

# Signaling, Microarrays, and Annotations

Michael Ochs

Information Science and Technology, Fox Chase Cancer Center  
School of Biomedical Engineering, Drexel University

# Microarrays and Biology

- Models by Physics
- Bayesian Decomposition - An Approach to Solve the Problem
- Results from Deletion Mutant Data

# What a Model Means to Me



Click here™ and a GIF decompressor are needed to see this picture.



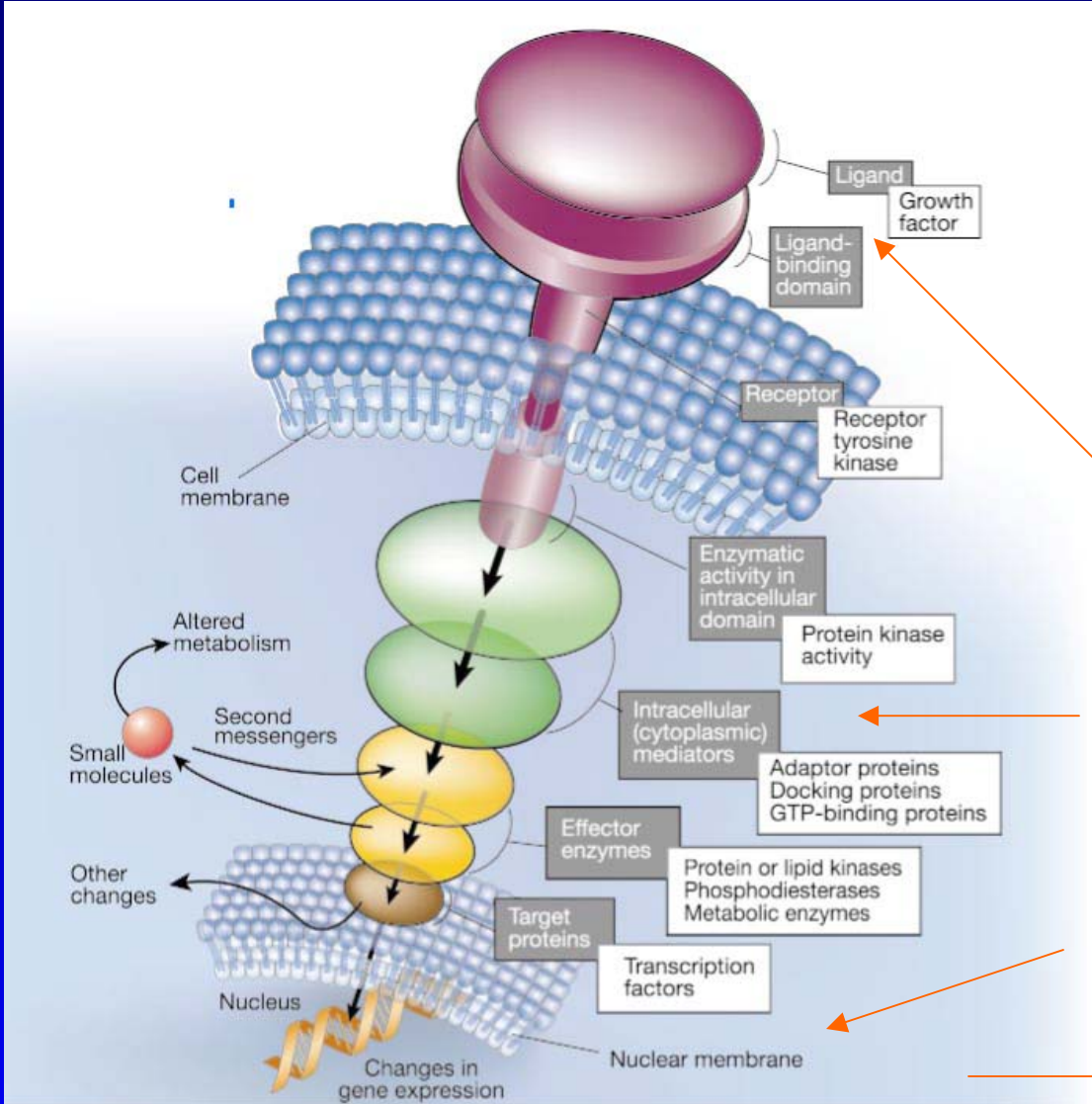
# Signalling Pathways

Stimulus

Signal Transduction

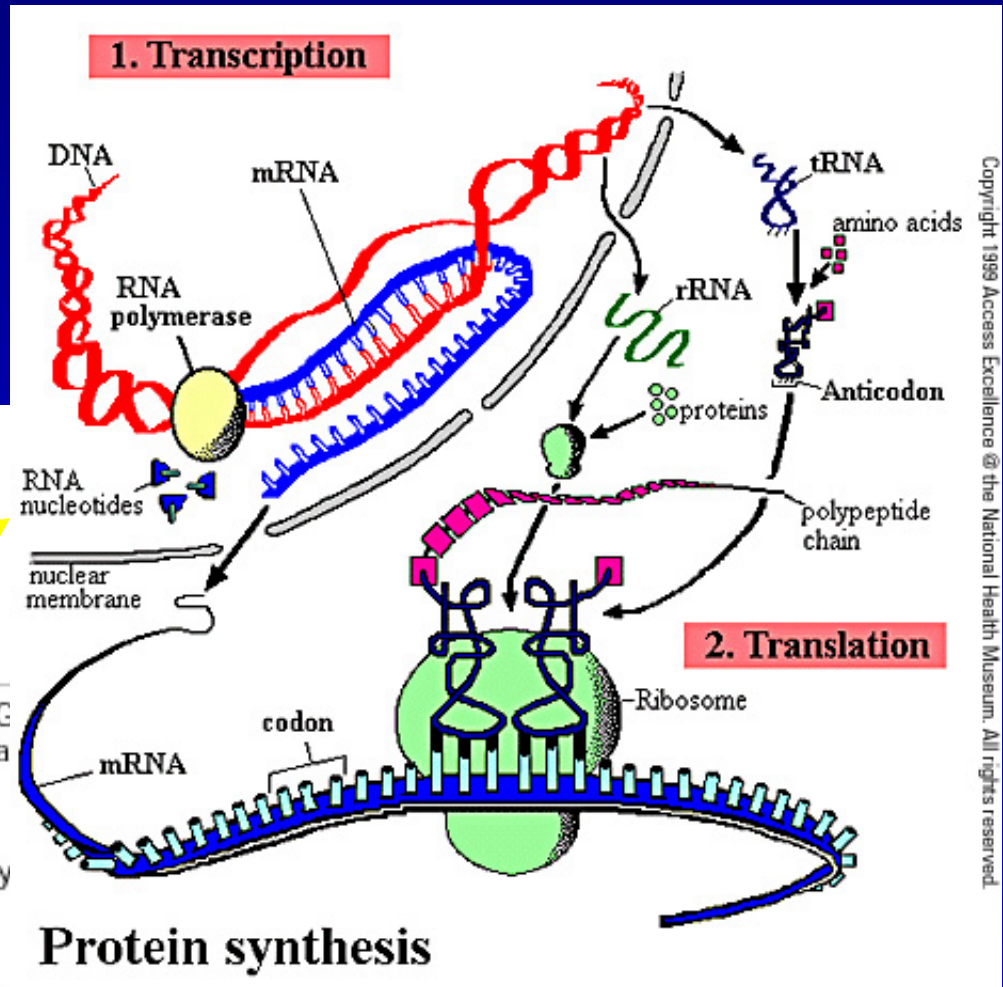
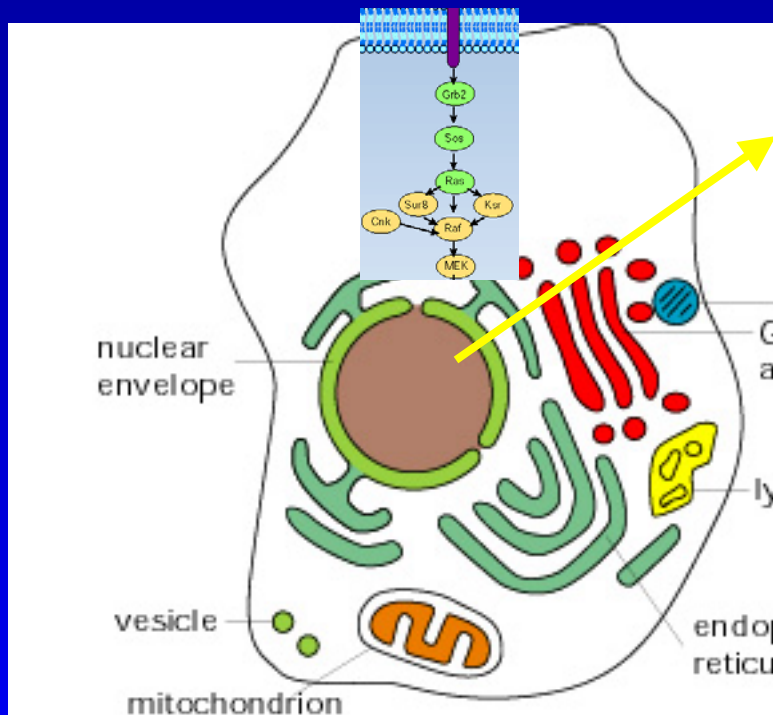
Transcription

