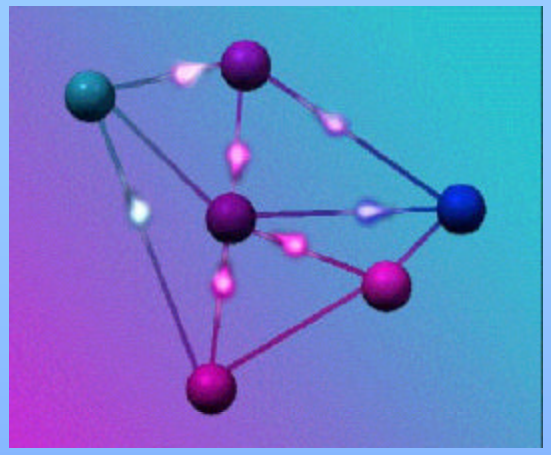


Institute for Pure and Applied Mathematics, UCLA

**Functional Genomics**

***Expression Arrays, Genetic Networks and Diseases***

November 8-12, 2000

# *From Genes to Dynamic Molecular Networks*

**Roland Somogyi, Ph.D.**

Molecular Mining Corporation

molecular mining

# Multigenic & pleiotropic regulation: the basis of genetic networks

**more realistic idealization**

**most complete model**

**simple idealization**

multigenic regulation

pleiotropic regulation

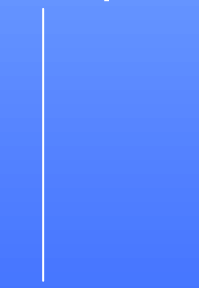
genetic network

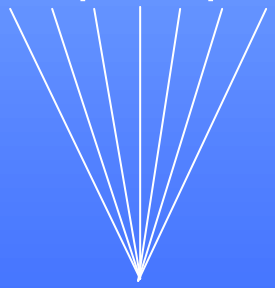
single input

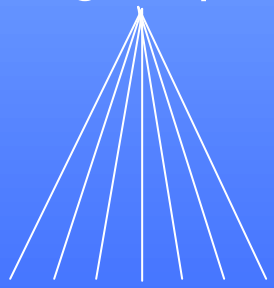
multiple inputs

