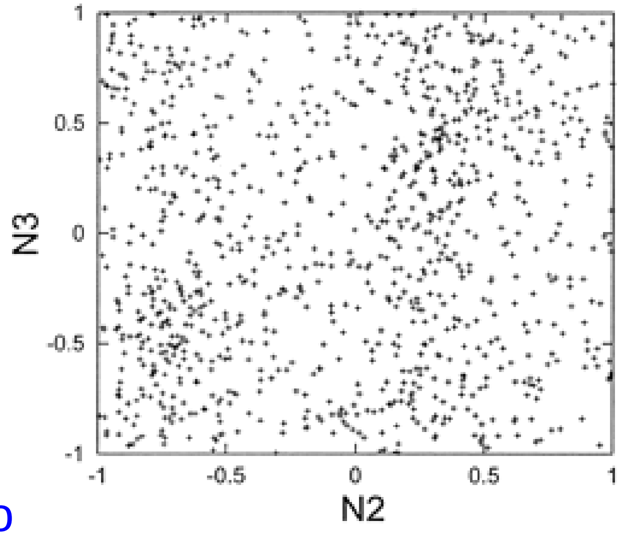


Time line

Systems
Biology
Group

- 1997, 1998 – budding yeast cell cycle expression
- 2000 – plants
- 1999, 2000 - human
- 2001 – mouse

- 2002 – Human data is noise !



Reproducibility of peak between two repeats

Time line

Systems Biology Group

- 1997, 1998 – budding yeast cell cycle expression
- 2000 – plants
- 1999, 2000 - human
- 2001 – mouse

- 2002 – Human data is noise !
- 2002 – Cancer cell cycle expression (approximation)

Time line

Systems Biology Group

- 1997, 1998 – budding yeast cell cycle expression
- 2000 – plants
- 1999, 2000 - human
- 2001 – mouse

- -

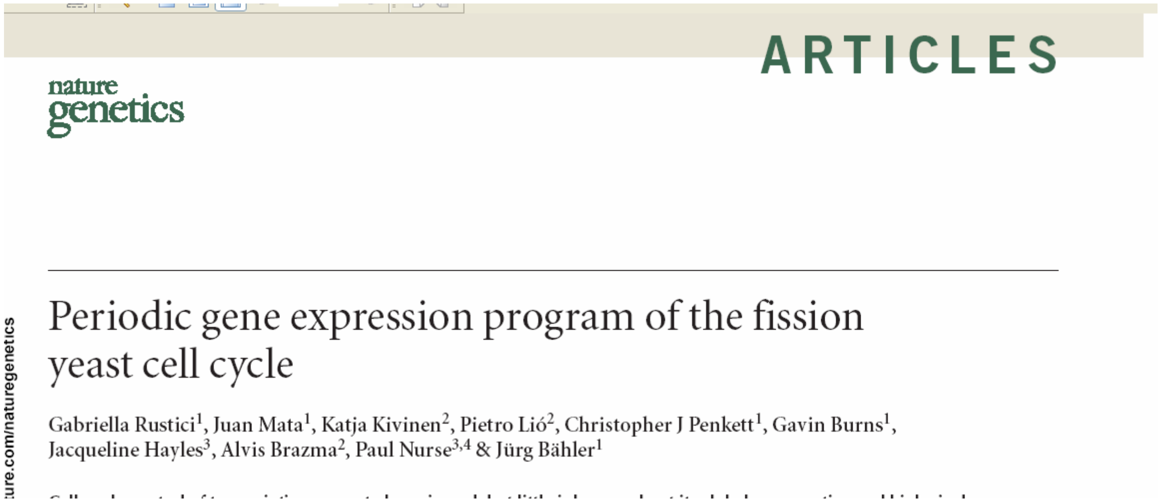
- 2002 – human data is noise !
- 2002 – cancer cell cycle expression (approximation)
- 2004, 2005 – deconvolution and Checksum
- 2006 (upcoming) – human cell cycle data

