Systematic determination of genetic network architecture

Saeed Tavazoie, Jason D. Hughes, Michael J. Cambell, Raymond J. Cho, and George M. Church

Nature Genetics 22: 281-285.

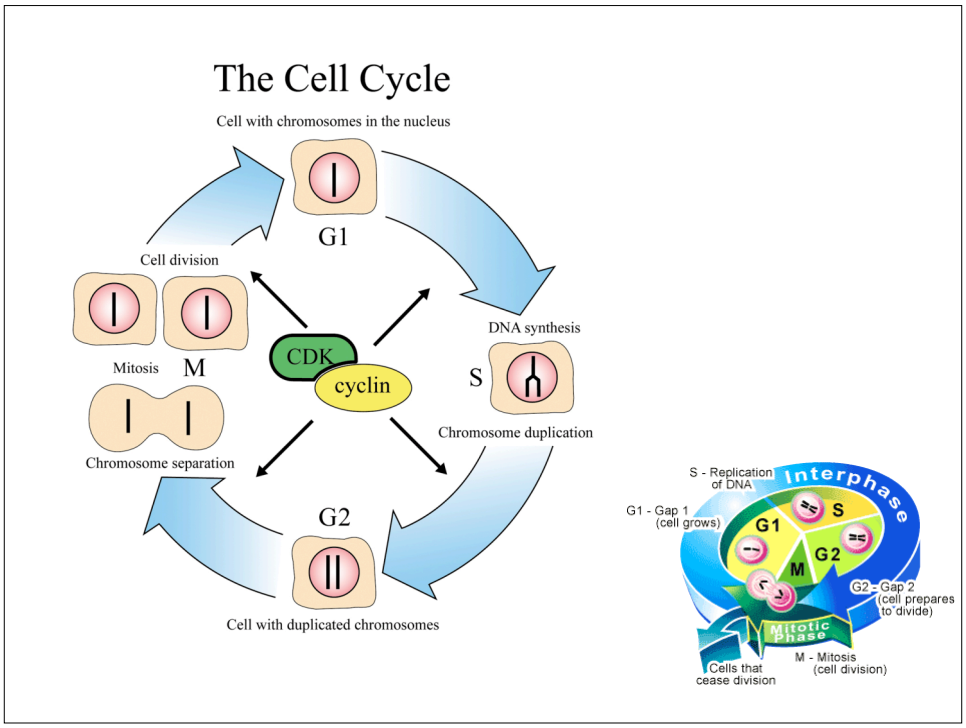
Identification of transcriptional regulatory networks from whole-genome mRNA expression data

Goals:

1. Recapitulate the structure of known networks without *a priori* knowledge of their biology.
2. Identify new components of the network and characterize their significance.

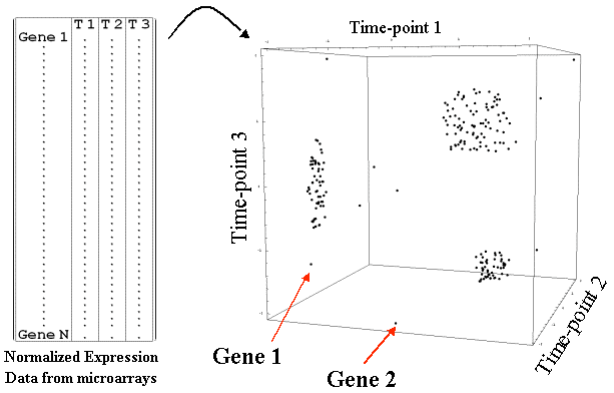
Approaches:

1. Identify **intrinsic mRNA expression patterns** within whole-genome transcriptional profiles.
2. Identify **intrinsic DNA sequence patterns** which are specific to each expression pattern.

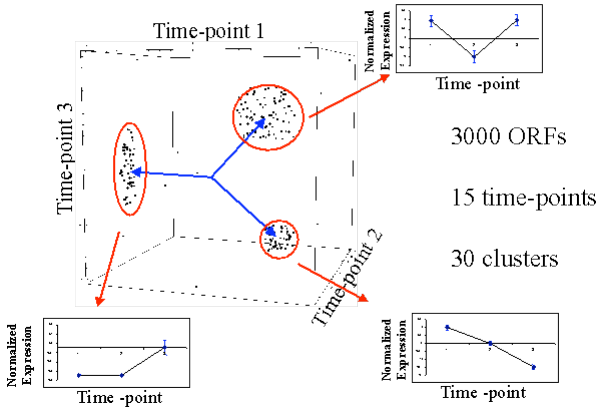

The Cell Cycle
Cell with chromosomes in the nucleus
G1
DNA synthesis
S
Chromosome duplication
G2
Cell with duplicated chromosomes
Mitosis
M
Chromosome separation
Cell division
CDK
cyclin
S - Replication of DNA
Interphase
G1 - Gap 1 (cell grows)
G1
S
G2
M
G2 - Gap 2 (cell prepares to divide)
Mitotic Phase
M - Mitosis (cell division)
Cells that cease division