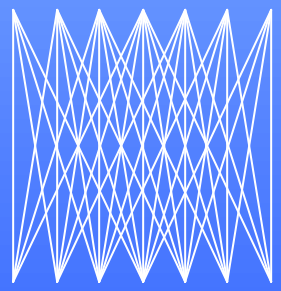
single input

multiple inputs









single output

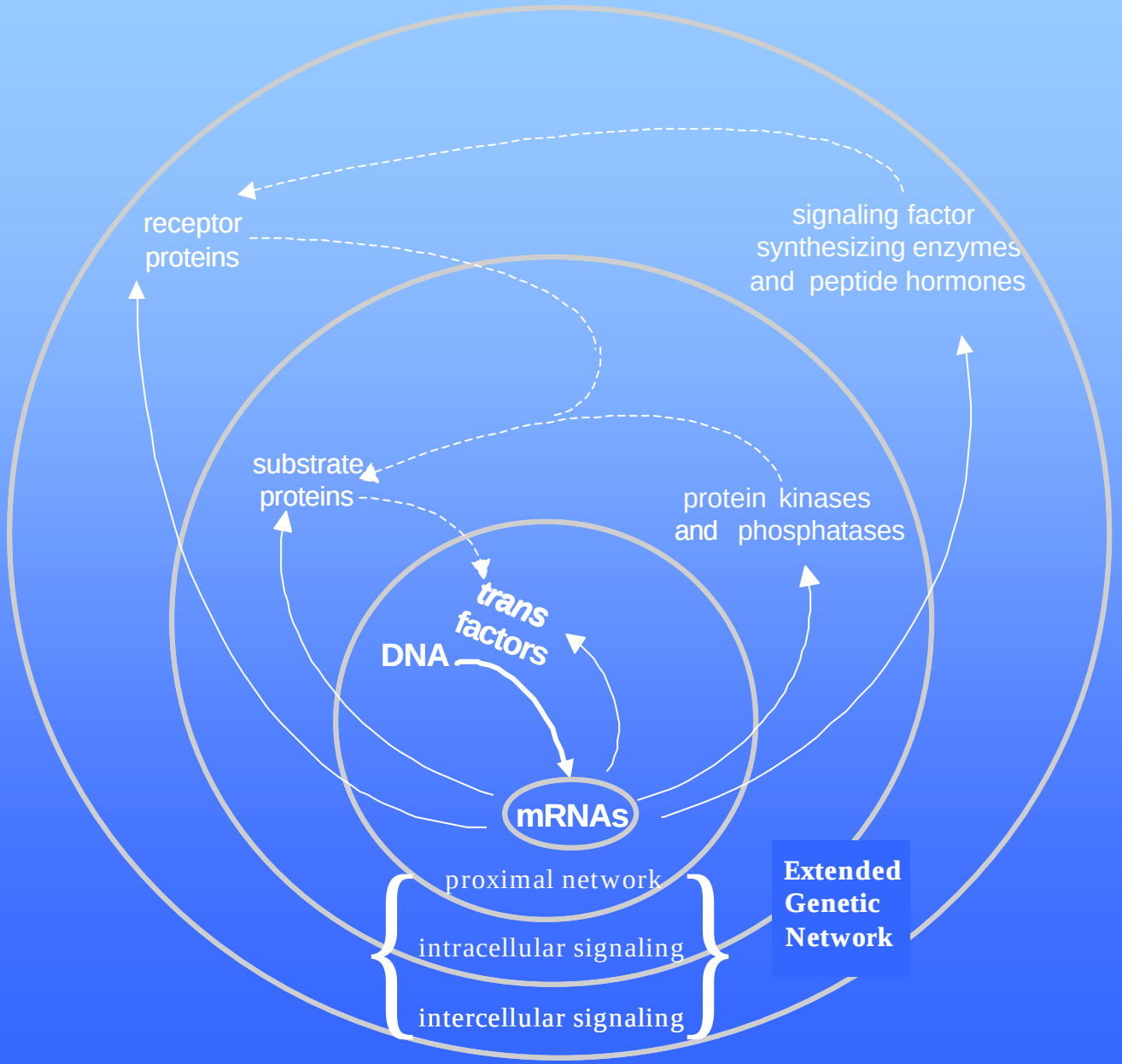
single output

multiple outputs

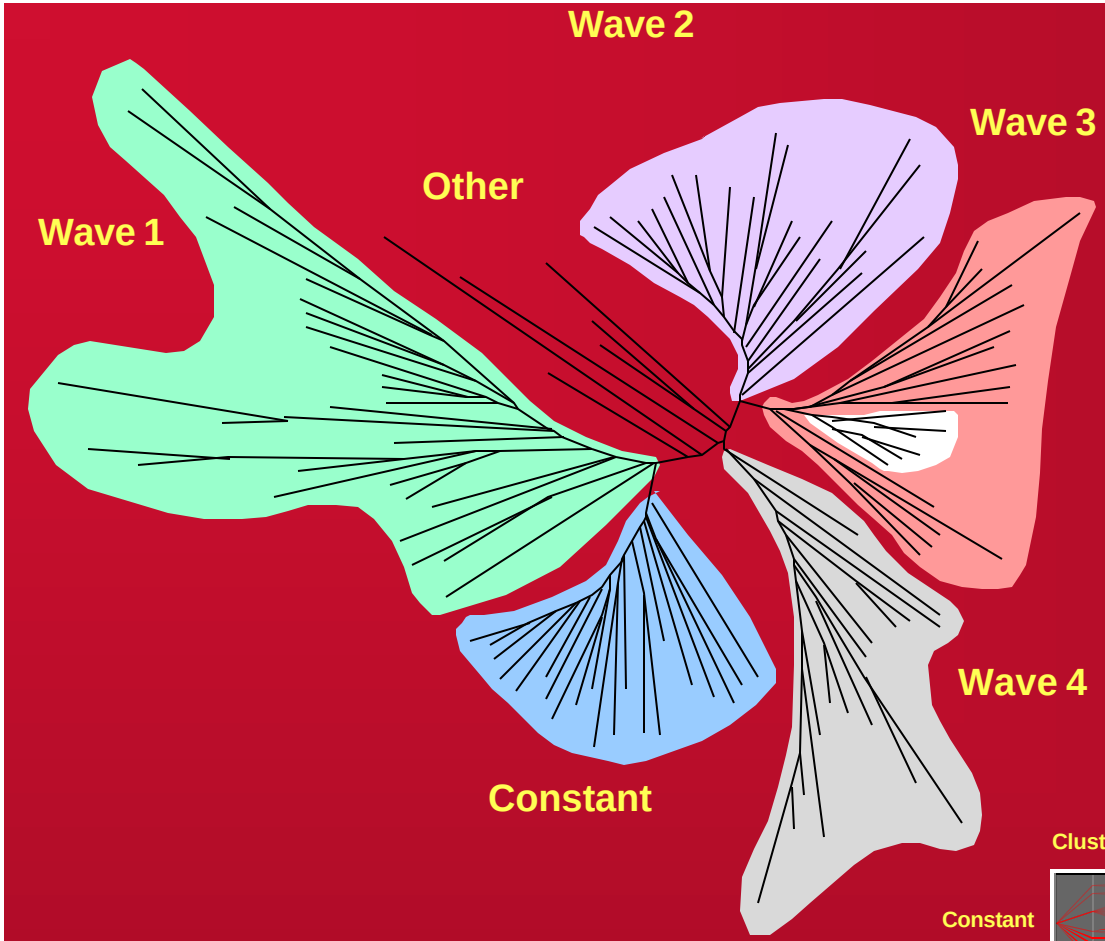
multiple outputs

molecular mining

# The Genetic Network



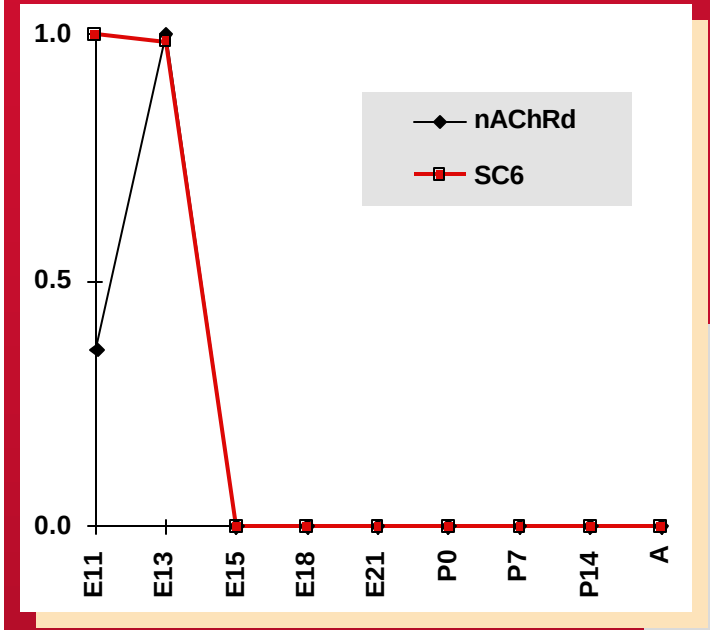

molecular mining

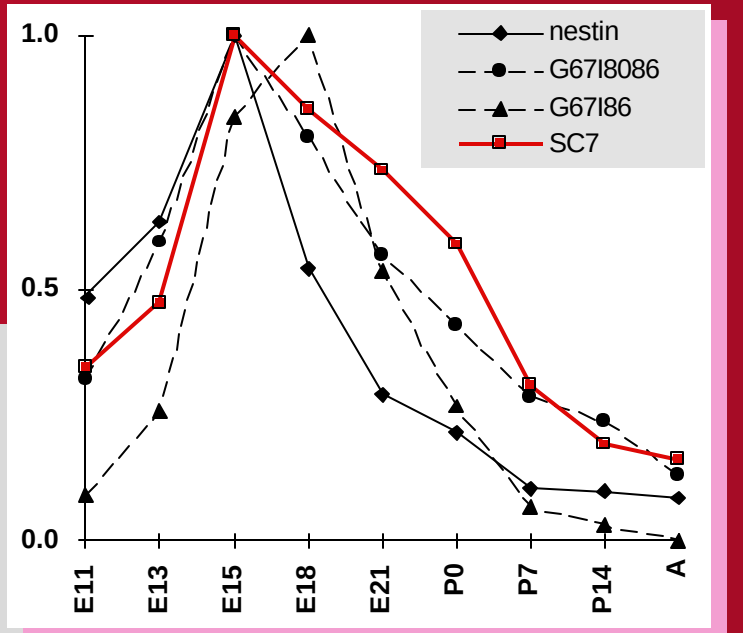


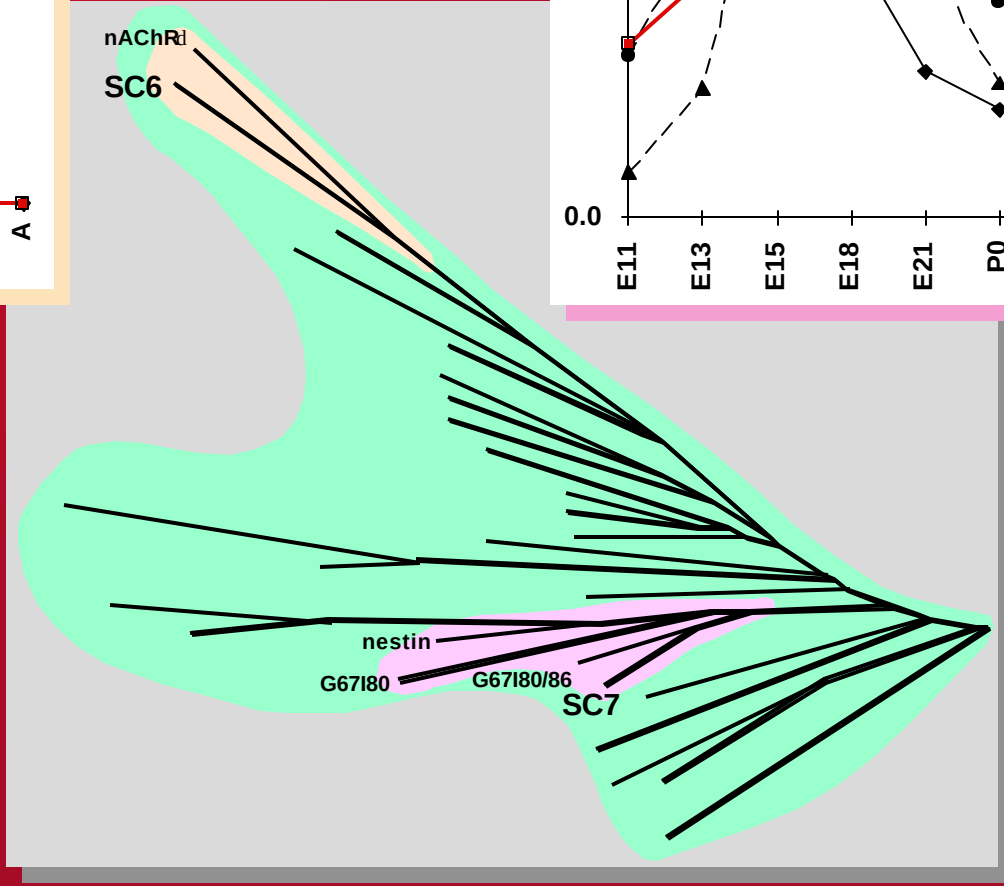

# Euclidean Cluster Analysis


molecular mining

# Clustering of Novel Genes in Wave 1



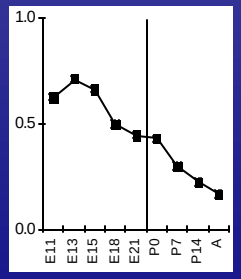



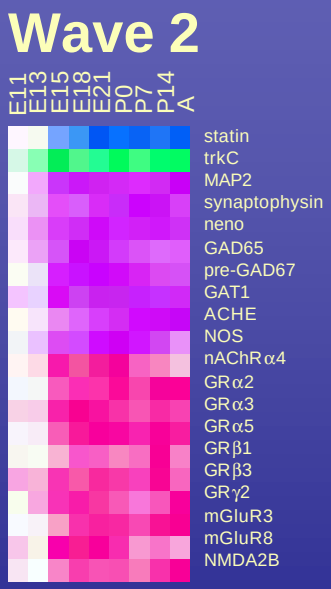

molecular mining

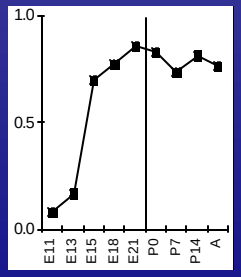
# Gene Expression Waves in the Developing Spinal Cord