Time line

- 1997, 1998 – budding yeast cell cycle expression
- 2000 – plants
- 1999, 2000 - human
- 2001 – mouse

- 2002 – human data is noise !
- 2002 – cancer cell cycle expression (approximation)
- 2004, 2005 – deconvolution and Checksum
- 2006 (upcoming) – human cell cycle data

- 2004 – fission yeast cell cycle data



ARTICLES

nature
genetics

Periodic gene expression program of the fission yeast cell cycle

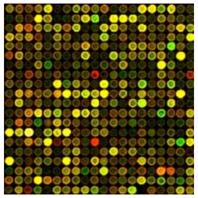
Gabriella Rustici[1], Juan Mata[1], Katja Kivinen[2], Pietro Lió[2], Christopher J Penkett[1], Gavin Burns[1], Jacqueline Hayles[3], Alvis Brazma[2], Paul Nurse[3,4] & Jürg Bähler[1]

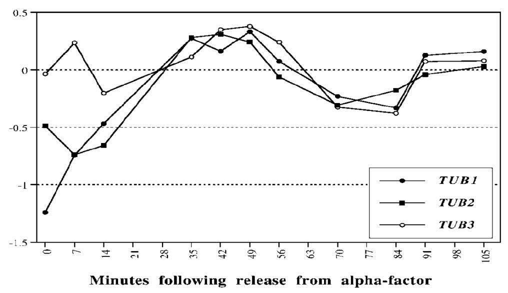
ure.com/naturegenetics

“Our comparisons with budding yeast data revealed a surprisingly small core set of genes that are periodically expressed in both yeasts.”

From expression values to score

- Cells are *synchronized* to the same phase
- Microarray experiments at *multiple time points* after release from synchronization
- Scores derived from multiple expression time series
- Rank genes based on their scores, and use a *cutoff score* to identify cycling genes

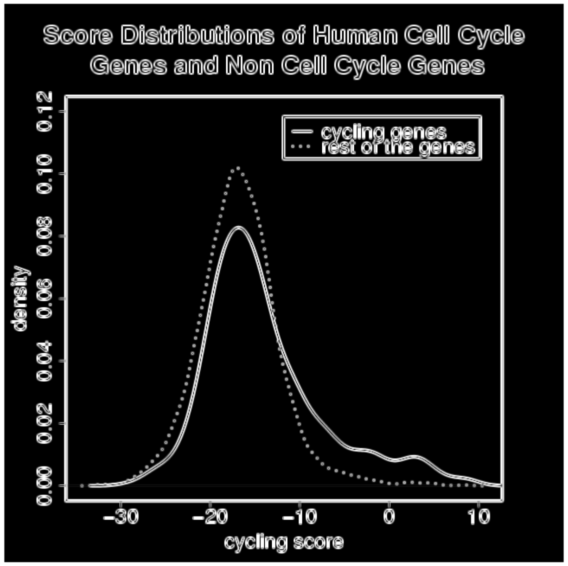




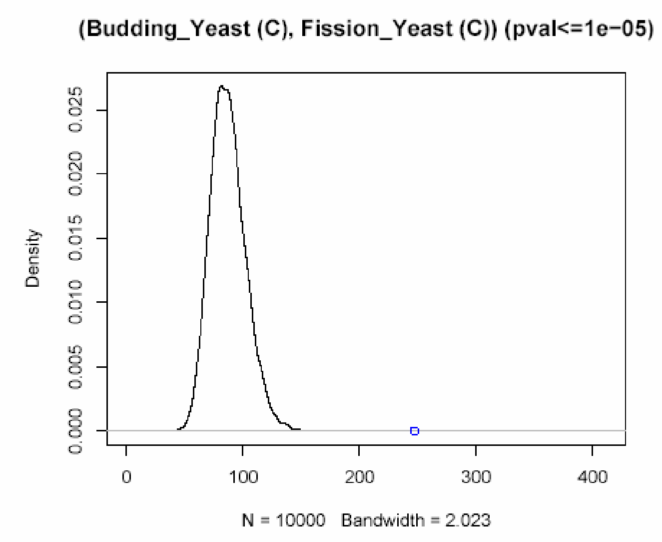
Spellman et al. (1998)