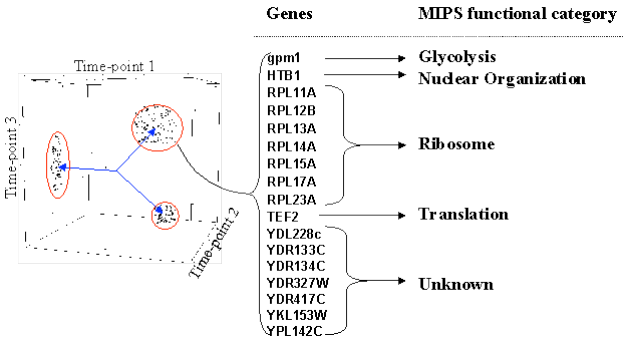
Clustered gene expression
gene
500
1000
1500
2000
2500
3000
2
4
6
8
10
12
14
time

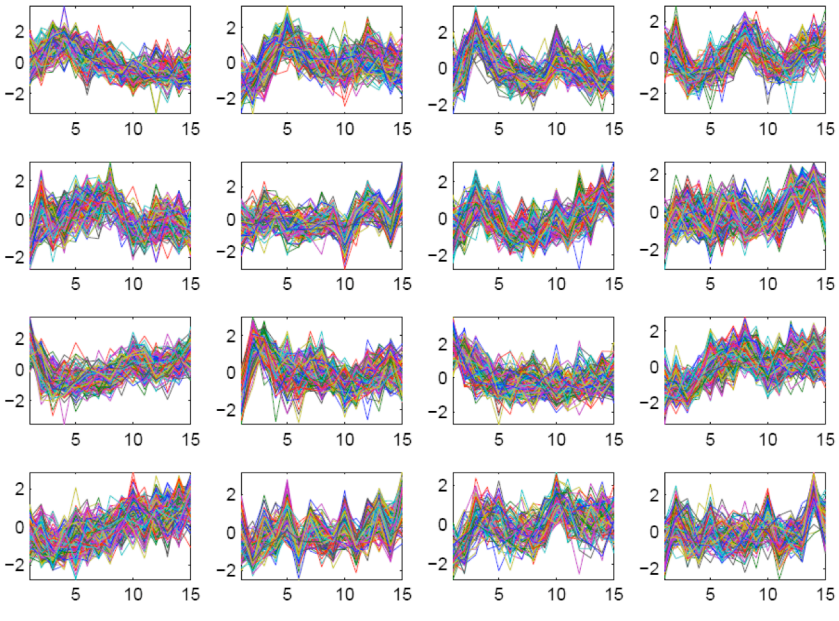
Representation of expression data



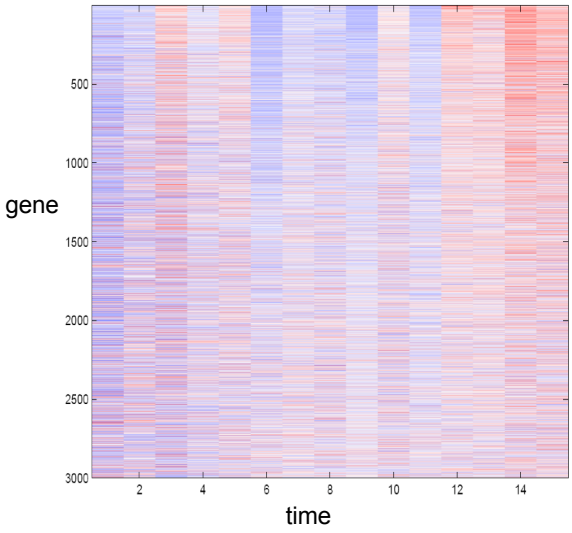
Identifying prevalent expression patterns (clusters)




Time-point 1
Time-point 3
Time-point 2
Genes
MIPS functional category
gpm1
HTB1
RPL11A
RPL12B
RPL13A
RPL14A
RPL15A
RPL17A
RPL23A
TEF2
YDL228c
YDR133C
YDR134C
YDR327W
YDR417C
YKL153W
YPL142C
Glycolysis
Nuclear Organization
Ribosome
Translation
Unknown