mRNA



Downward, *Nature*, 411, 759, 2001

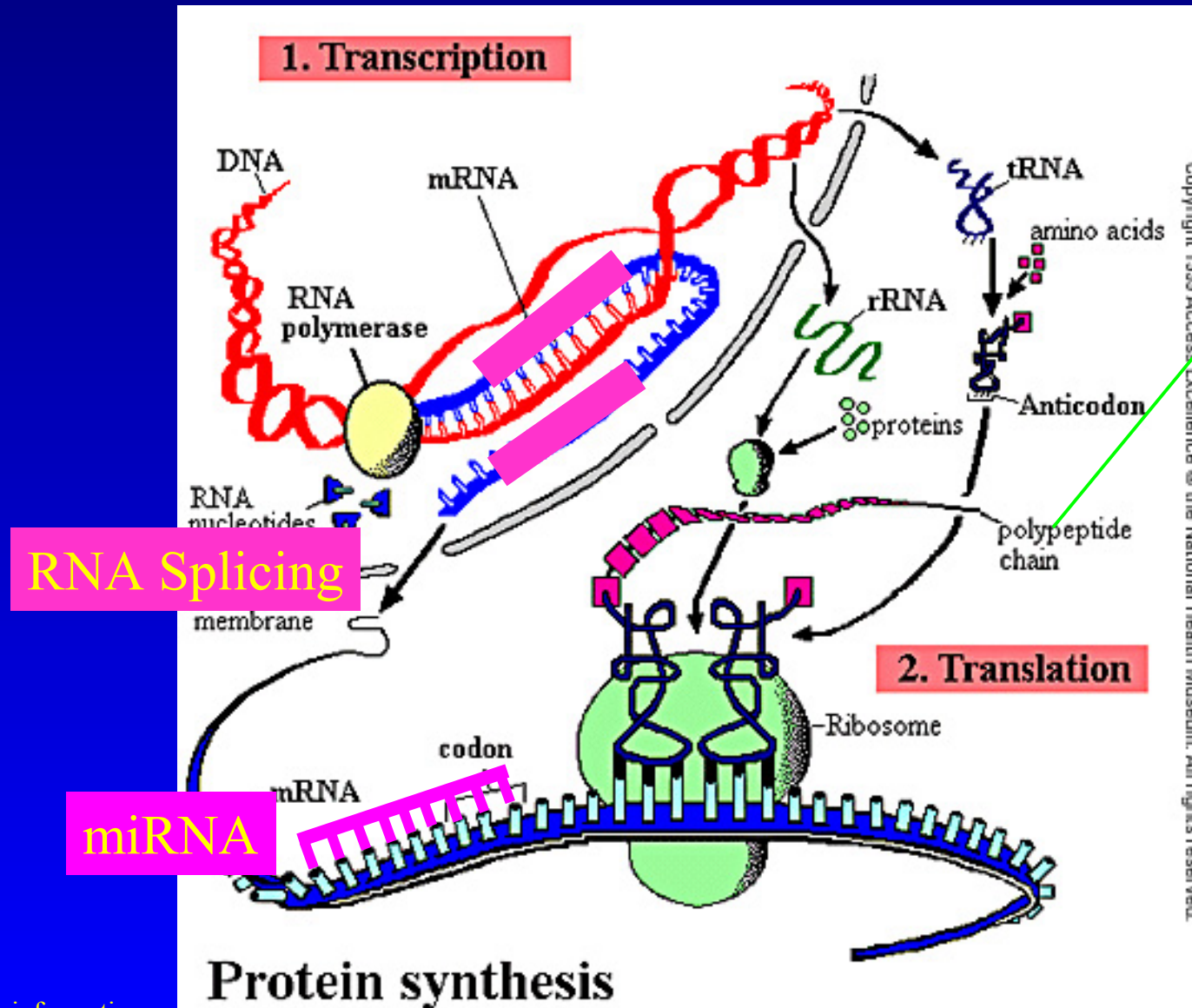
# Making Proteins



Copyright 1999 Access Excellence @ the National Health Museum. All rights reserved.

©1999 by Alberts, Bray, Johnson, Lewis, Raff, Roberts, Waller. <http://www.essentialcellbiology.com>  
 Published by Garland Publishing, a member of the Taylor & Francis Group.

