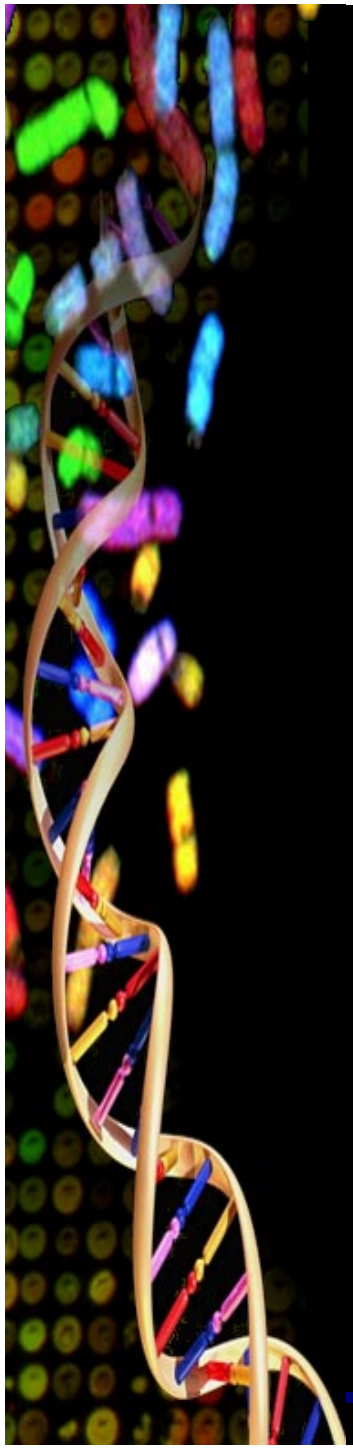
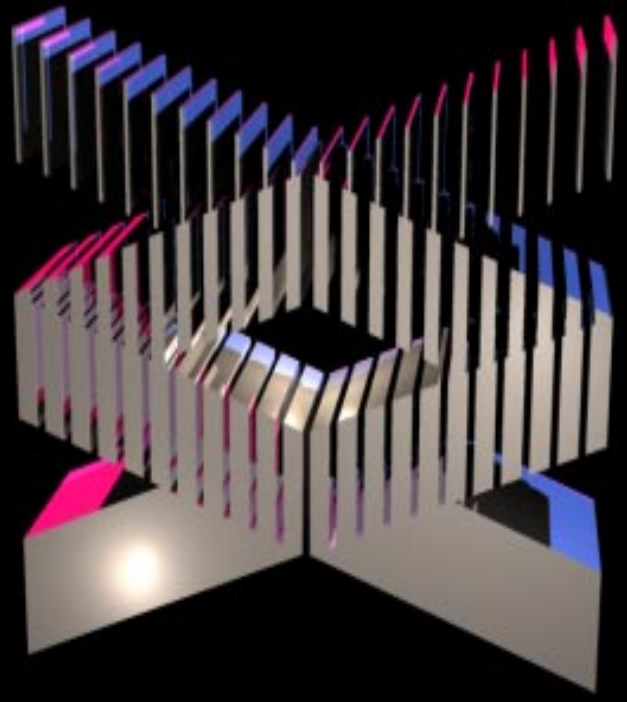




National Human Genome Research Institute

Cancer Genetics Branch

National Institutes of Health

Visualization for Cancer Classification By Gene Expression Profiling

National Human Genome Research Institute
Cancer Genetics Branch

Acknowledgement

NHGRI
J. Trent
P.Meltzer
M. Bittner
Y. Chen
Sheungchan Kim
Suja Panavally
Kim Gayton
Art Glatfelter
Chris Gooden
Yuan Jiang
John Lueders
Spyro Mousses
Bob Walker

Many many others

NHGRI
Andy Baxevanis
Tyra Wolfburg
Anthony Masiello
John Pearson
Ken Trout

Former NHGRIers
D. Duggan (NIAA)
Javed Khan(NCI)
Jun Wei(NCI)
H. Zhang (NCI)

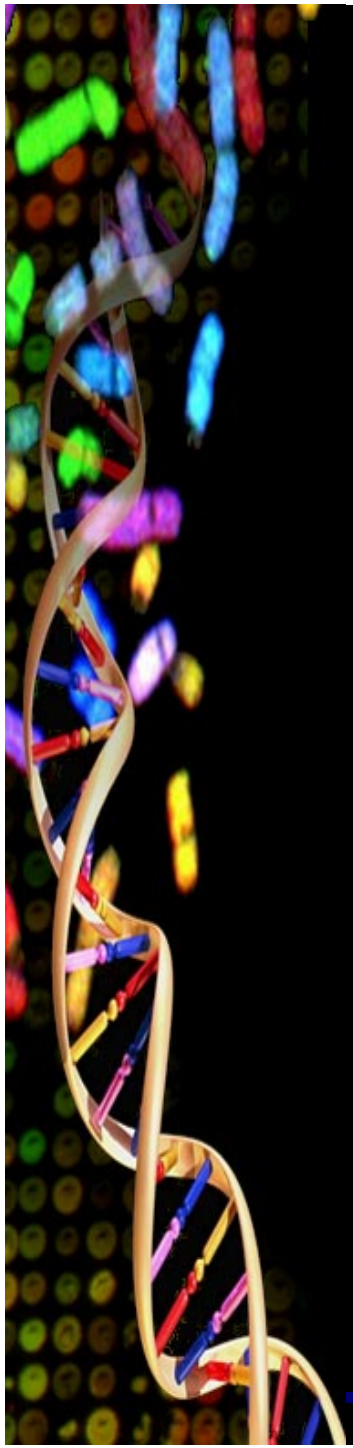
Lund Univ. Sweden
Carsten Peterson
Markus Ringer