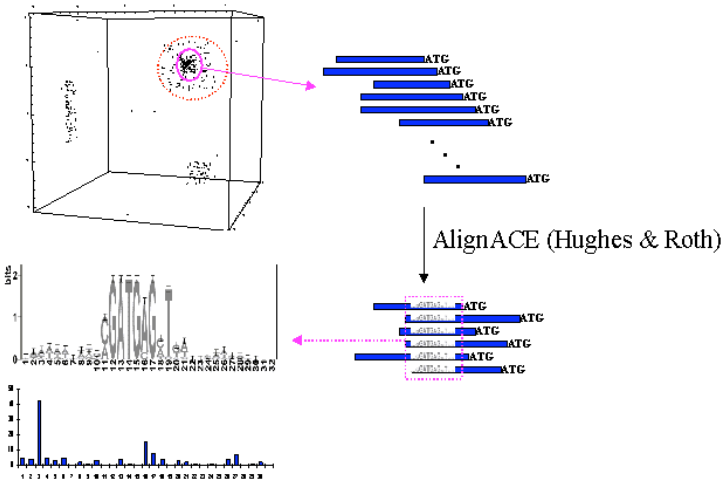

K−Means Clustering of Profiles

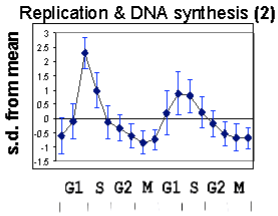
Cell cycle gene expression



3000 genes, 15 time points, (roughly 10 mins interval)

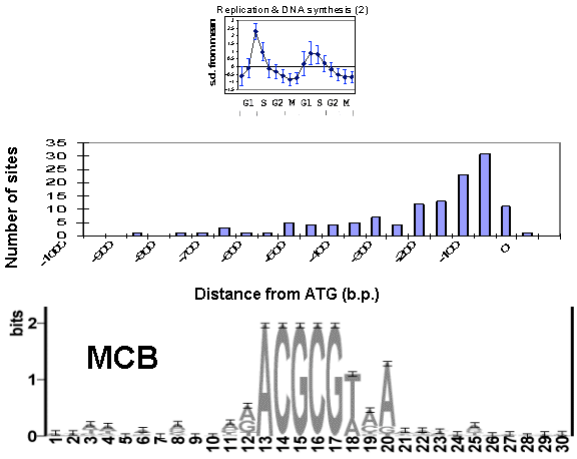
From **expression patterns** to DNA **sequence patterns**