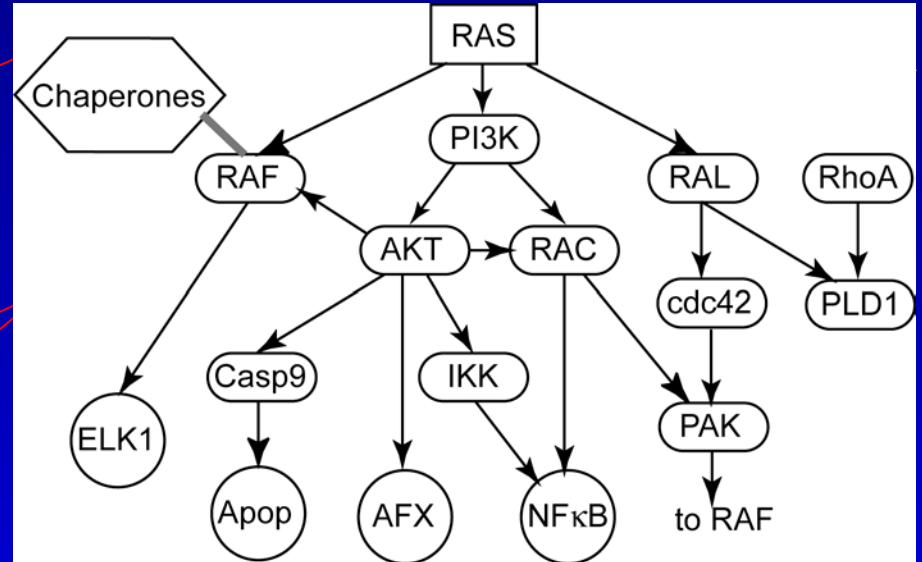
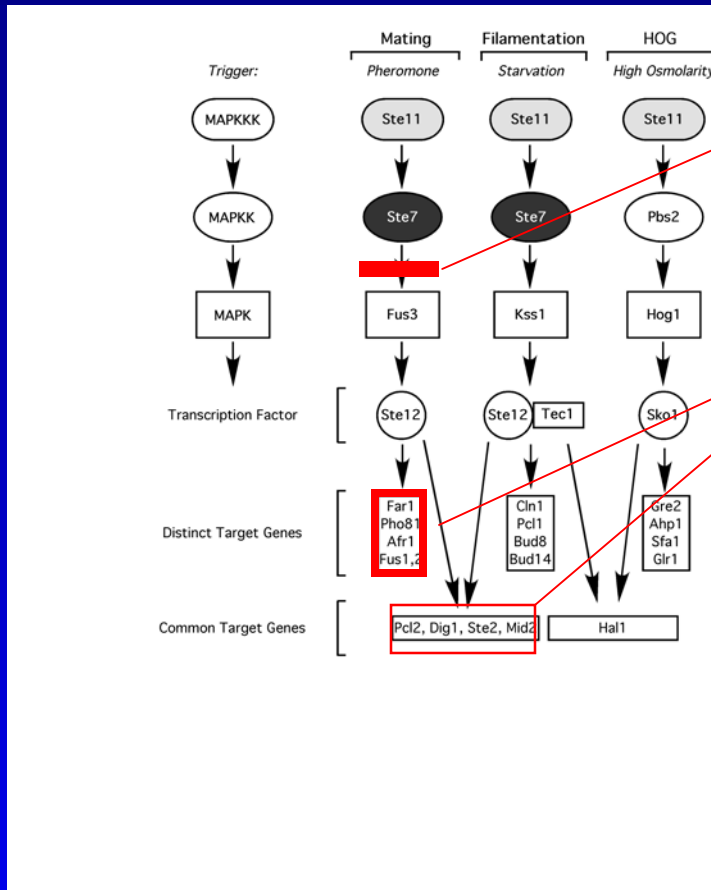
# A Closer Look at Translation



Copyright 1999 Access Excellence @ the National Health Museum. All rights reserved.

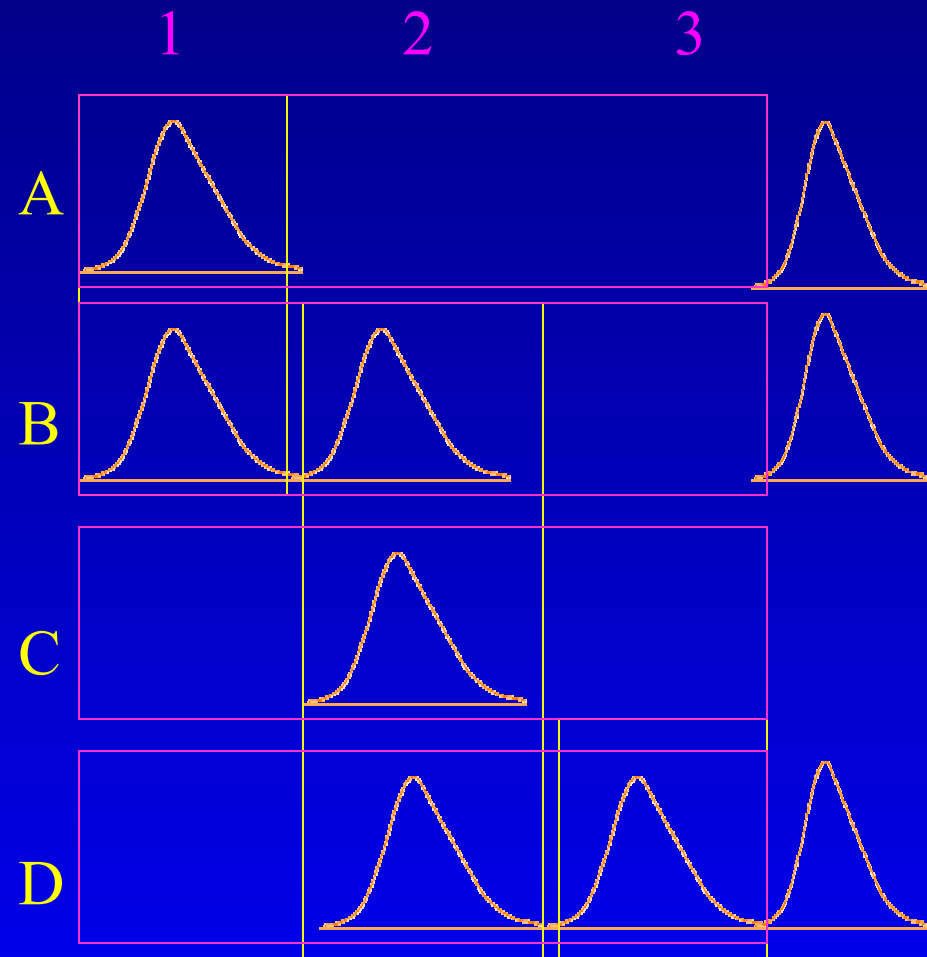
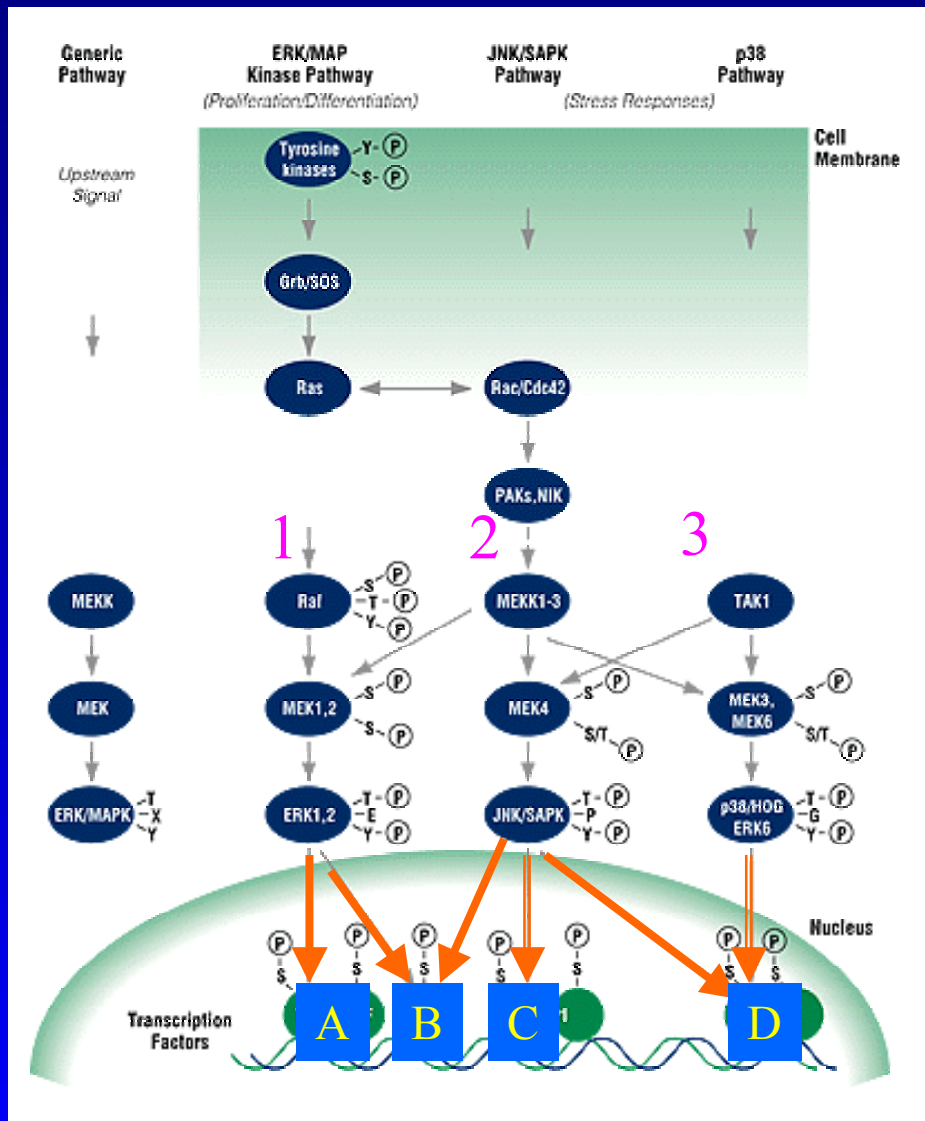
Post-Translational Modification