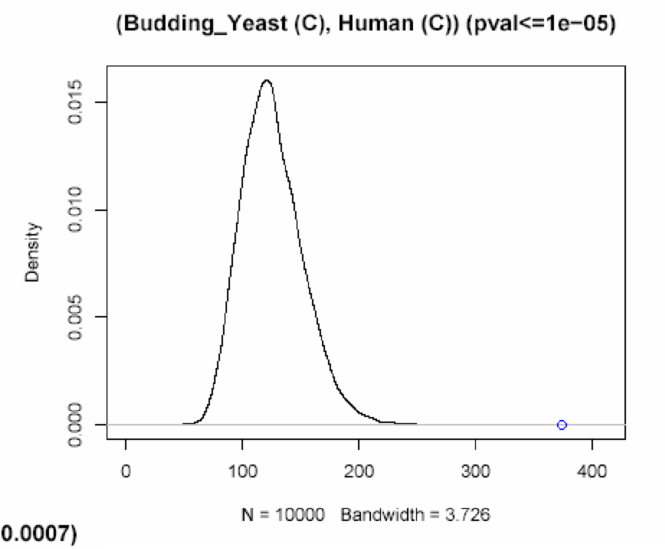
Problems

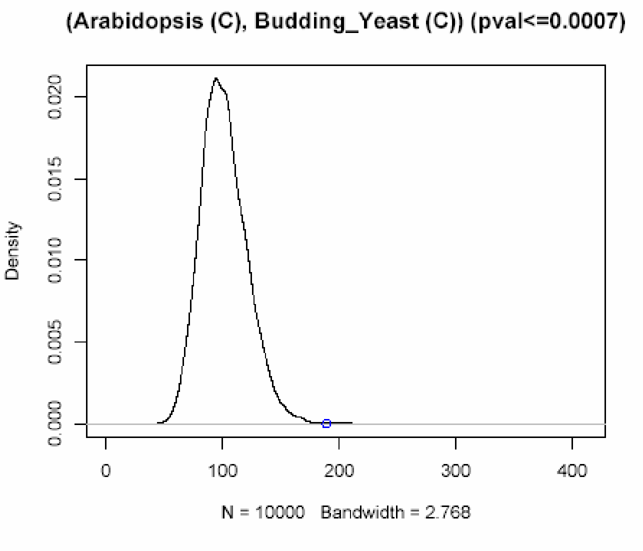
- Different scoring methods result in different lists
- Microarray data are noisy
- Hard to separate scores for cycling and non-cycling genes
    - – Score distribution of cell cycle genes (derived from GO) versus the rest
        - solid curve: cycling genes
        - dotted curve: the rest



Significance of homolog overlap

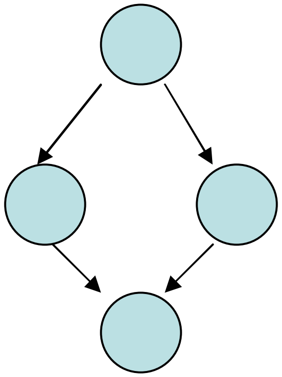






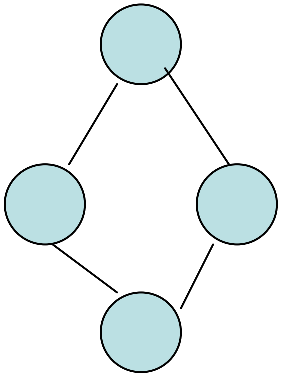
Graphical models

- Efficient way to represent and reason about *joint distributions*
- Graphs in which nodes represent random variables and edges correspond to dependency assumptions
- Two major types: Directed and undirected