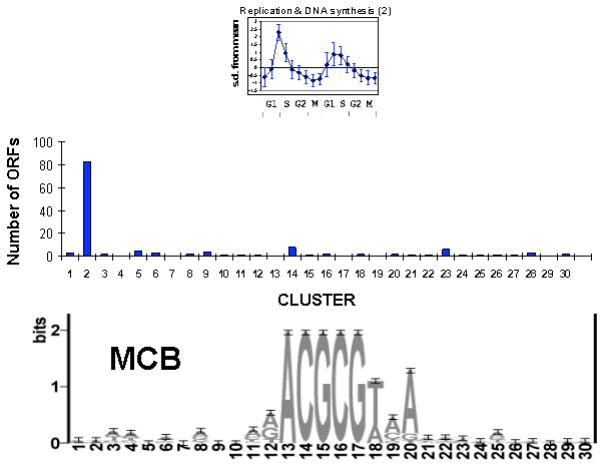


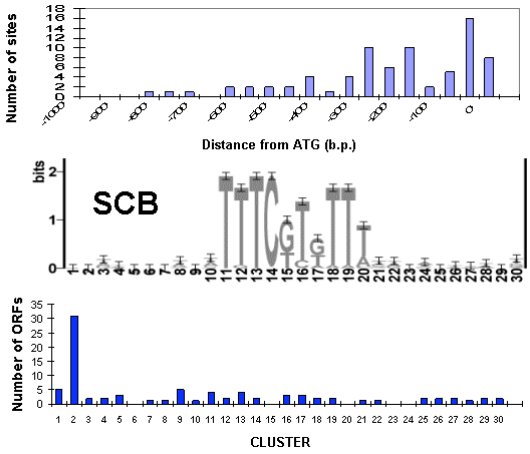


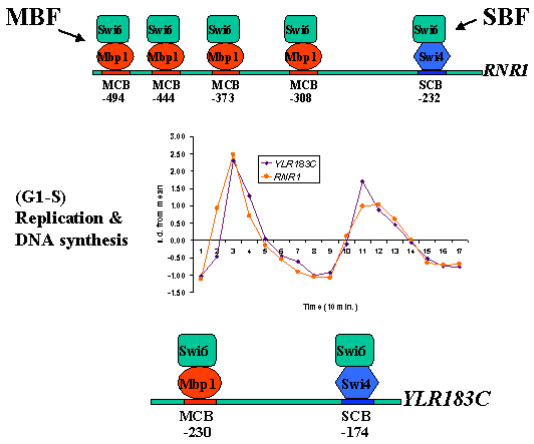
N = 186

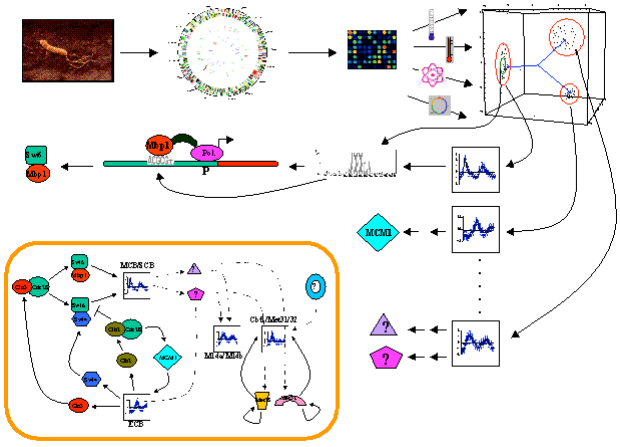
MIPS Functional category (total ORFs)	ORFs within functional category (k)	P-value $-\text{Log}_{10}$
DNA synthesis and replication (82)	23	16
Cell cycle control and mitosis (312)	30	8
Recombination and DNA repair (84)	11	5
Nuclear organization (720)	40	4




Replication & DNA synthesis (2)
sd. from mean
G1 S G2 M G1 S G2 M
Number of ORFs
CLUSTER
bits
MCB

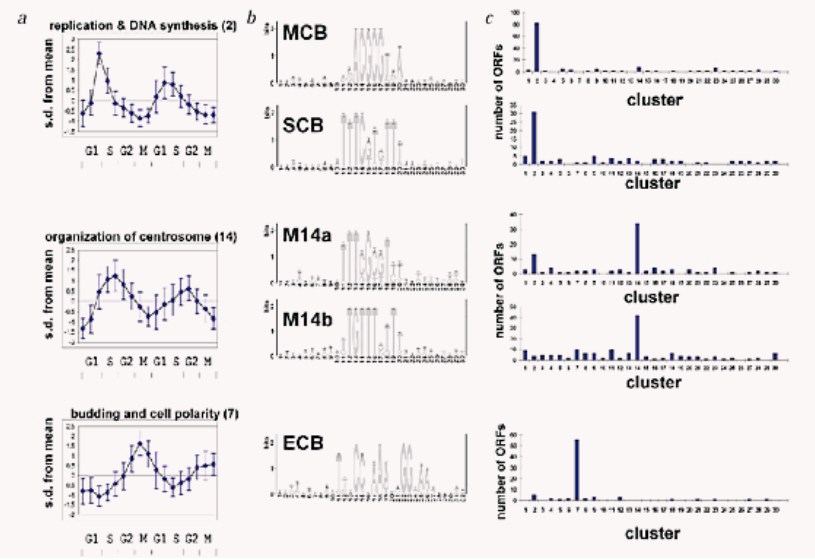

Number of sites
Distance from ATG (b.p.)
bits
SCB
Number of ORFs
CLUSTER


MBF
SBF
Swi6
Mbp1
Swi4
RNR1
MCB -494
MCB -444
MCB -373
MCB -308
SCB -232
(G1-S) Replication & DNA synthesis
YLR183C
RNR1
Time (10 min.)
MCB -230
SCB -174
YLR183C


MCM1
MCB/SCB
ECB

Method

- Used microarray data by Cho et al. 1998 that consisted of expression data for 6000 genes at 15 times points during two *S. cerevisiae* mitotic cell cycles.
- Analyzed 3000 "most variable ORFs" and normalized data by subtracting the mean expression level value across all time points for each gene. $$Y_{ij} = \frac{X_{ij} - \langle X_i \rangle}{[\frac{1}{15}\sum_{j=1}^{15}(X_{ij} - \langle X_i \rangle)^2]^{\frac{1}{2}}}$$
- Clustered genes by expression pattern using Euclidean distance metric values in the k-means algorithm



Used AlignACE to align 600bp upstream regions in order to determined common nucleotide motifs.

Results

- Found 18 motifs in 12 different clusters
  - Seven characterized transcription factor binding sites that are known to regulate many of the genes in their respective cluster.
- Clusters with known regulons have cis-regulatory elements emerged as the highest scoring motif in every case.
  - examples include MCB box and SCB cell-cycle box.
- Motifs that have not been previously described demonstrate strong correlation with clusters that are enrichment for genes with specific functions.
  - Cluster 3 motifs M3a and M3b and their association with RNA and translation related genes within and outside of cluster 3.
- "Half of the 30 clusters were significantly enriched for functional categories or had significant motifs."