## Wave 1

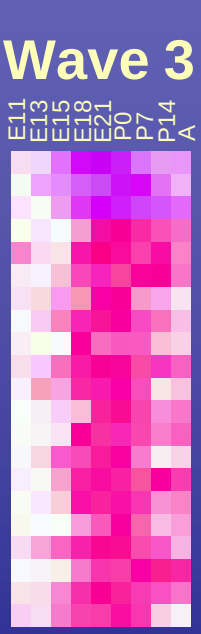



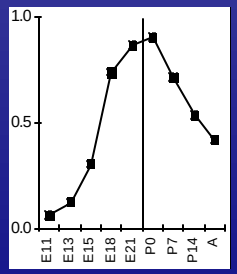

## Wave 2



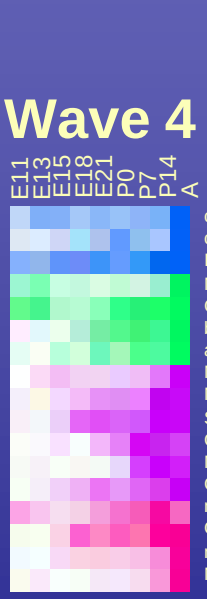



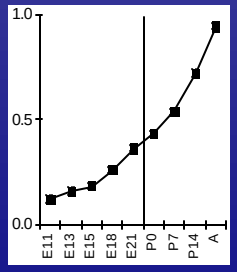

## Wave 3



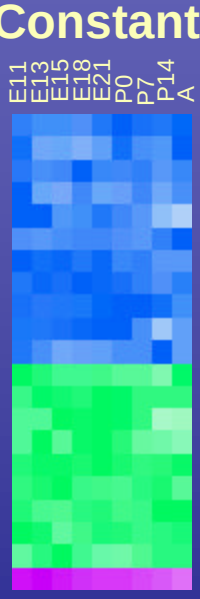



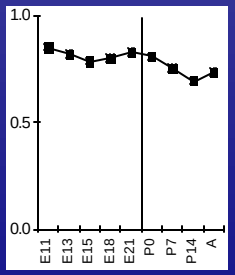

## Wave 4




## Constant



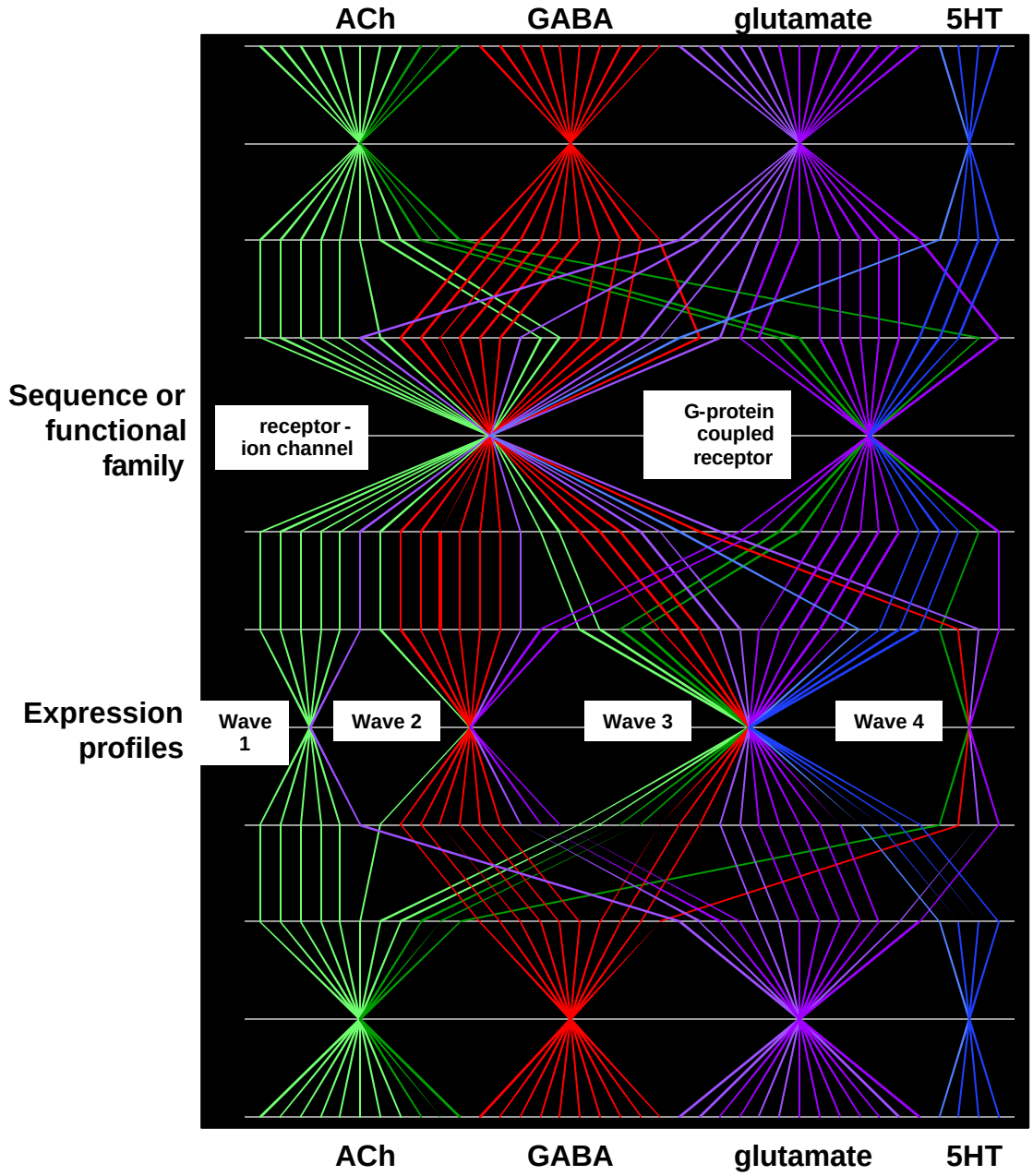




## Other



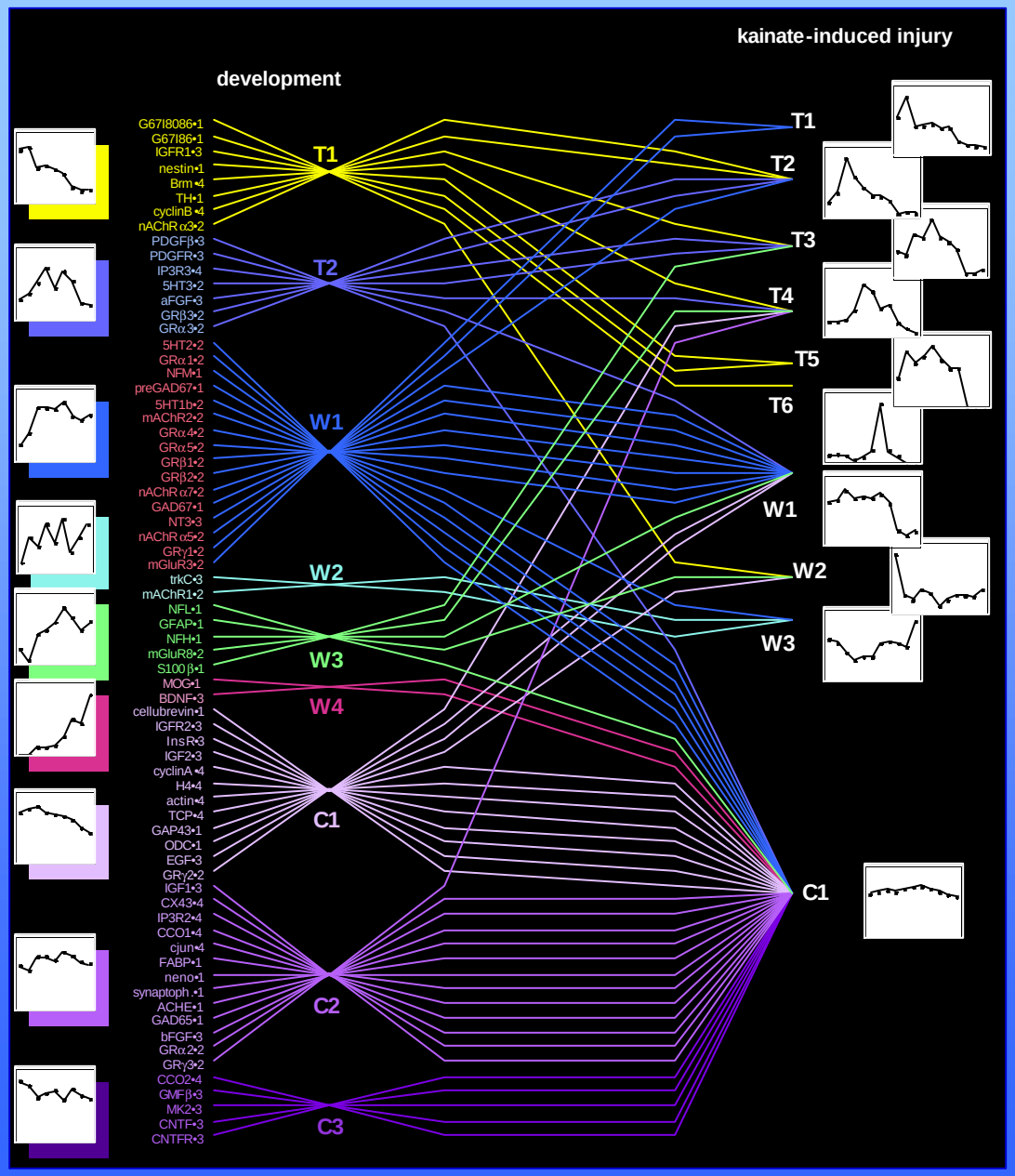
molecular mining

# Neurotransmitter receptor expression pathways in spinal cord development




molecular mining

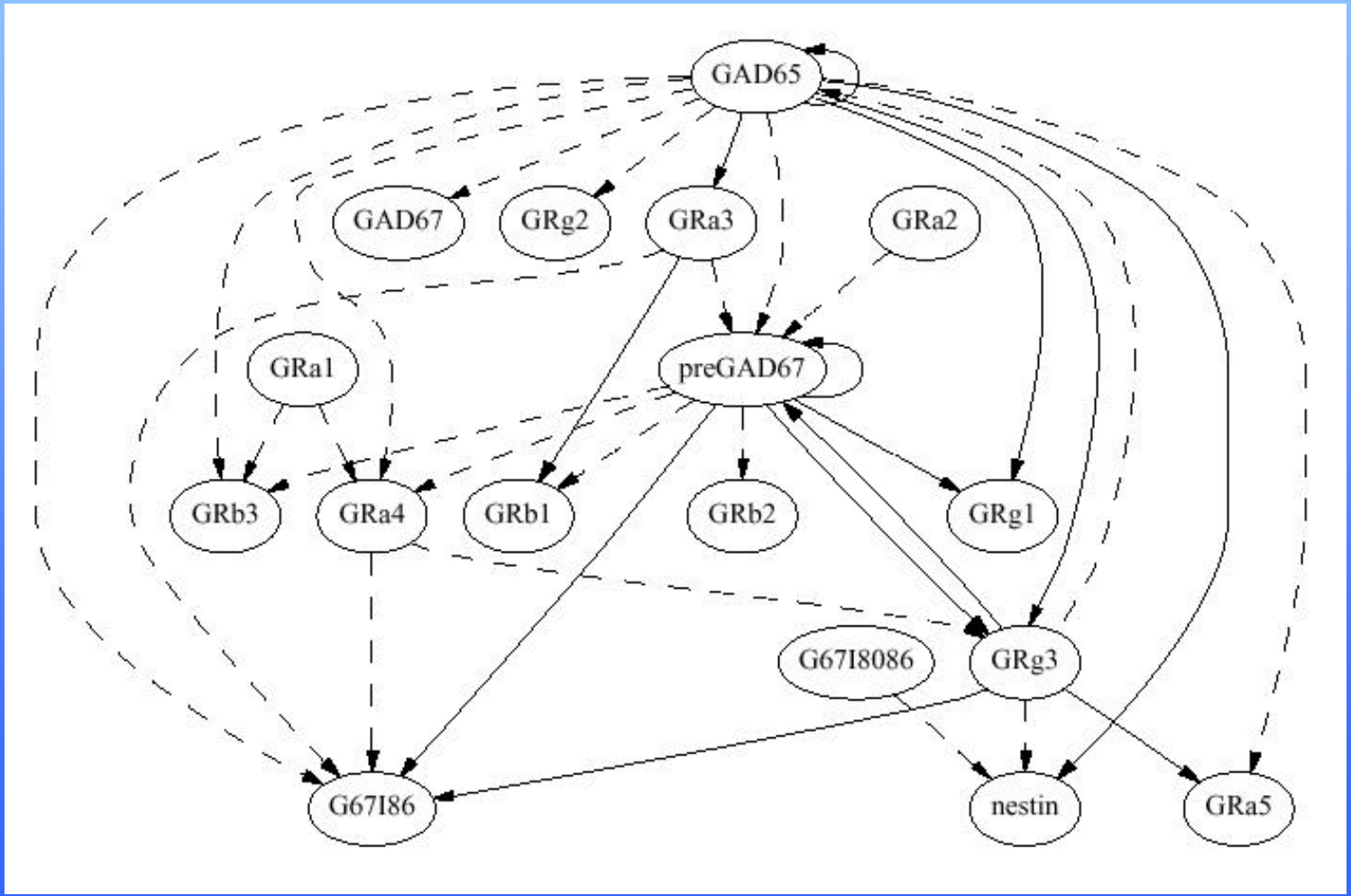
# Hippocampal injury perturbs developmentally regulated genes