$$\prod_i p[x_i \mid Pa(x_i)]$$

- Bayesian networks
- Hidden Markov models



$$\prod_{i,j} \psi_{i,j}(x_i, x_j)$$

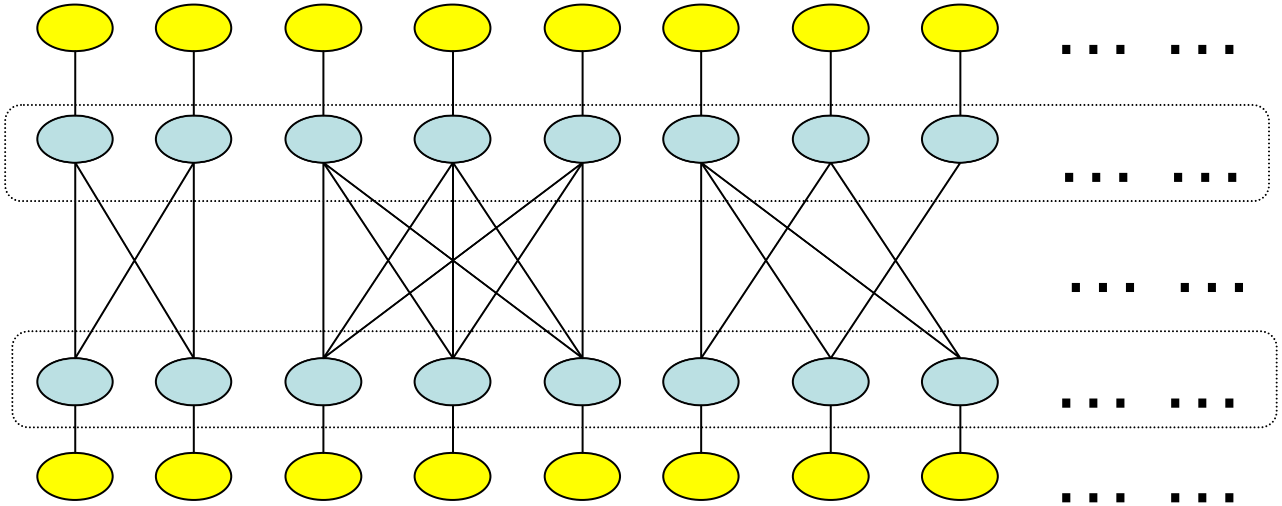
- Markov random fields

Graphical models (cont)

- Parameters are used to specify the conditional probability distribution (directed graphs) or the potential functions (undirected graphs)
- Computational questions:
  - Determining the structure of the model (sometimes)
  - Estimating the parameters of the model
  - Inference

Probabilistic graphical model for combining expression and sequence homology

Systems
Biology
Group



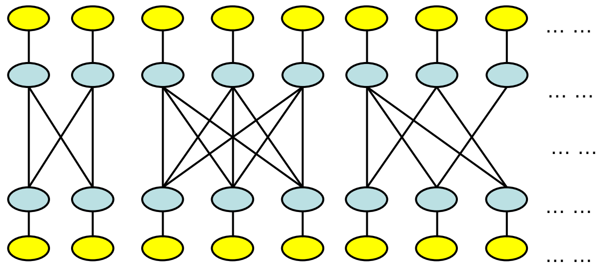
● : C_i : Cycling Status Nodes (unobserved)

● : S_i : Score Nodes (observed)