# Model



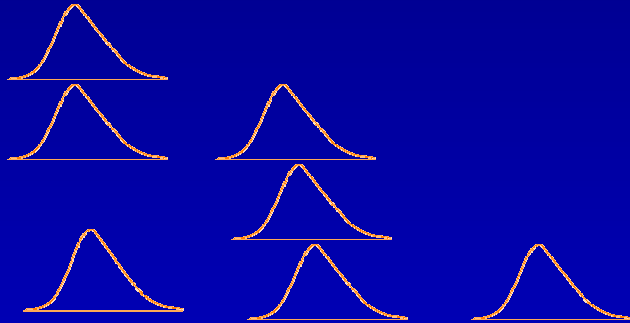
**But the Gene Lists are Incomplete as are the Network Diagrams!**

# Identifying Pathways



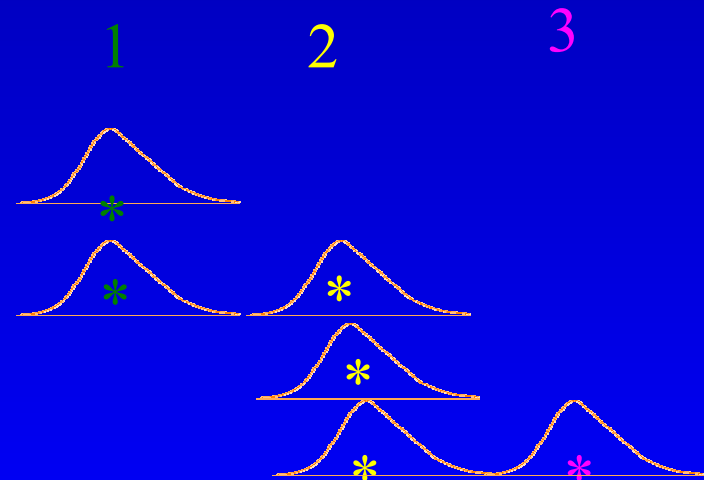
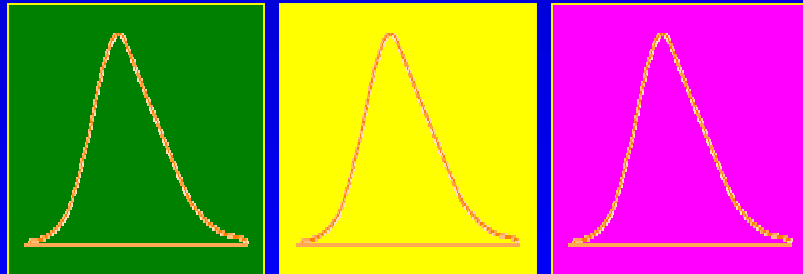
www.promega.com

# Goal of Analysis



Take measurements of thousands of genes, some of which are responding to stimuli of interest

And find the correct set of basis vectors that link to pathways

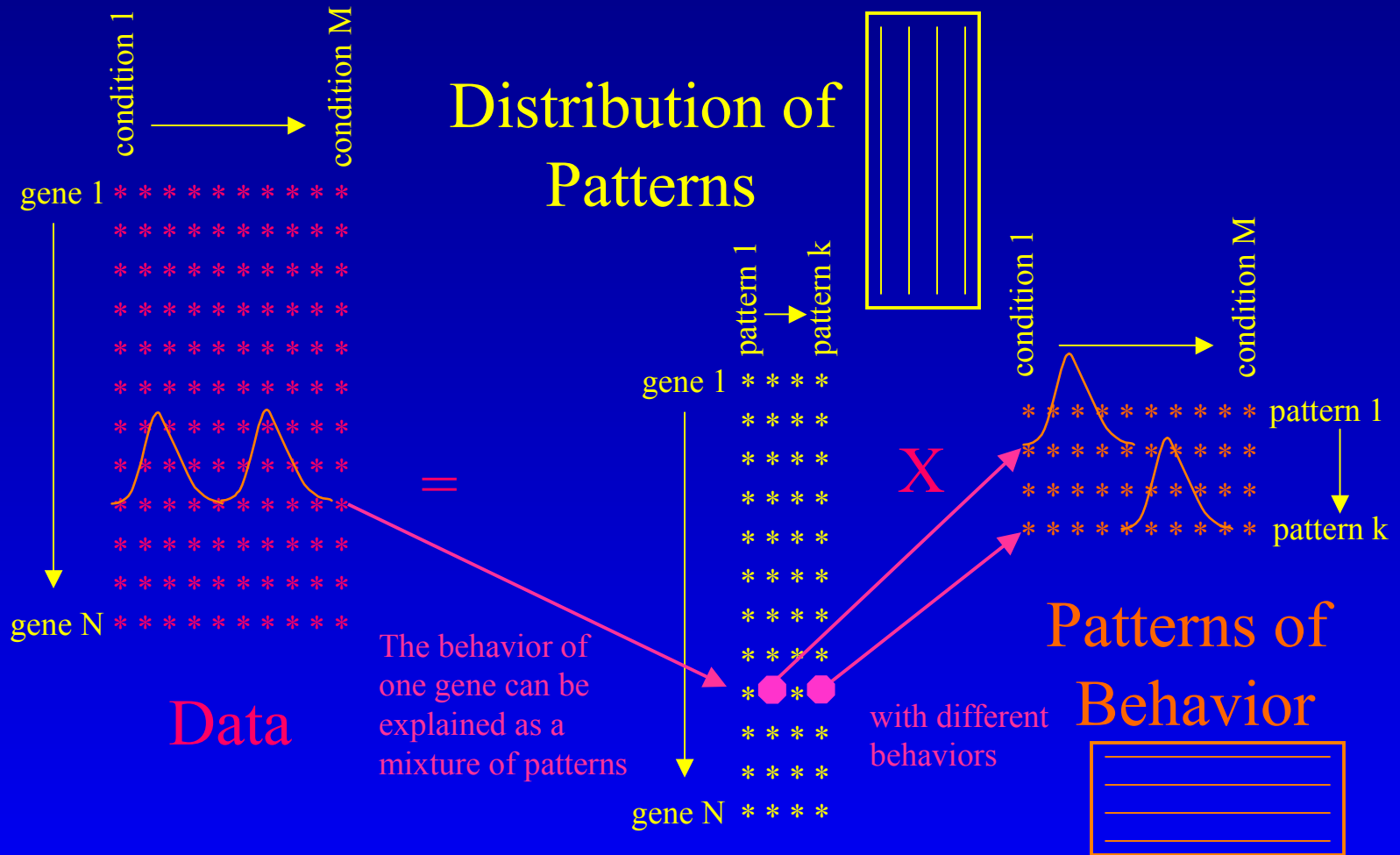


then identify the pathways

# Microarrays and Biology

- Models by Physics
- Bayesian Decomposition - An Approach to Perform Analysis
- Results from Deletion Mutant Data