molecular mining

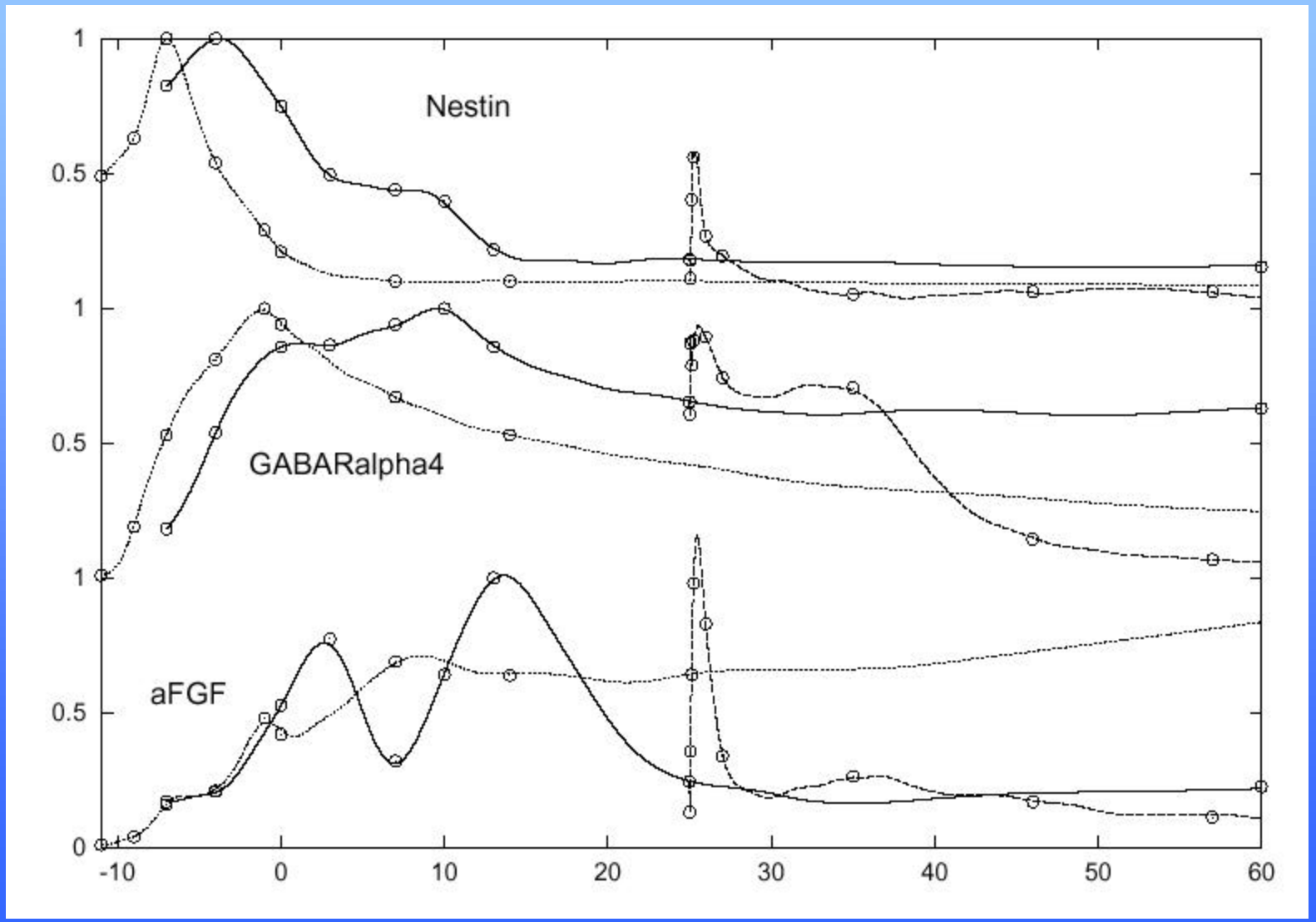
The Molecular Networks Company™

# Reverse engineering the GABAergic expression network



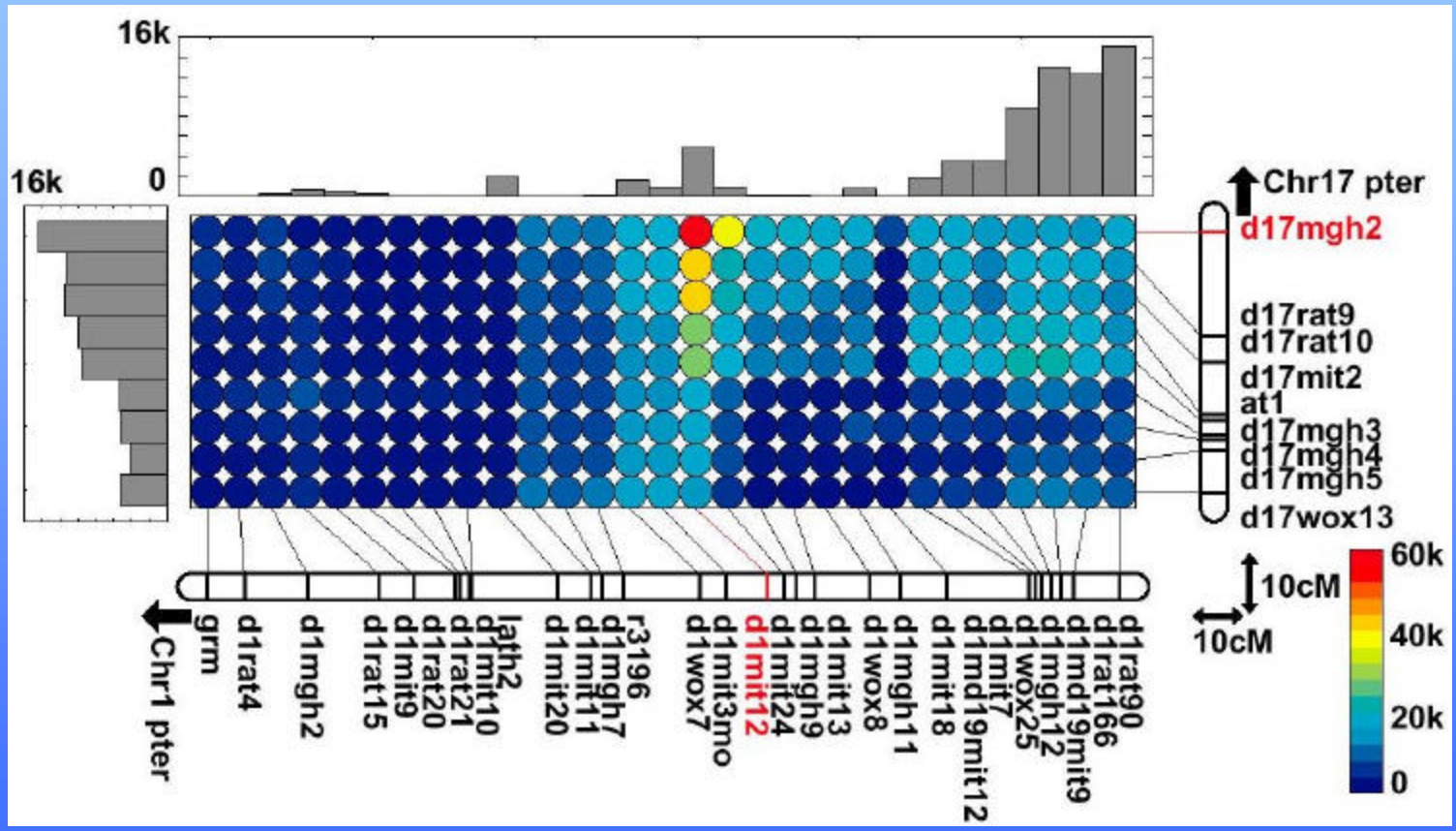

molecular mining

# Reconstruction of expression time series using a linear model




molecular mining

# Non-linear combinations of loci determine oral glucose tolerance in rat




molecular mining

From Nakaya *et al.*, PSB 2000

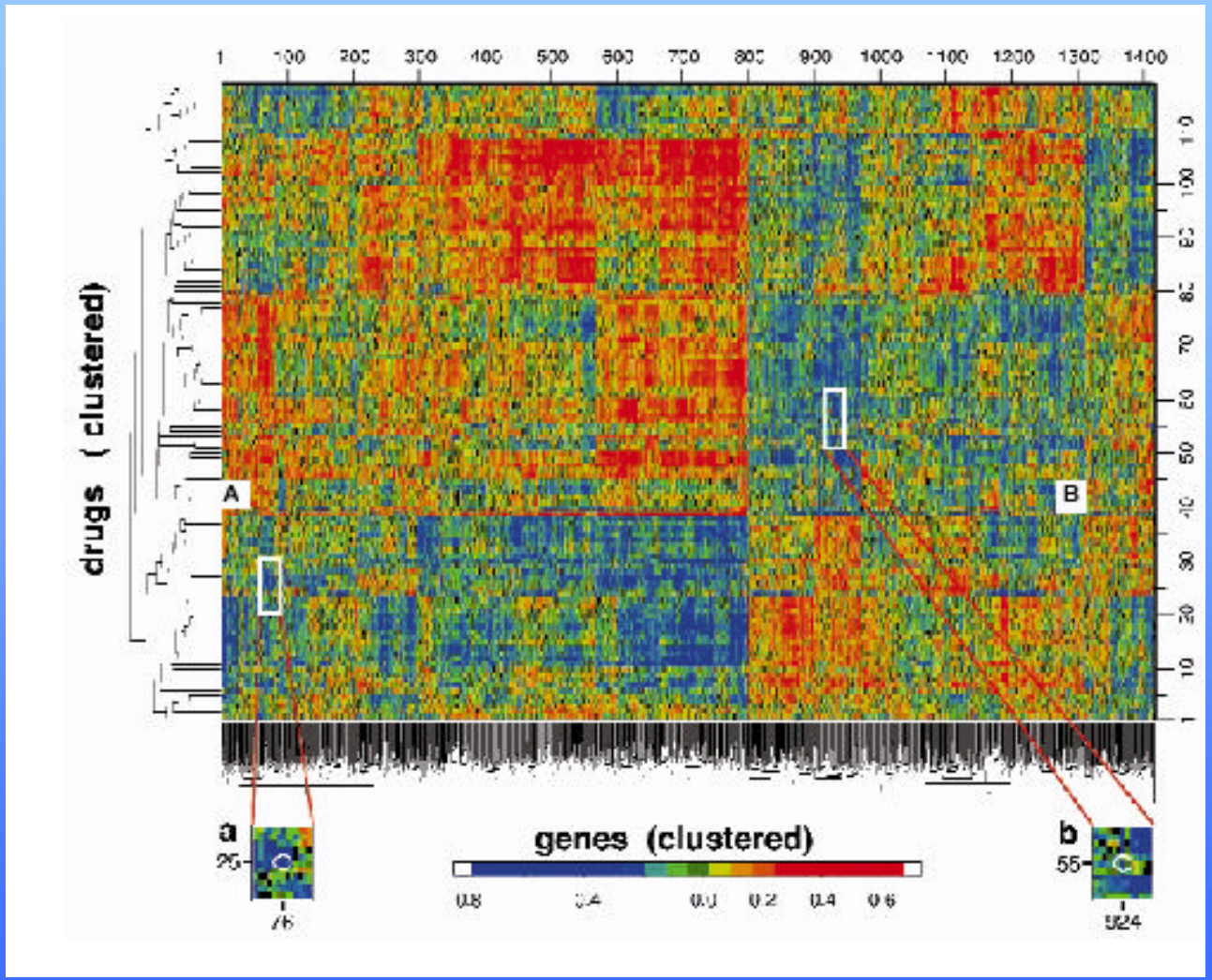
# NCI60:
## Exploring the molecular basis of cancer drug response

Data from John Weinstein of the National Cancer Institute

- 60 human cancer cell lines
- Growth inhibition response of all cell lines to 90 drugs
- Basal expression of 1400 genes for all cell lines

molecular mining

# Clustering of NCI60 According to Genes and Drugs




molecular mining

Source:

Scherf et al., 2000

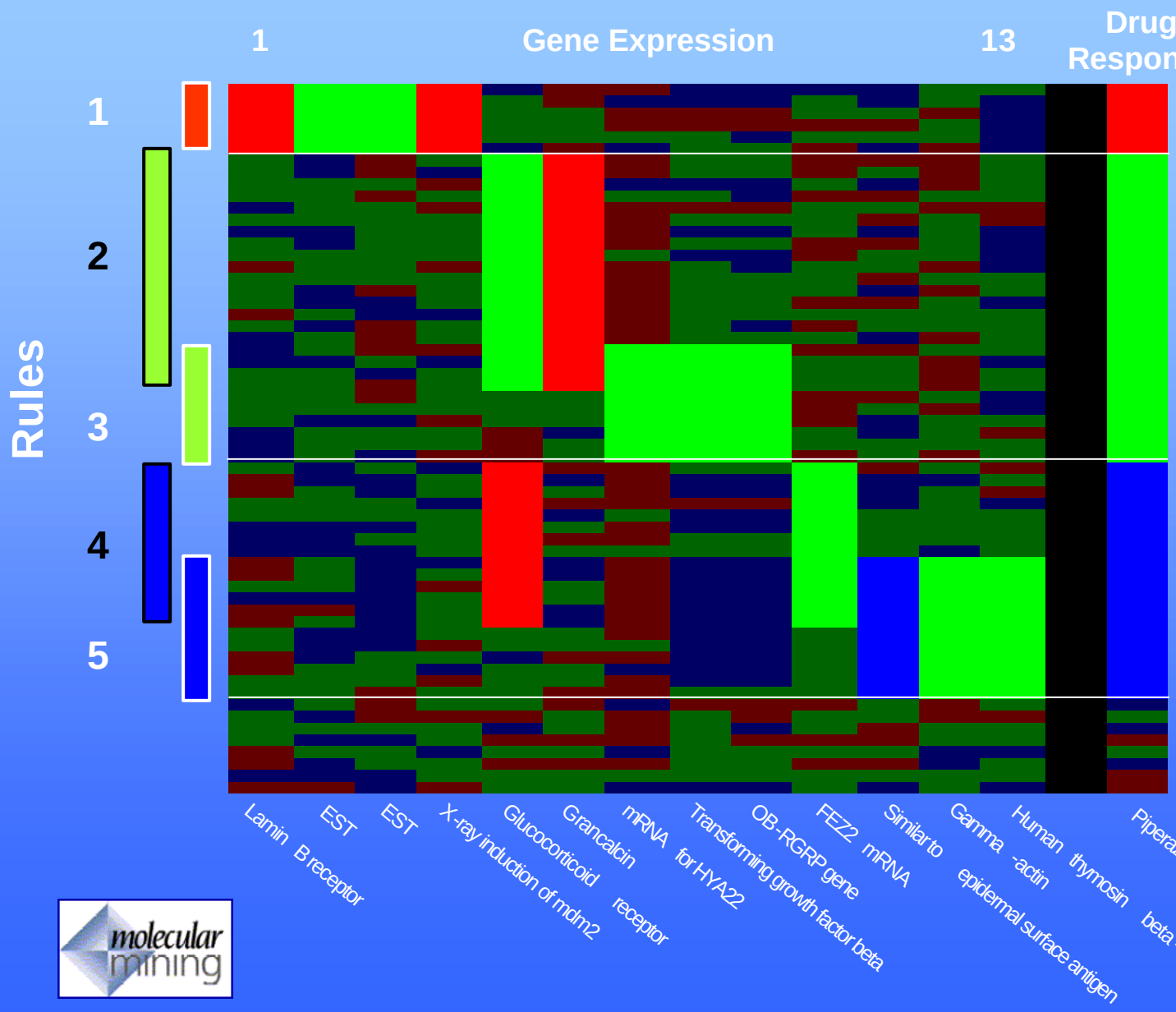
# Analysis of NCI60

Can we find answers to explicit questions using advanced datamining methods?