——— : Encodes Dependency Relations

Numeric summary of expression time series

Likelihood of the model



need to be learned from data

- Node Potential: $\psi_i(C_i) = Pr(C_i|S_i)$
- Edge Potential: $\psi_{ij}(C_i, C_j) = 2^{-\lambda w_{ij}(C_i - C_j)^2}$

controls contribution from each source

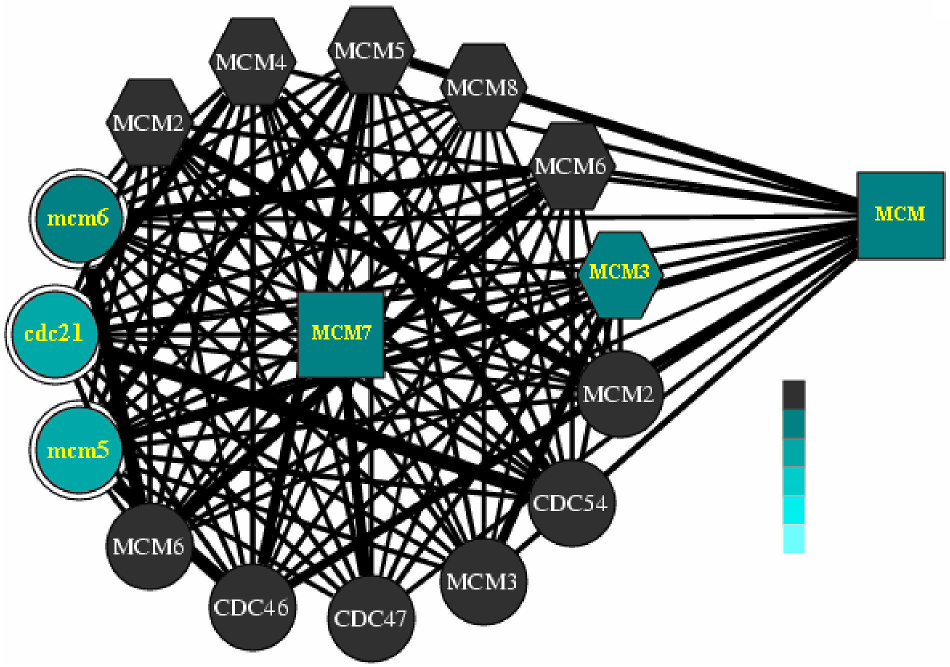
weight (from homology)

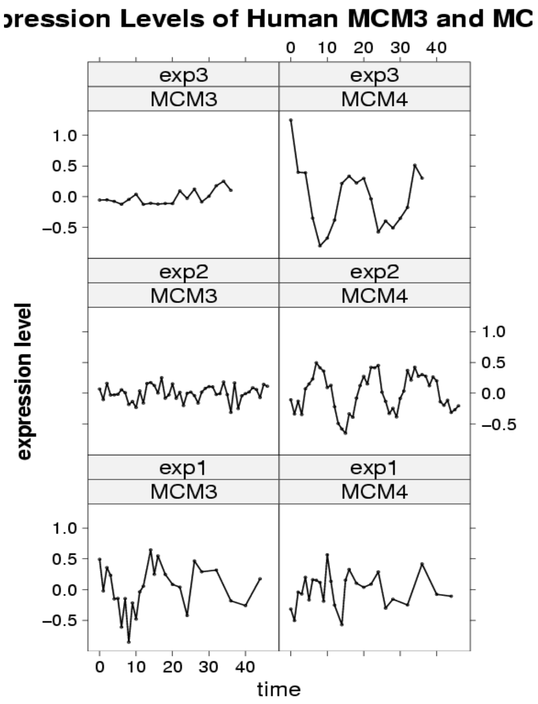
- Joint probability distribution

$$L = \frac{1}{Z}\prod_i \psi_i(C_i)\prod_{i,j}\psi_{ij}(C_i, C_j)$$

In what cases can we improve?