# BD: Matrix Decomposition

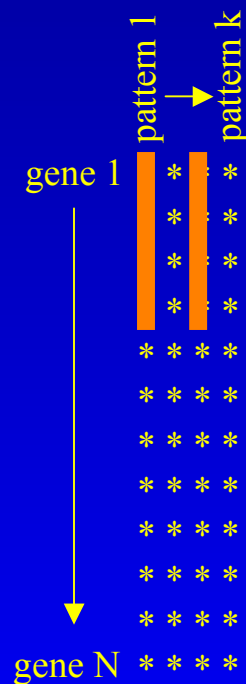


# The Model

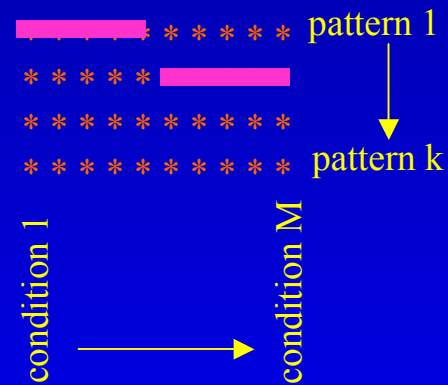
- Pathways Linked to Multiply Regulated Genes
- Positivity (No Negative Expression)
- Classification
  - Group 1 is Tumor
  - Group 2 is Normal
- Regulation
  - Genes Regulated by a Single Transcription Factor
  - Genes Known to be Coregulated (e.g., ribosomal proteins)

# Correlations and Biology

## Distribution of Patterns



## Patterns of Behavior



# Microarrays and Biology

- Models by Physics
- Bayesian Decomposition - An Approach to Perform Analysis
- Results from Deletion Mutant Data

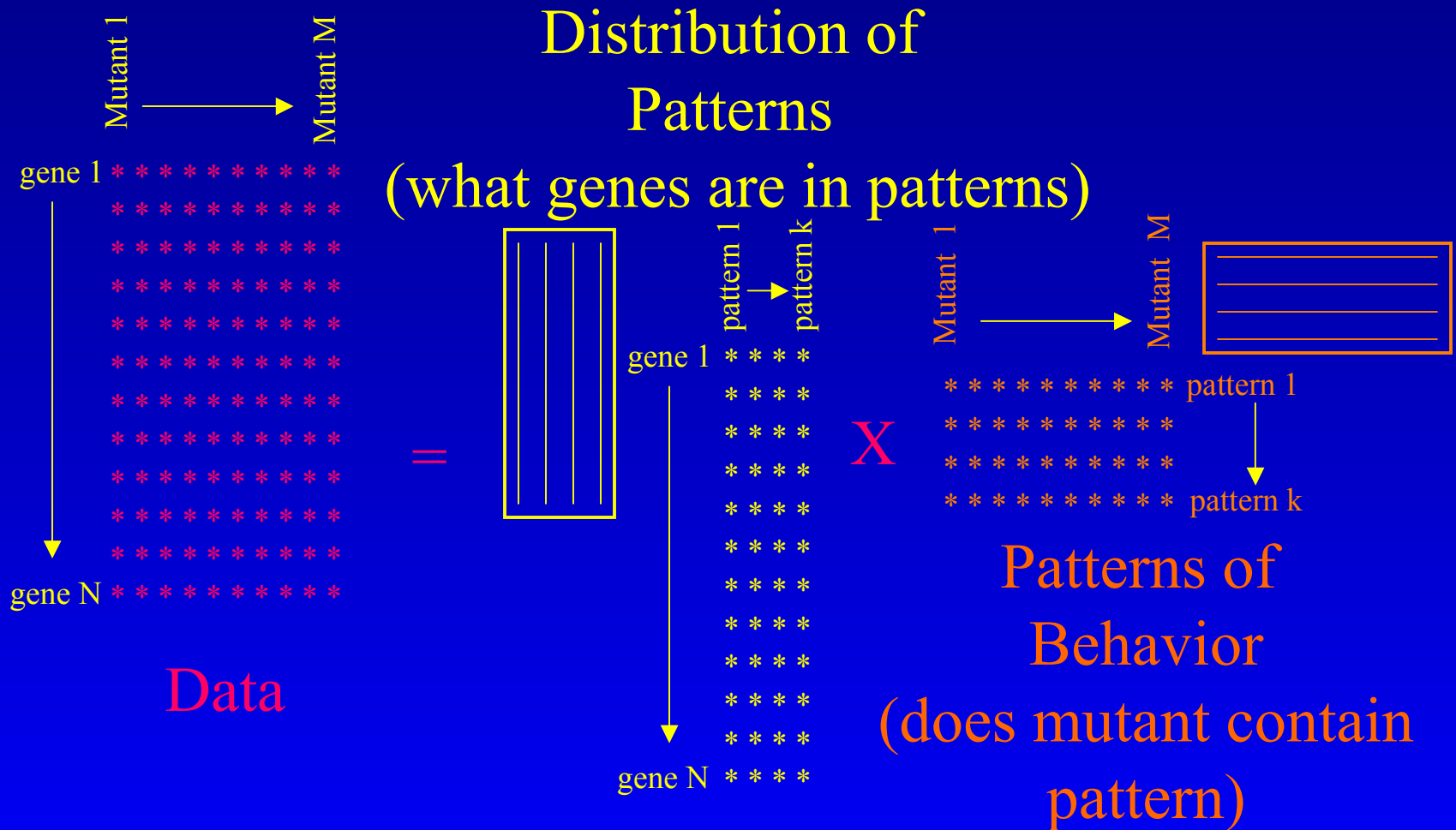
# Deletion Mutant Data Set

(Hughes *et al*, Cell, 102, 109, 2000)

- 300 Deletion Mutants in *S. cerevisiae*
  - Biological/Technical Replicates with Gene Specific Error Model
  - Filter Genes
    - >25% Data Missing in Ratios or Uncertainties
    - < 2 Experiments with 3 Fold Change
  - Filter Experiments
    - < 2 Genes Changing by 3 Fold