1. Can cancer **drug response** be predicted from **gene expression**?
2. Is the number of predictive genes **small** enough to be **relevant to clinical diagnostics**?
3. Can we identify **novel drug targets** from existing drug response?

molecular mining

# MMC Tools Predict Cancer Drug Response from a Small Set of Genes




**Example: Rule 1**

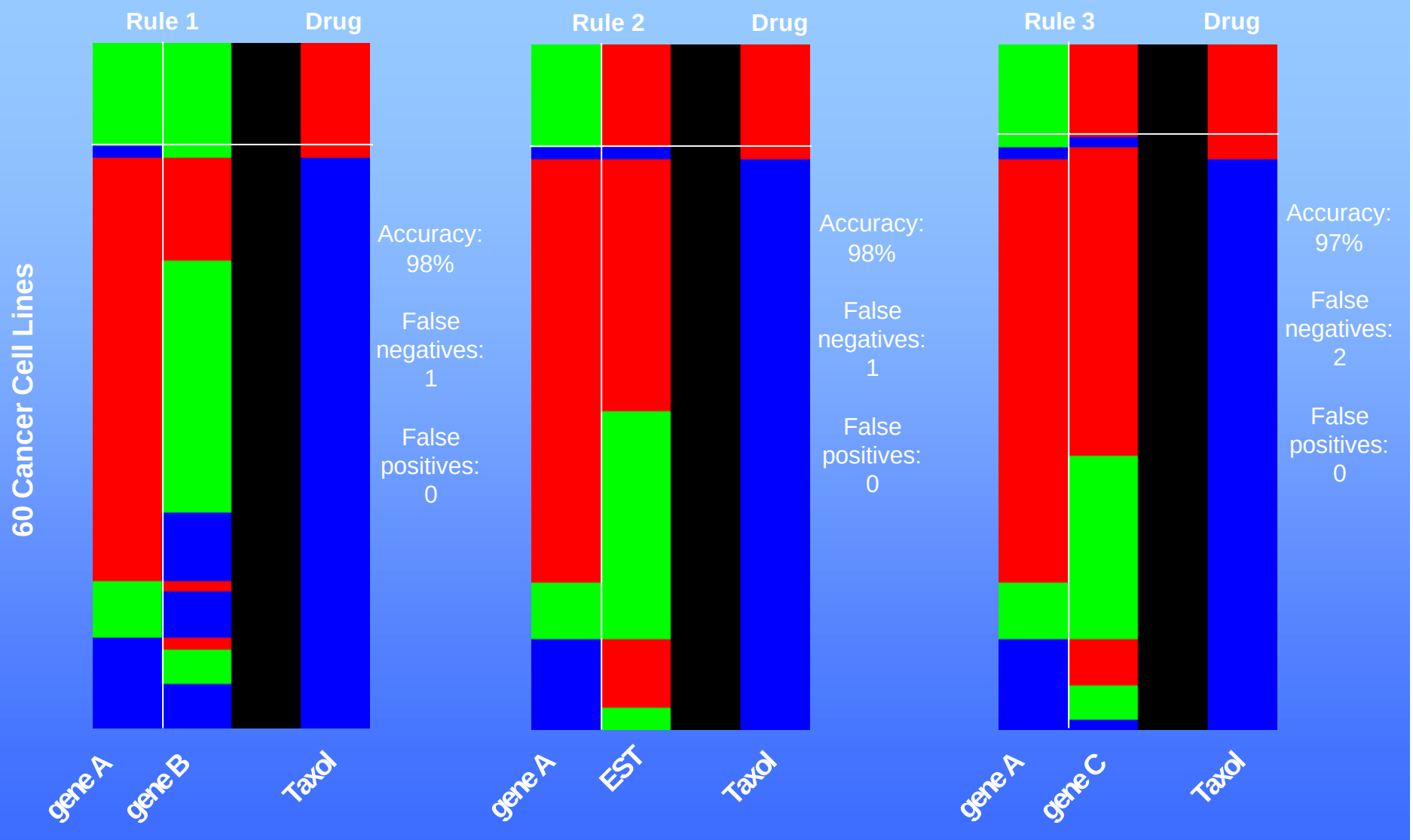
If
Gene 1 is *HIGH*
Gene 2 is *MEDIUM*
Gene 3 is *MEDIUM*
Gene 4 is *HIGH*
then
***Piperazine mustard*** response is *HIGH*

Note that gene 5 (Glucocorticoid receptor) participates in the prediction of the *MEDIUM* and the *LOW* states of the drug.

molecular mining

Data source: NCI60
John Weinstein
National Cancer Institute

# Optimized MMC Rules Predicting Response to **Taxol**




molecular mining

Data source: NCI60
John Weinstein
National Cancer Institute

# Analysis of NCI60

Predictive genes fall into several functional groups

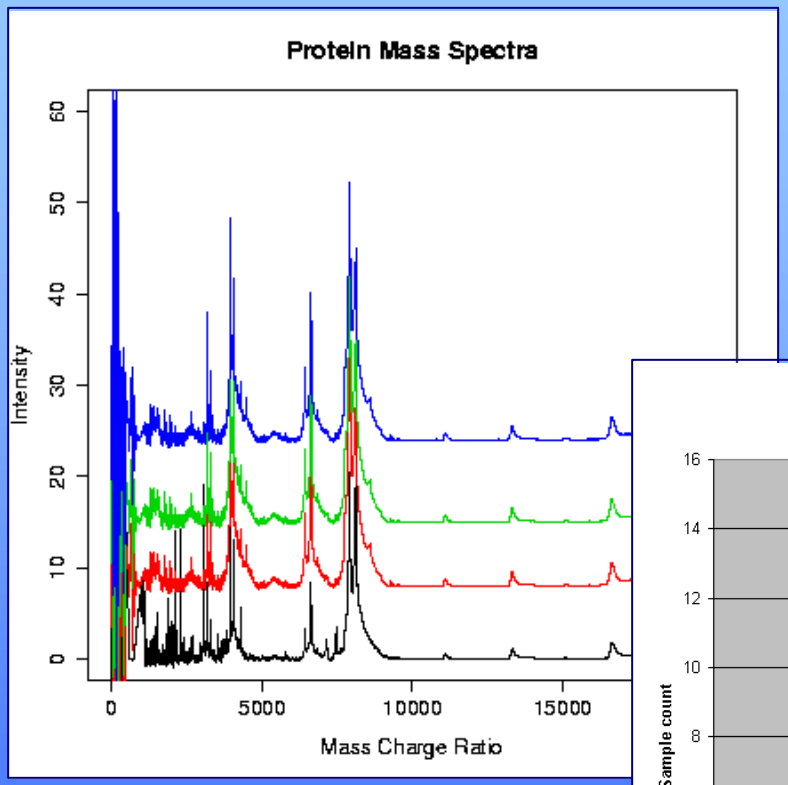
- Cytoskeleton
- Protein phoshorylation
- Transcription factor
- Protease
- Scaffold
- EST

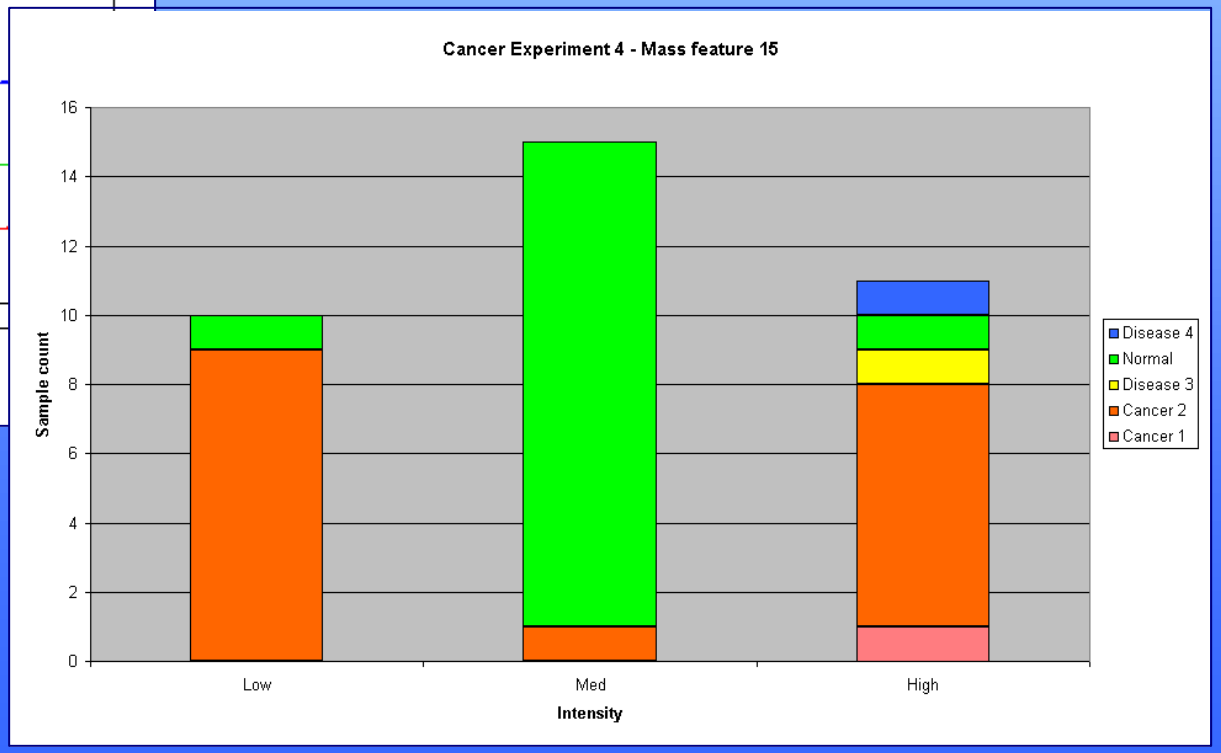
64 Drugs with MMC rules

5-6-Dihydro-5-azacytidine, 5-Hydroxypicolinaldehyde-thiose, alpha-2'-Deoxythioguanosine, Amonafide, Amsacrine, Anthrapyrazole-derivative, Asaley, Azacytidine, Carboplatin, Chlorambucil, Clomesone, CPT, CPT,10-OH, CPT,20-acetate, CPT,20-ester, CPT,20-ester, Cyanomorpholinodoxorubicin, Cyclocytidine, Cytarabine, Daunorubicin, Deoxydoxorubicin, Diaminocyclohexyl-Pt-II, Diaziridinylbenzoquinone, Dichloroallyl-lawsone, Dolastatin-10, Doxorubicin, DUP785-brequinar, Etoposide, Fluorodopan, Fluorouracil, Geldanamycin, Guanazole, Halichondrin, Hycanthone, Hydroxyurea, Inosine-glycodialdehyde, Iproplatin, L-Alanosine, Lomustine, Mechlorethamine, Melphalan, Mitozolamide, Morpholino-adriamycin, N-N-Dibenzyl-daunomycin, N-phosphonoacetyl-L-aspartic-ac, Oxanthrazole, Paclitaxel---Taxol, PCNU, Piperazine, Pipobroman, Porfiromycin, Pyrazofurin, Pyrazoloimidazole, Semustine, Spiromustine, Teroxirone, Thioguanine, Thiopurine, Thiotepa, Triethylenemelamine, Trimetrexate, Trityl-cysteine, Uracil, Yoshi-864

molecular mining

# A Proteomics Example: Extracting Disease Markers from Mass Spec Data

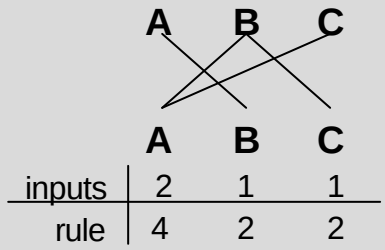



molecular mining

# Characteristics of a simple Boolean network

## Wiring and rules

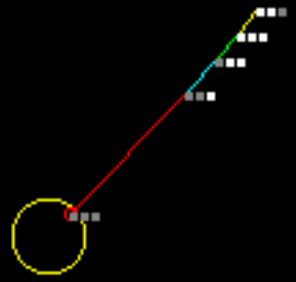



Basis for rules:

1. A activates B
2. B activates A and C
3. C inhibits A

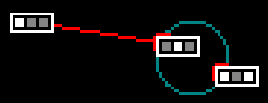
| iteration | A | B | C |
|---|---|---|---|
| 1 | 1 | 1 | 0 |
| 2 | 1 | 1 | 1 |
| 3 | 0 | 1 | 1 |
| 4 | 0 | 0 | 1 |
| 5 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 |