Systems Biology Group

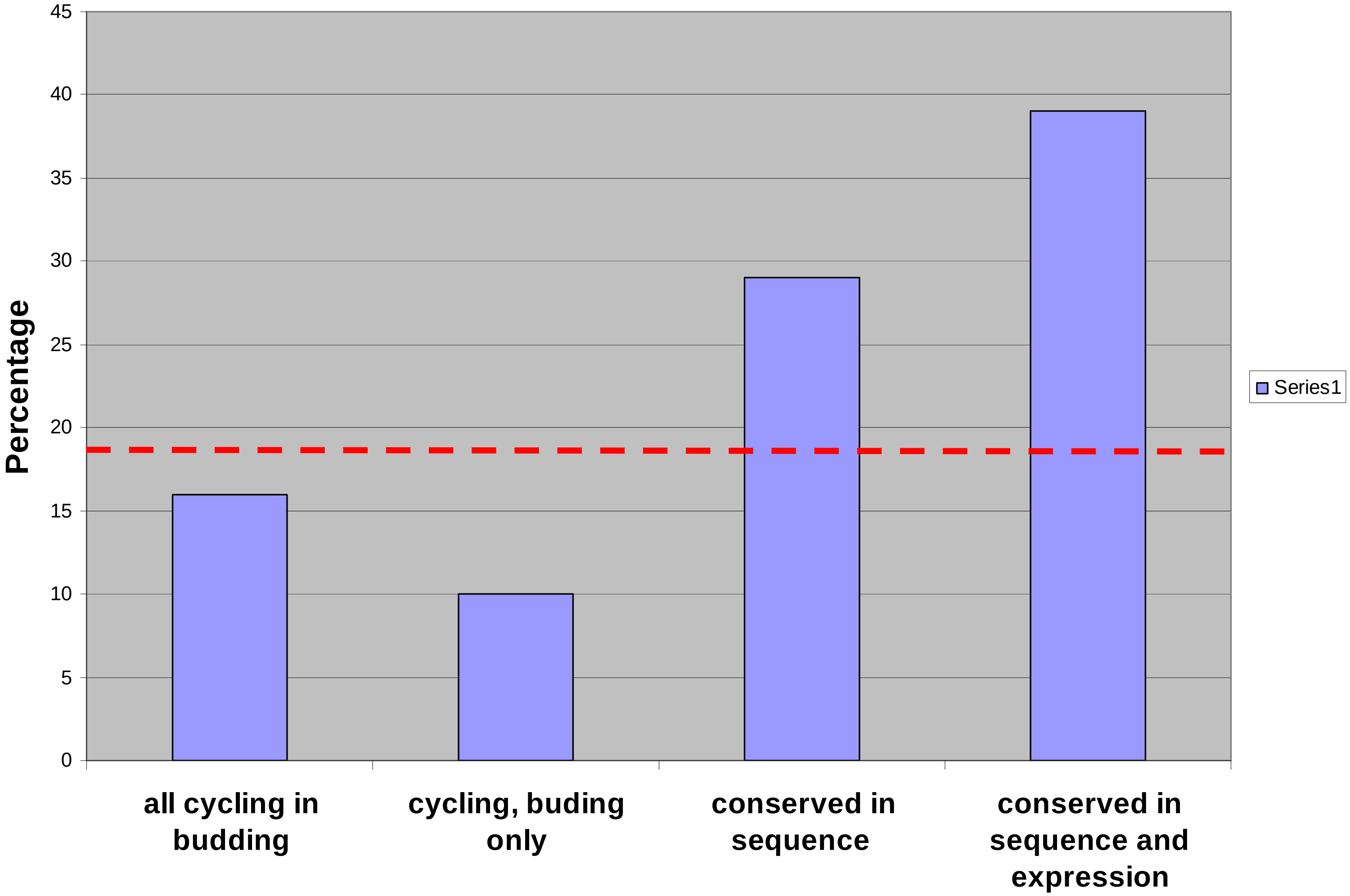




- ⯃ human
- ○ budding yeast
- ◎ fission yeast
- □ Arabidopsis

Essential yeast genes

Systems Biology Group



10-810: Graduate Computational Genomics

Systems
Biology
Group

- Spring 2007
- Bar-Joseph, Benos, Xing
- TR 10:30-11:50, Scaife Hall, 208

- Intro to computational biology emphasizing machine learning, sequence analysis and systems biology.