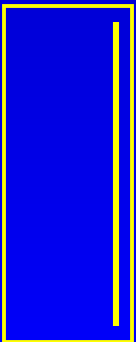
228 Experiments/764 Genes

# BD: Matrix Decomposition



# Genes in Patterns

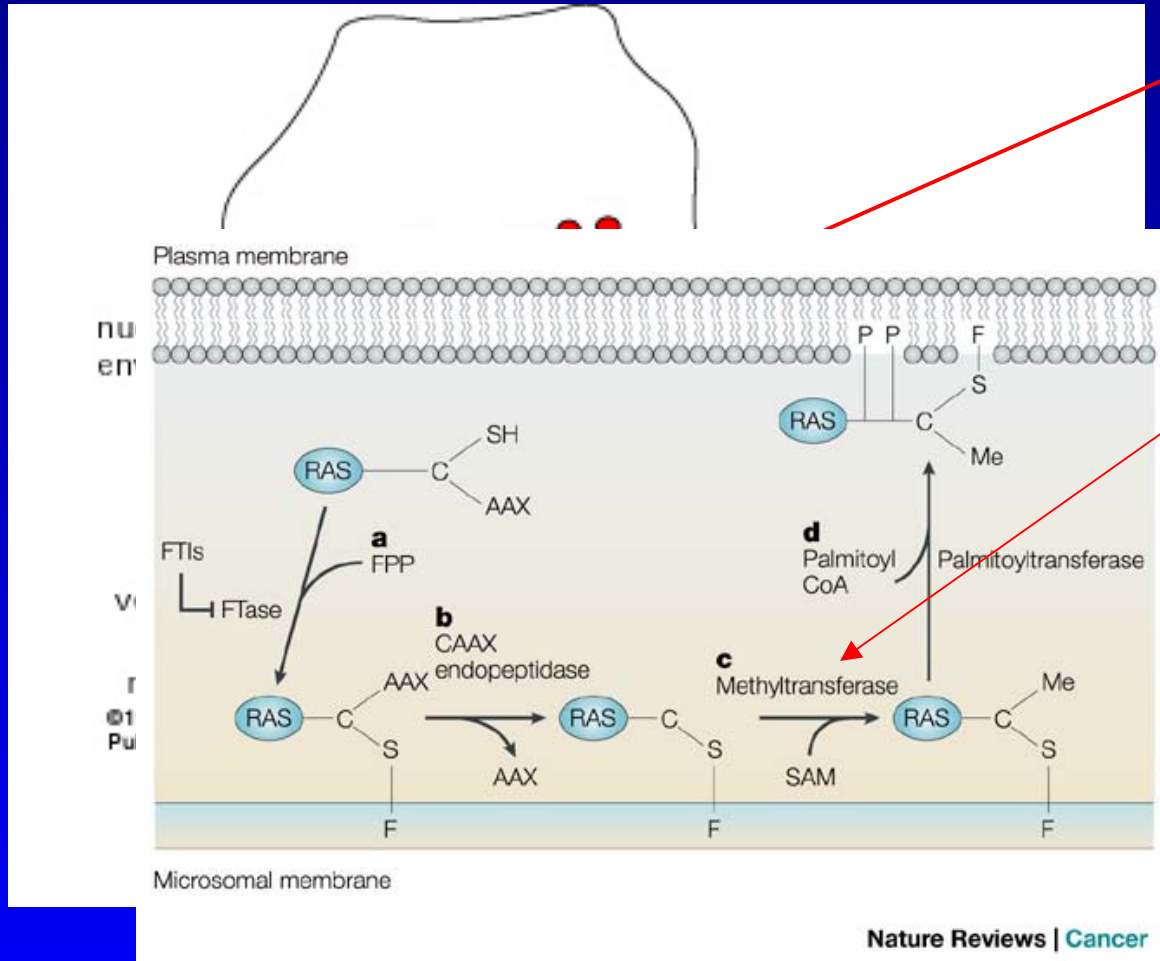
- Pattern 1
  - 403 Genes
- Pattern 2
  - 410 Genes
- Pattern 3
  - 390 Genes
- Pattern 4
  - 276 Genes
- Pattern 5
  - 355 Genes
- Pattern 6
  - 297 Genes
- Pattern 7
  - 223 Genes



# Annotating Genes

- Goals Being Left Behind
  - Identifying a List of Differentially Expressed Genes
  - Discriminating Classes
- Goals Now of Interest
  - Identifying Changes in Pathways
  - Identifying Active Biological Processes
  - Identifying Active Biological Functions

# Gene Ontology



Location

Function

Process

# Those are all PROTEINS!

- ESTs and Oligonucleotides
  - Short Sequences, Not Proteins, Not Genes
  - Need to Link these to Genes
- Clustering Sequences
  - UNIGENE/LocusLink
  - TIGR Gene Indices
  - BLAST
- Annotating Genes
  - Experimental
  - Computational

# UNIGENE

- Take ESTs, Align Together

- EST

- ~400 nucleotides

- Mismatch Allowed Reasonably High

- 123,995 “Genes”