Trajectory 1 results in a point attractor



| iteration | A | B | C |
|---|---|---|---|
| 1 | 1 | 0 | 0 |
| 2 | 0 | 1 | 0 |
| 3 | 1 | 0 | 1 |
| 4 | 0 | 1 | 0 |

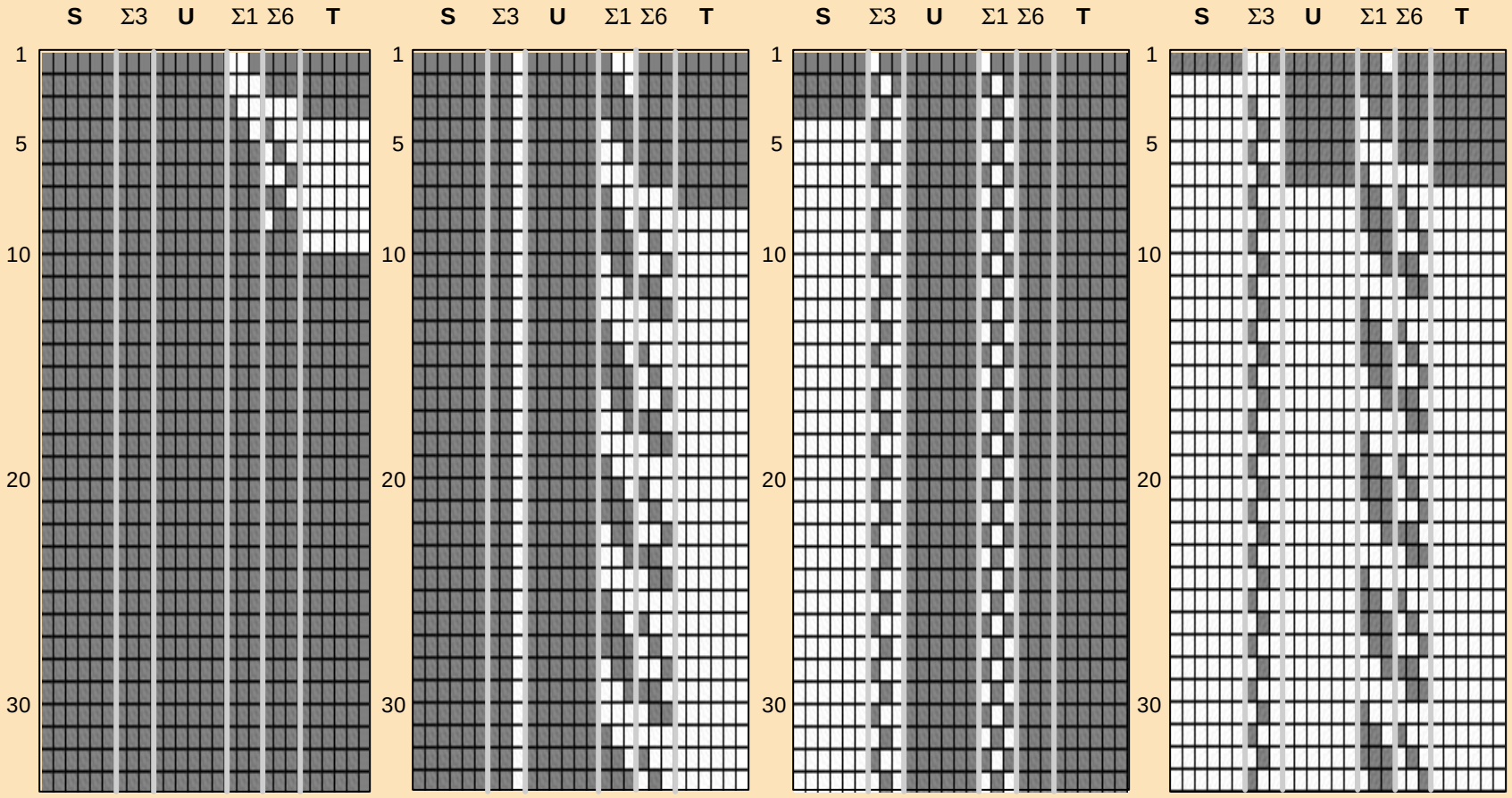
Trajectory 2 results in a 2-state dynamic attractor



molecular mining

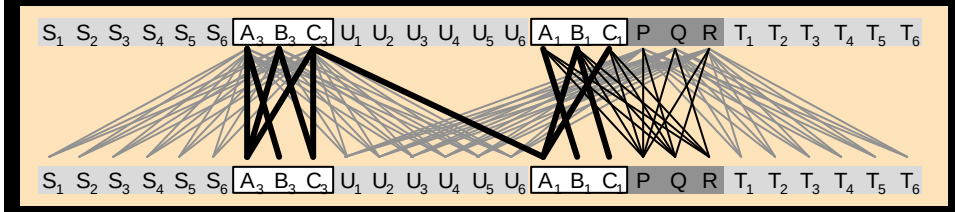
# Cell types as alternative attractors in genetic networks

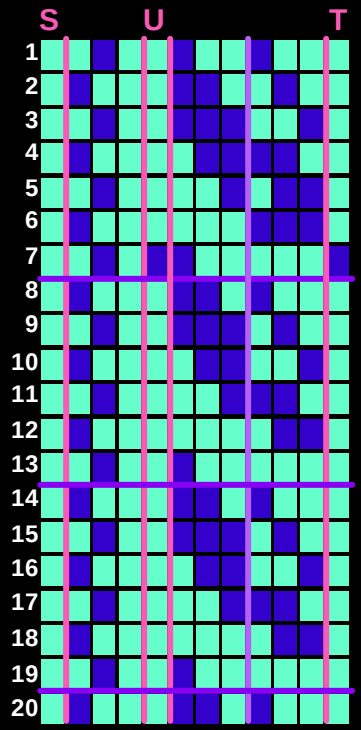
## Trajectories leading to alternative attractors