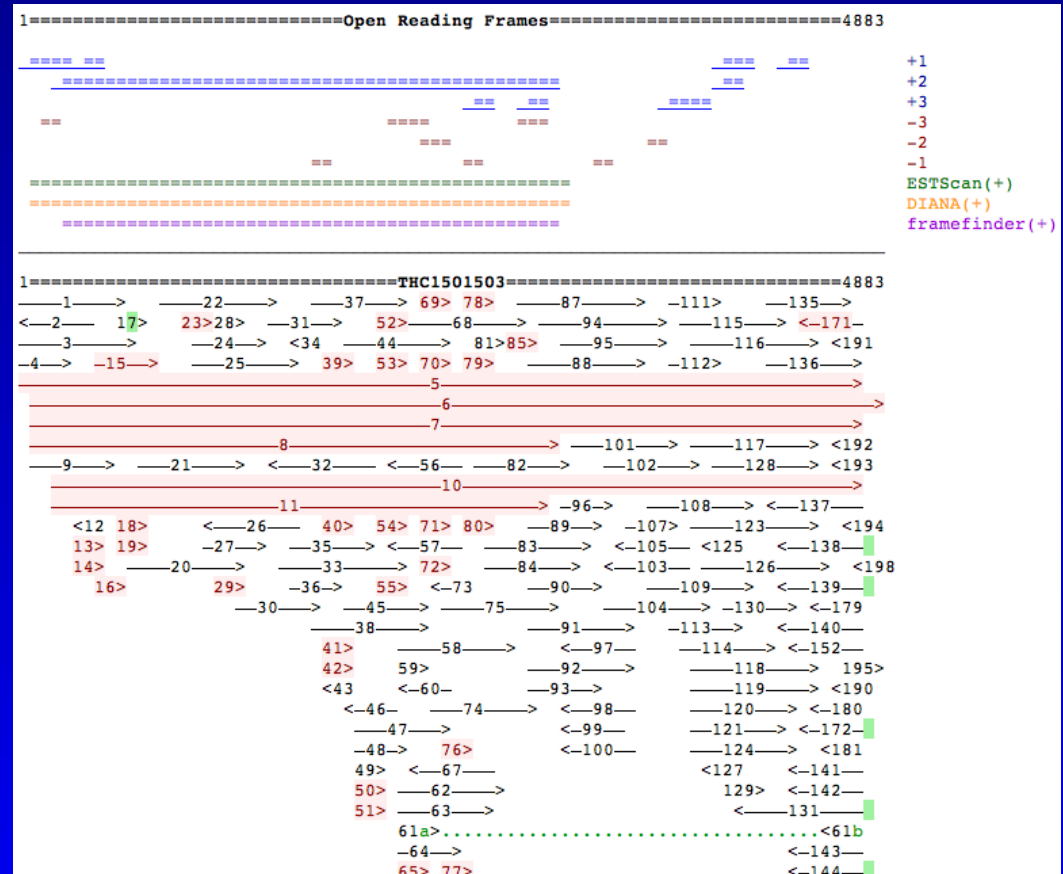
- ~10,000 Experimental Genes

- ~few thousand Estimated Genes

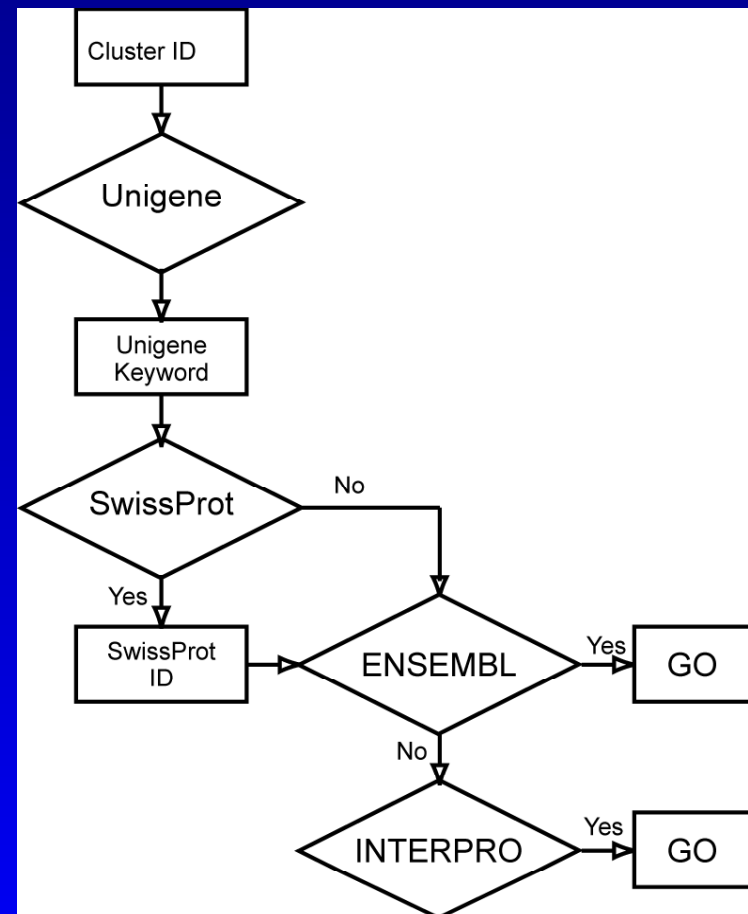
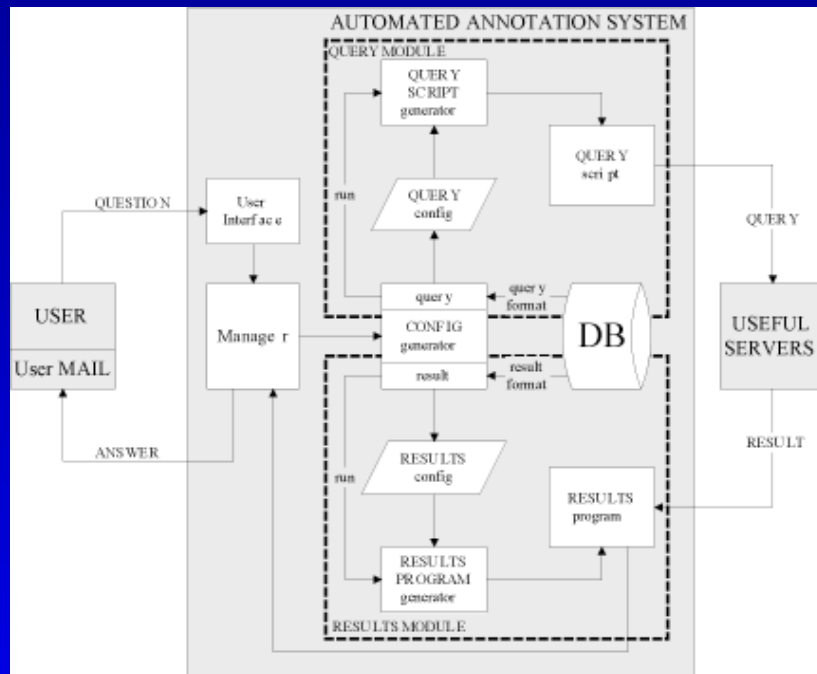
EST SEQUENCES (206)			
<a href="#">CB988872.1</a>	cDNA clone IMAGE:30340819	Human Placenta	5' read M
<a href="#">BM802736.1</a>	cDNA clone IMAGE:5575023	embryonal carcinoma, cell line	5' read PM
<a href="#">BI553937.1</a>	cDNA clone IMAGE:5265145	hippocampus	5' read PM
<a href="#">BI460646.1</a>	cDNA clone IMAGE:5267110	testis	5' read M
<a href="#">BI562738.1</a>	cDNA clone IMAGE:5298790	testis	5' read PM
<a href="#">BI335137.1</a>	cDNA clone IMAGE:5142020	cervical carcinoma cell line	5' read PM
<a href="#">AI051878.1</a>	cDNA clone IMAGE:1659317		3' read 5.2 kb A
<a href="#">AI872006.1</a>	cDNA clone IMAGE:2439762	moderately-differentiated endometrial adenocarcinoma, 3 pooled tumors	3' read 3.2 kb A
<a href="#">AA765882.1</a>	cDNA clone IMAGE:1304914	germinal center B cell	3' read 2.4 kb P
<a href="#">AI476550.1</a>	cDNA clone IMAGE:2156845	moderately-differentiated adenocarcinoma	3' read 2.4 kb A
<a href="#">AI923933.1</a>	cDNA clone IMAGE:2451885	well-differentiated endometrial adenocarcinoma, 7 pooled tumors	3' read 2.3 kb A
<a href="#">AA576266.1</a>	cDNA clone IMAGE:1072813	breast tumor	3' read 2.3 kb A
<a href="#">AA532733.1</a>	cDNA clone IMAGE:993160	adrenal adenoma	3' read 2.2 kb A
<a href="#">AA806119.1</a>	cDNA clone IMAGE:1350743	germinal center B cell	3' read 2.2 kb A
<a href="#">AA594282.1</a>	cDNA clone IMAGE:1057969	colon tumor RER+	3' read 2.2 kb A
<a href="#">AA857282.1</a>	cDNA clone IMAGE:1474855	squamous cell carcinoma	3' read 2.1 kb A
<a href="#">AA805390.1</a>	cDNA clone IMAGE:1334224	germinal center B cell	3' read 2.1 kb A
<a href="#">AA814509.1</a>	cDNA clone IMAGE:1427238	substantia nigra	3' read 1.9 kb A
<a href="#">AI539104.1</a>	cDNA clone IMAGE:2205404	poorly-differentiated endometrial adenocarcinoma, 2	3' read 1.9 kb A

# TIGR

- Take ESTs, Align Together into TC
  - EST
  - ~400 nucleotides
  - Highly Restrictive Match
    - 40 bp, 90% match,
    - max 30 bp gap