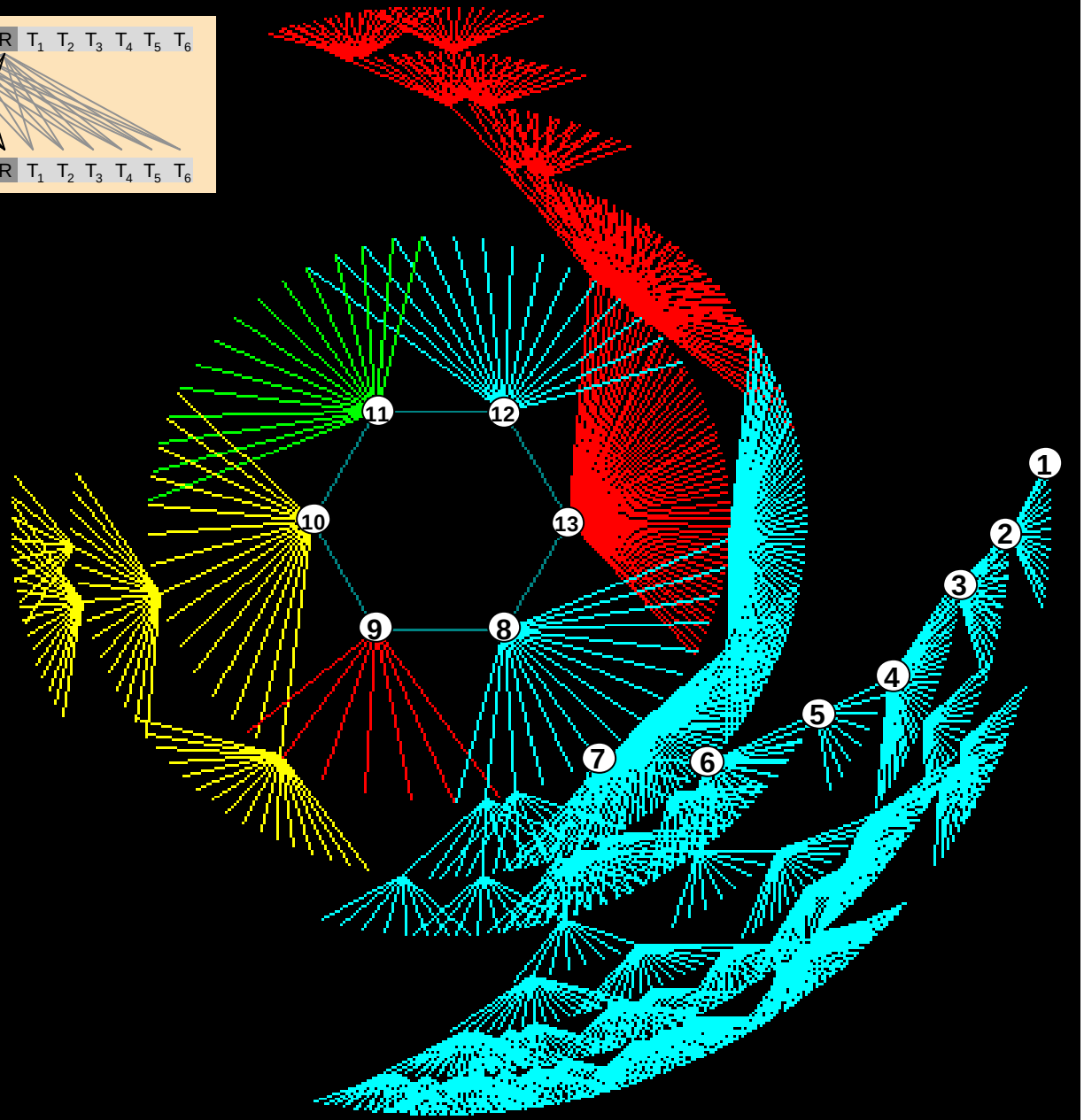



molecular mining

# A hypothetical genetic network



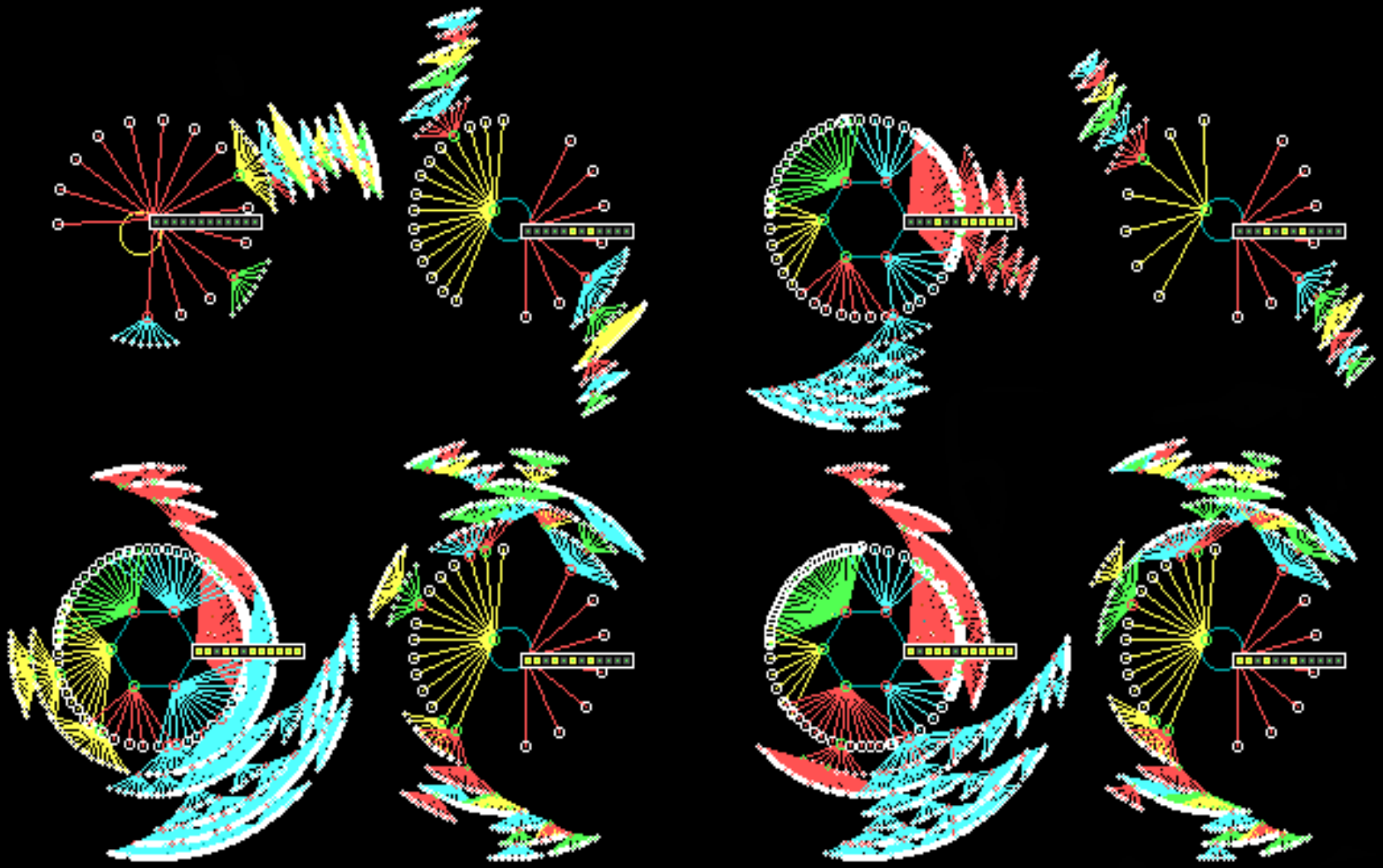




molecular mining

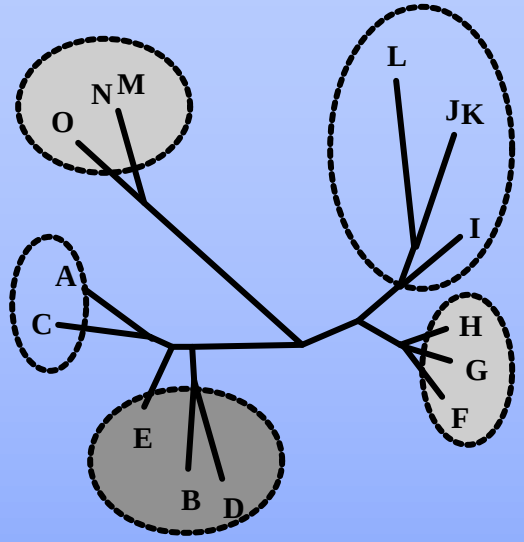
# The Expanded Network: Complete Basin of Attraction Fields

**4096 states mapping to 8 attractors**



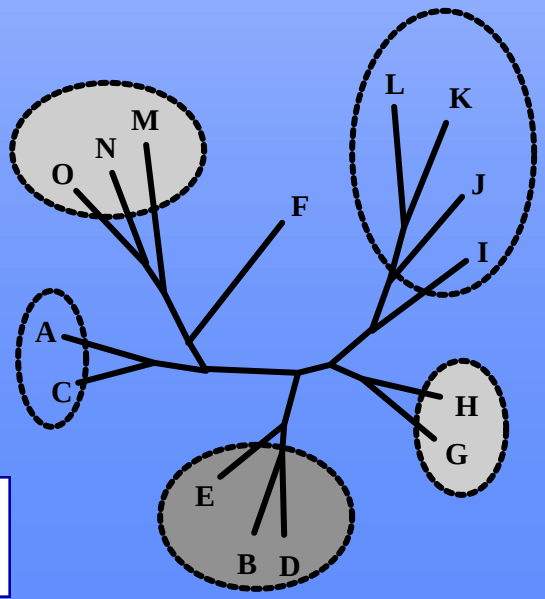
molecular mining

# Cluster analysis captures shared control rules




## Wiring (Molecular Interaction) Clusters

| gene | Boolean rule |
|---|---|
| A | F and H and J |
| B | G and H and J |
| C | F and H and I |
| D | G and H and I |
| E | H and I and J |
| F | I and J and K and L and (not G) |
| G | I and J and K and L and (not O) |
| H | I and J and K and L |
| I | J and K and L |
| J | K and L |
| K | K or L |
| L | L or M |
| M | N or O |
| N | N and O |
| O | N and O and (not E) |




## Trajectory (Gene Expression) Clusters

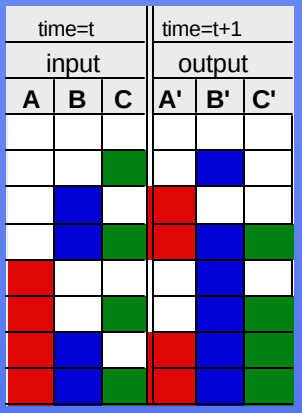
| trajectory | I | | | | | | | | | | | II | | | | III | | | | IV | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| time | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| A | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| D | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| E | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| F | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| G | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| H | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| I | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| J | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| L | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| M | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| N | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| O | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

# REV.E.AL. – Reverse Engineering Algorithm

**How to go from**

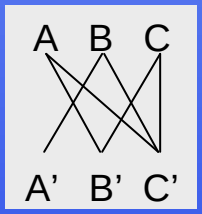
state transition measurements

| input | | | output | | |
|---|---|---|---|---|---|
| **A** | **B** | **C** | **A'** | **B'** | **C'** |
| 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 1 | 0 |
| 0 | 1 | 0 | 1 | 0 | 0 |
| 0 | 1 | 1 | 1 | 1 | 1 |
| 1 | 0 | 0 | 0 | 1 | 0 |
| 1 | 0 | 1 | 0 | 1 | 1 |
| 1 | 1 | 0 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 | 1 |




**to**

connections




**and**

functions or rules

A' = B
B' = A or C
C' = (A and B) or (B and C) or (A and C)

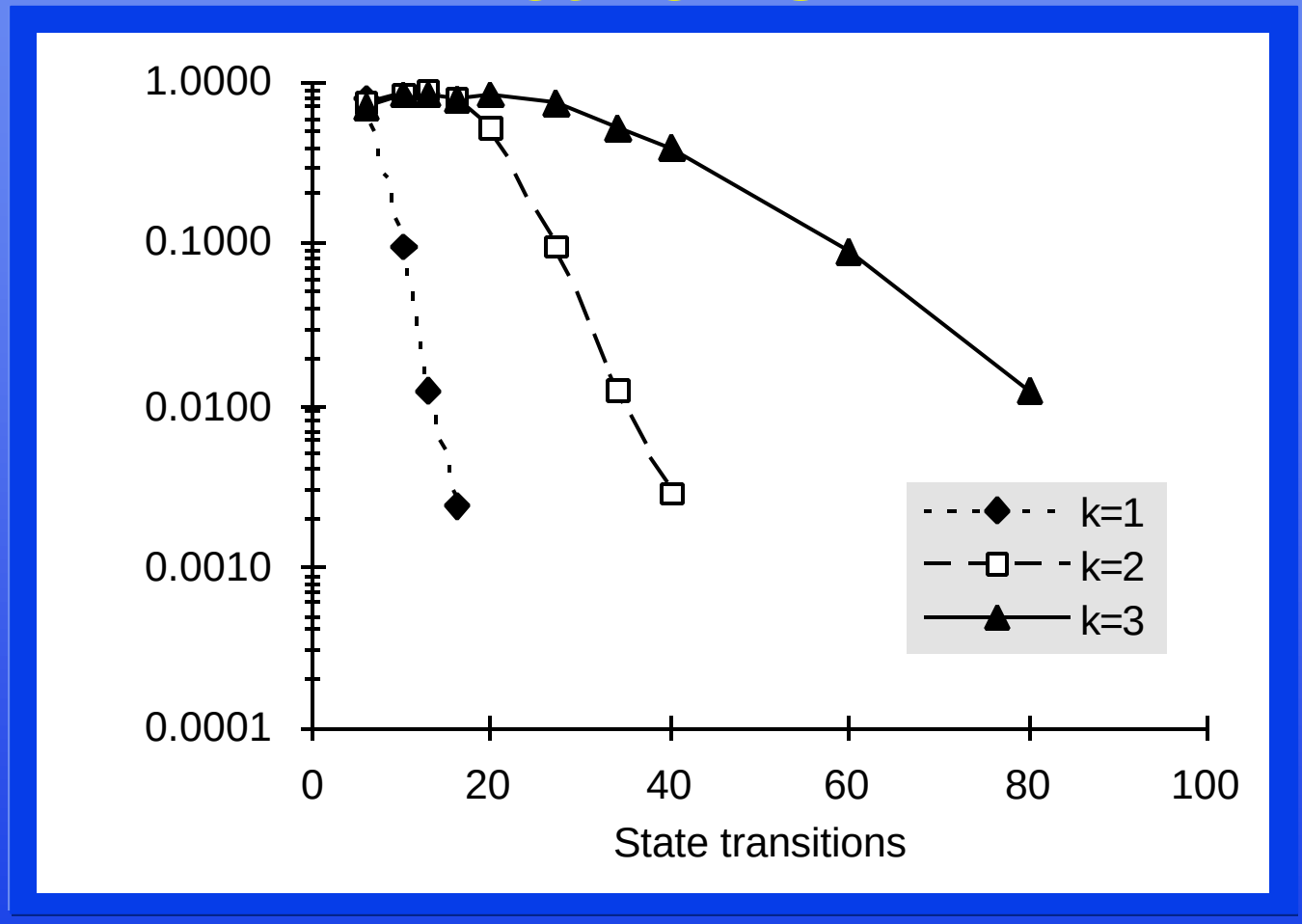
**using**

mutual information analysis

$$H(X) = -\sum p(x) \log p(x)$$
$$H(X,Y) = -\sum p(x,y) \log p(x,y)$$
$$M(X,Y) = H(X) + H(Y) - H(X,Y)$$
$$M(X,[Y,Z]) = H(X) + H(Y,Z) - H(X,Y,Z)$$

molecular mining

# REVEAL : Performance for n=50 gene networks




- REVEAL will always find a solution!
- But is the solution correct?

# Network Terminology

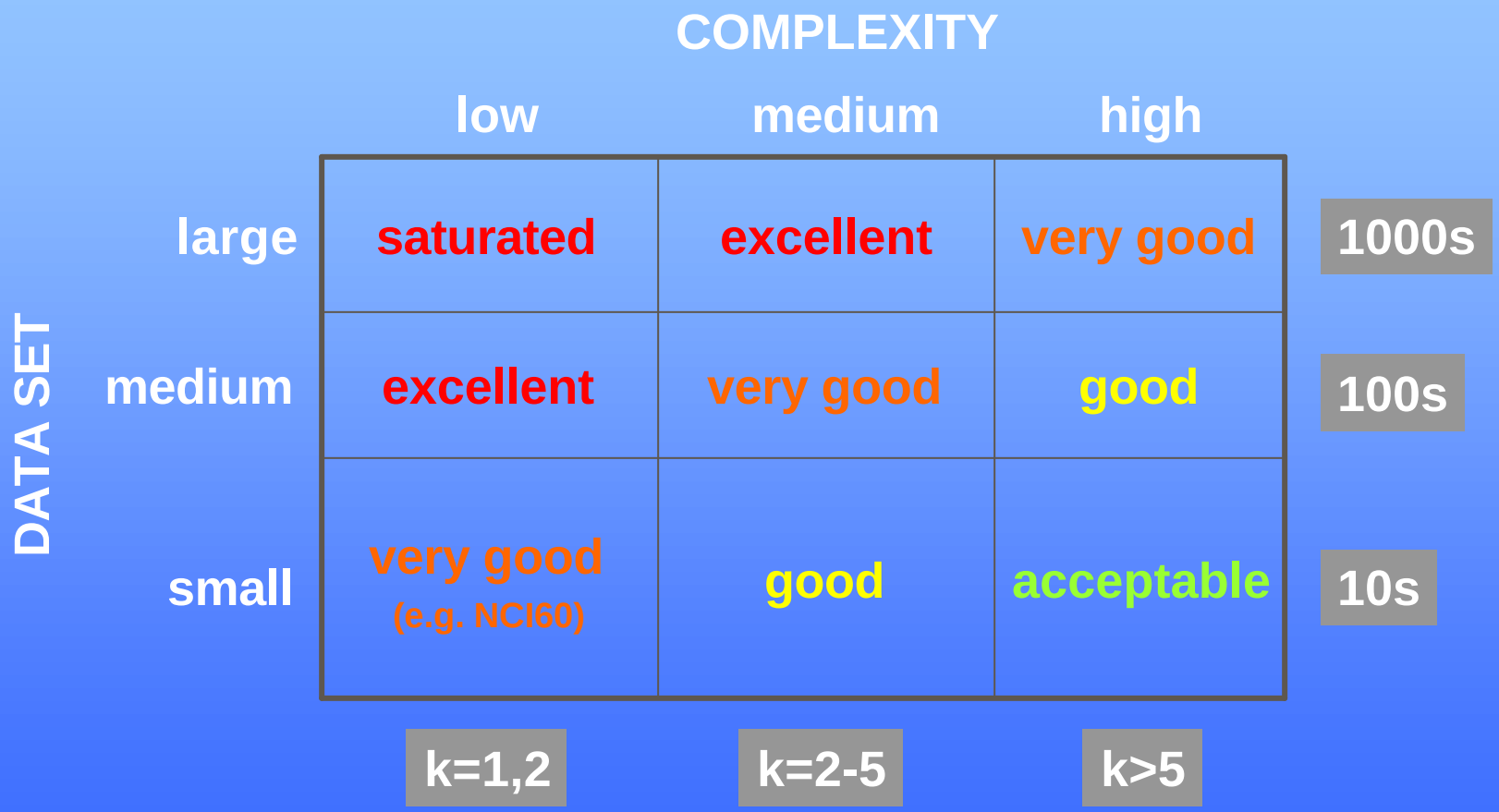
## Architecture

| | | |
|---|---|---|
| wiring | <-> | regulatory connections |
| rules (functions, codes) | <-> | regulatory interactions |

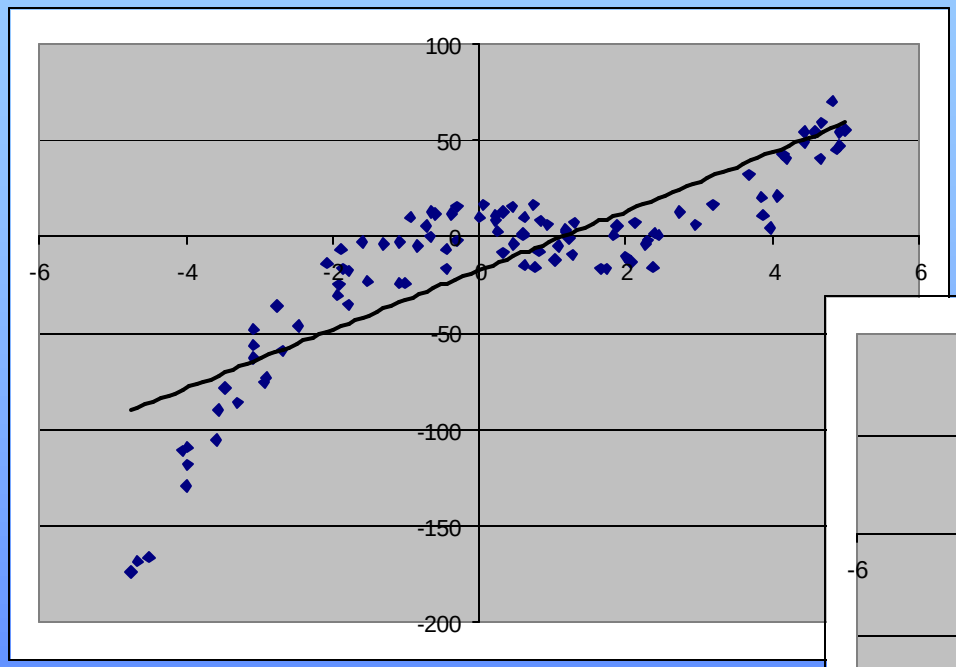
## Dynamics

| | | |
|---|---|---|
| state | <-> | set of molecular activities |
| state transition | <-> | response to previous state |
| trajectory | <-> | series of state transitions |
| attractor | <-> | final outcome, phenotype |

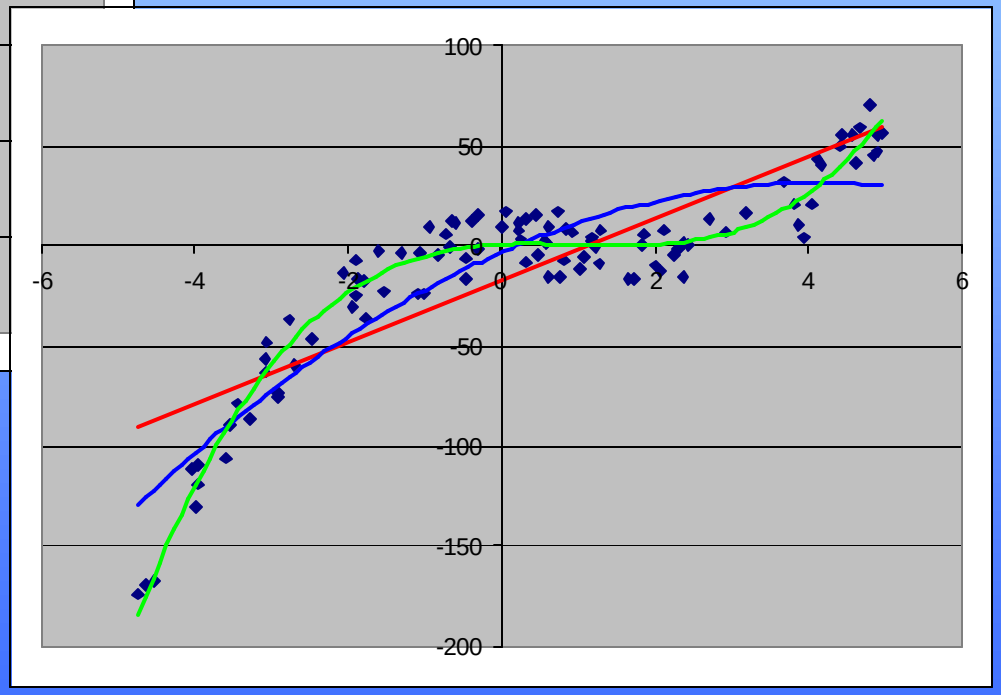
# Predictive power depends on complexity of relationships and depth of data



# Classical Statistics *vs.* Data Mining



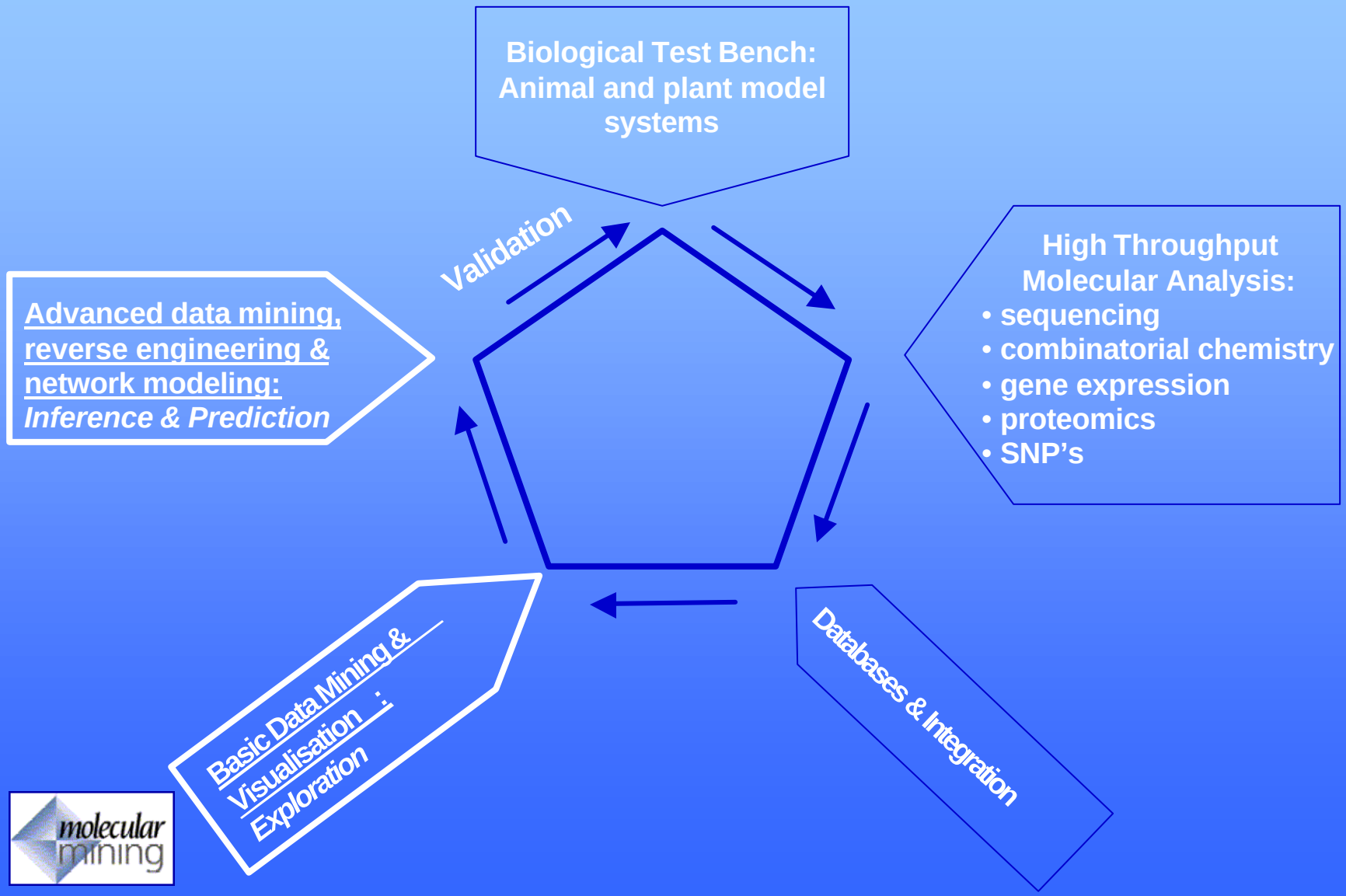

***Fit*** a model to a dataset




***Search*** through the space of possible models, to fit to a dataset

molecular mining

# The Discovery Algorithm



# The MMC Opportunity: Information + Model -> Prediction

- There is no model from which to predict function completely from sequence information.
- Simply “gene overexpressed here - underexpressed there” information does not reflect complexity of gene function

- We require **extensive activity information**, a **model framework** and **efficient algorithms** for prediction.

- MMC partnerships aim to integrate information and build valuable models for guidance of therapeutic strategies.

molecular mining

# Molecular Mining Corporation

## The Molecular Networks Company™

- Access reprints at **www.molecularmining.com**
- Contact info:

Molecular Mining Corporation
128 Ontario St.
Kingston, Ont, K7L 2Y4
Canada
P: 613-547-9752
F: 613-547-6835