# Annotating Genes



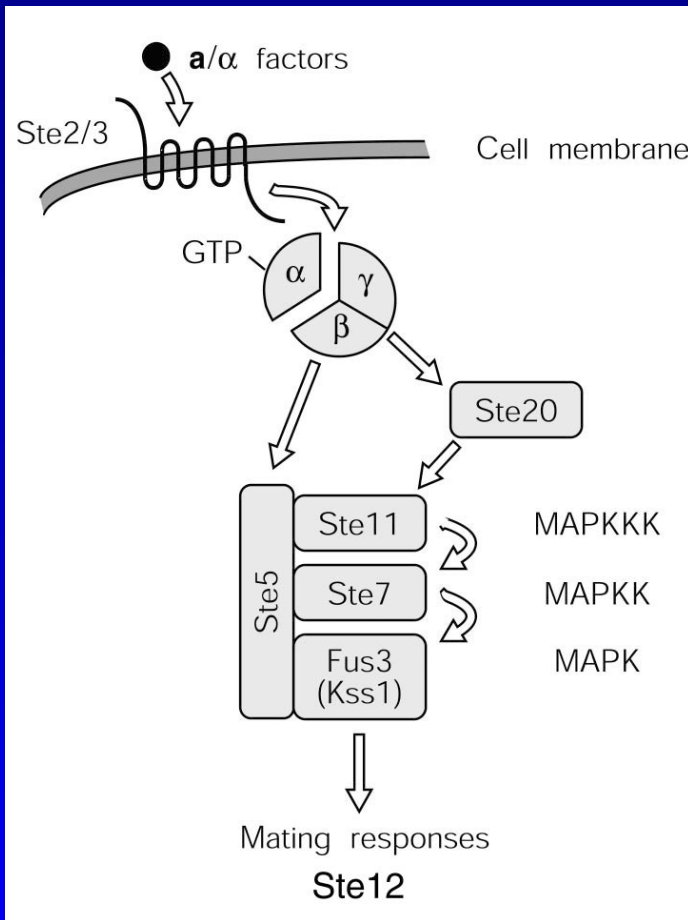
# Gene Ontology (Process)

$$\text{Enhancement} = \frac{N_{patt}^{GO} / N_{patt}^{TOTAL}}{N_{data}^{GO} / N_{data}^{TOTAL}}$$

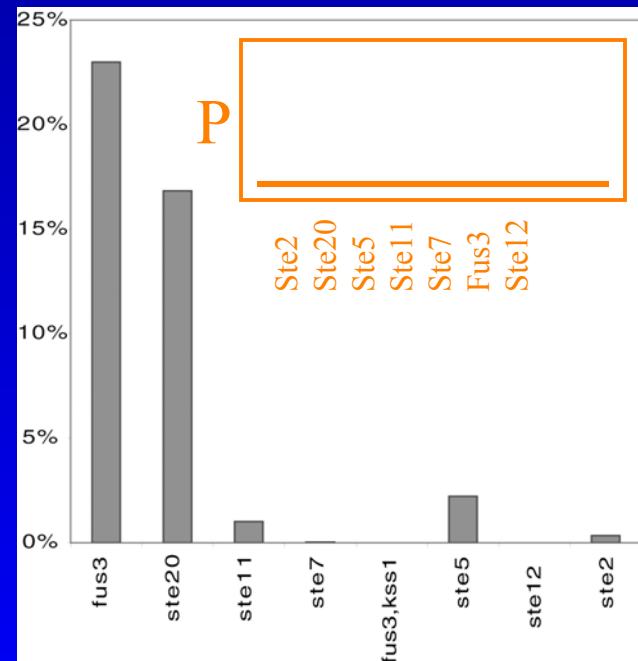
Class - GO	Inst...	P.1	P.2	P.3	P.4	P.5	P.6	P.7
755:cytogamy	2	0.0	0.0	0.0	4.41	0.0	0.97	1.03
6944:membrane fusion	6	0.0	0.0	0.0	4.41	0.0	1.61	1.03
45026:plasma membrane fusion	4	0.0	0.0	0.0	4.41	0.0	1.45	1.03
920:cell separation during cytokinesis	3	3.34	0.0	4.91	4.41	0.0	1.29	0.0
752:agglutination during conjugation with c...	2	0.0	1.19	0.0	4.41	0.0	0.0	2.05
771:agglutination	2	0.0	1.19	0.0	4.41	0.0	0.0	2.05
7157:heterophilic cell adhesion	2	0.0	1.19	0.0	4.41	0.0	0.0	2.05
15833:peptide transport	2	0.0	2.38	2.46	4.41	1.29	1.93	1.03
749:response to pheromone during conjug...	13	0.51	1.46	0.76	3.73	0.6	0.74	1.42
746:conjugation	25	0.53	1.05	0.79	3.52	0.52	0.93	1.39
747:conjugation with cellular fusion	25	0.53	1.05	0.79	3.52	0.52	0.93	1.39
19953:sexual reproduction	25	0.53	1.05	0.79	3.52	0.52	0.93	1.39
7600:sensory perception	15	0.45	1.27	0.65	3.52	0.52	0.77	1.5
7606:chemosensory perception	15	0.45	1.27	0.65	3.52	0.52	0.77	1.5
9581:perception of external stimulus	15	0.45	1.27	0.65	3.52	0.52	0.77	1.5
9582:perception of abiotic stimulus	15	0.45	1.27	0.65	3.52	0.52	0.77	1.5
9593:perception of chemical substance	15	0.45	1.27	0.65	3.52	0.52	0.77	1.5
19236:response to pheromone	15	0.45	1.27	0.65	3.52	0.52	0.77	1.5
3:reproduction	28	0.6	1.02	0.88	3.3	0.46	1.04	1.32
750:signal transduction during conjugation ...	8	0.84	1.78	1.23	3.3	0.65	0.97	1.54
7166:cell surface receptor linked signal tran...	9	1.11	1.85	1.09	2.94	0.86	1.07	1.37
7186:G-protein coupled receptor protein si...	9	1.11	1.85	1.09	2.94	0.86	1.07	1.37
16337:cell-cell adhesion	3	0.0	1.58	0.0	2.94	0.0	0.64	2.05
7275:development	41	0.65	1.16	0.96	2.58	0.57	1.18	1.35
6116:NADH oxidation	2	0.0	1.19	2.46	2.2	2.59	0.97	0.0
8272:sulfate transport	2	1.67	1.19	0.0	2.2	2.59	0.0	1.03
9628:response to abiotic stimulus	35	0.96	1.09	1.26	2.01	0.44	1.1	1.29
7127:meiosis I	9	1.11	0.79	1.64	1.96	0.29	0.86	0.91
9605:response to external stimulus	40	0.92	1.13	1.23	1.87	0.45	1.06	1.23



# Mating Response



Amount of Behavior Explained by Mating Pathway for Mutants



(Posas, *et al*, Curr Opin Microbiology, **1**, 175, 1998)

# Conclusions

- BD Identifies Patterns Related to Underlying Physiology
- BD Uses Prior Knowledge to Guide Data Analysis
- With Adequate Information, BD Links Expression Changes to Pathway Activity
- Proteomics, TF Binding Data, and Future Data Types are Easily Included

# Acknowledgements

## Fox Chase

Tom Moloshok

Jeffrey Grant

Yue Zhang

Elizabeth Goralczyk

Luke Somers

Michael Slifker

Ghislain Bidaut

Andrew Kossenkov

Vladimir Minayev

Garo Toby

Bill Speier (Johns Hopkins)

Daniel Chung

DJ Datta (UCSF)

## Collaborators

A. Godwin (FCCC)

B. Eisenberg (FCCC > Dartmouth)

J.-M. Claverie (CNRS)

G. Parmigiani (JHU)

E. Korotkov (RAS)

Frank Manion

Bob Beck