NCI
R. Simon
Michael Radmacher

CIT
Edward Suh
Phani

Texas A&M
E. Dougherty
Ronaldo Hashimoto
Yoga Balagurunathan

Agilent Labs
Z. Yakhini
Allan Kuchinski
A. Ben-Dor
N. Sampas



OUTLINES

- Microarray Image Visualization
  - Color image, target detection, display.
  - Microarray image simulation.
- Microarray Data Visualization
  - Color-coded data
  - Clusterred view
  - MDS/PCA visualization
  - Gene Selections
  - Classification
- Microarray Analysis Visualization
  - Gene association by CoD analysis
  - Gene network exploration

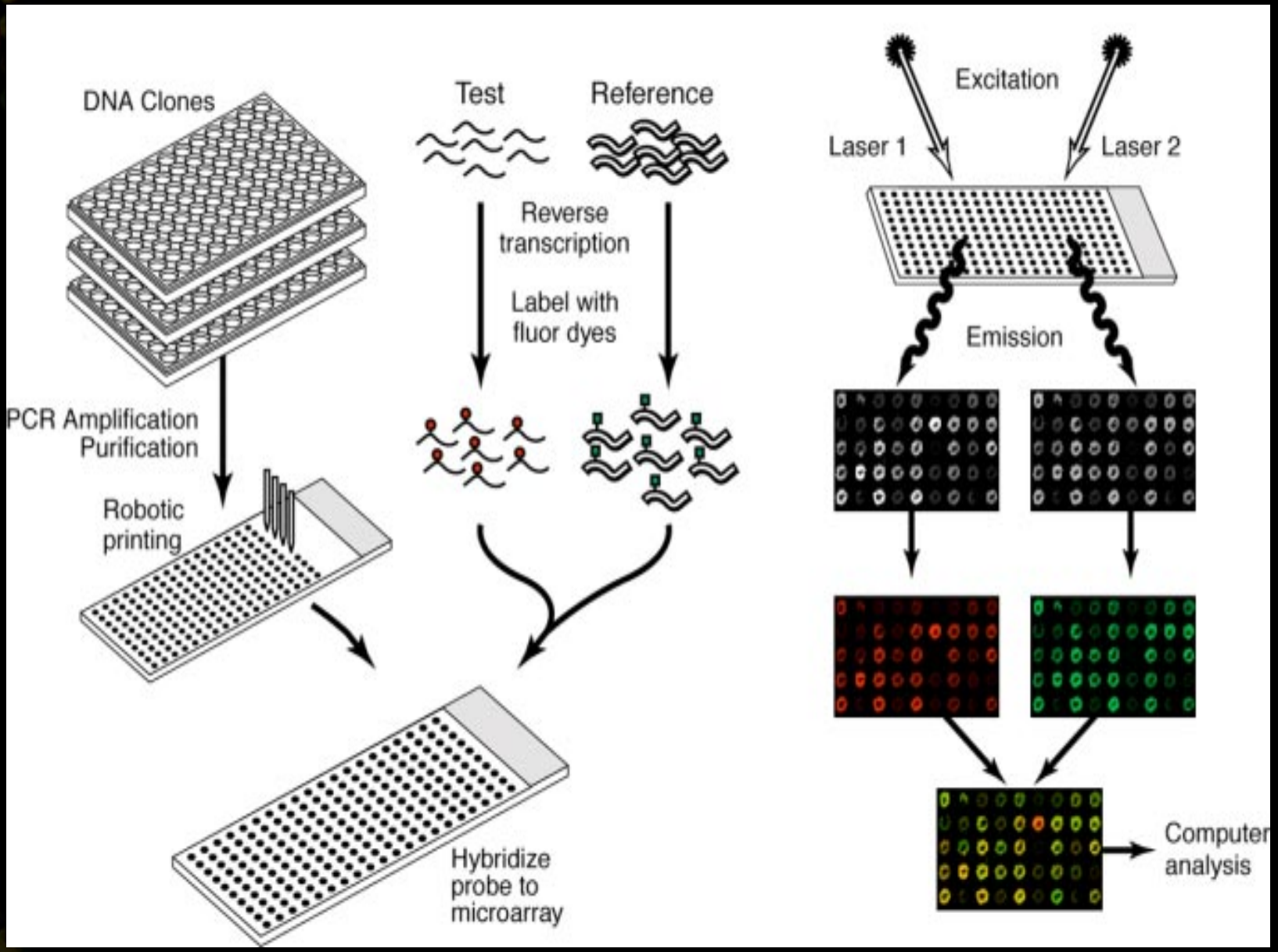
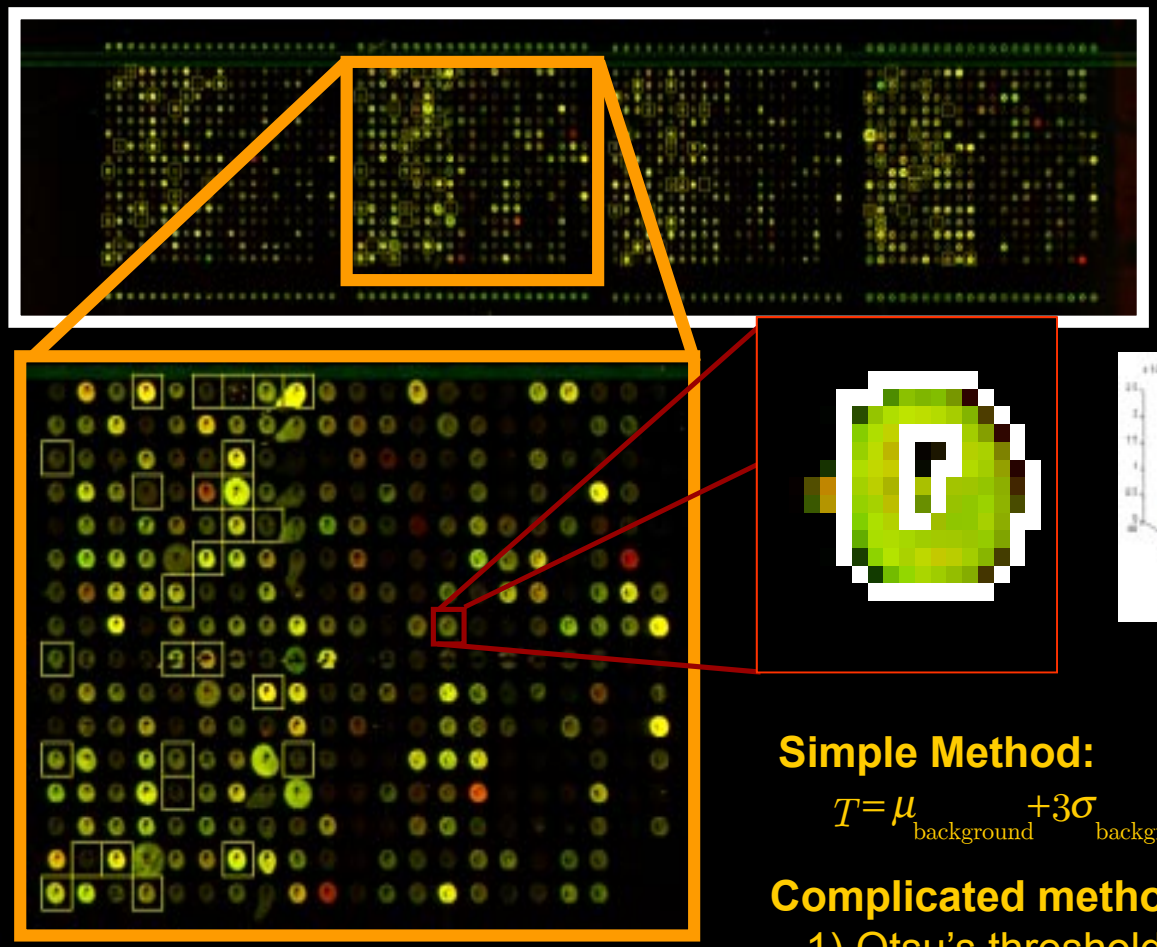
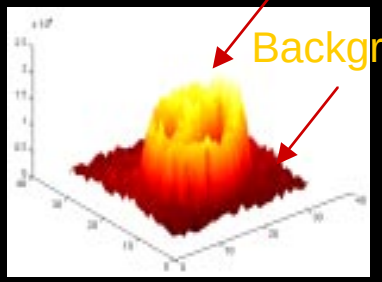

DNA Clones
PCR Amplification
Purification
Robotic
printing
Test
Reference
Reverse
transcription
Label with
fluor dyes
Hybridize
probe to
microarray
Excitation
Laser 1
Laser 2
Emission
Computer
analysis

Image Analysis: Details



cDNA target Detection



Simple Method:

$$T = \mu_{\text{background}} + 3\sigma_{\text{background}}$$

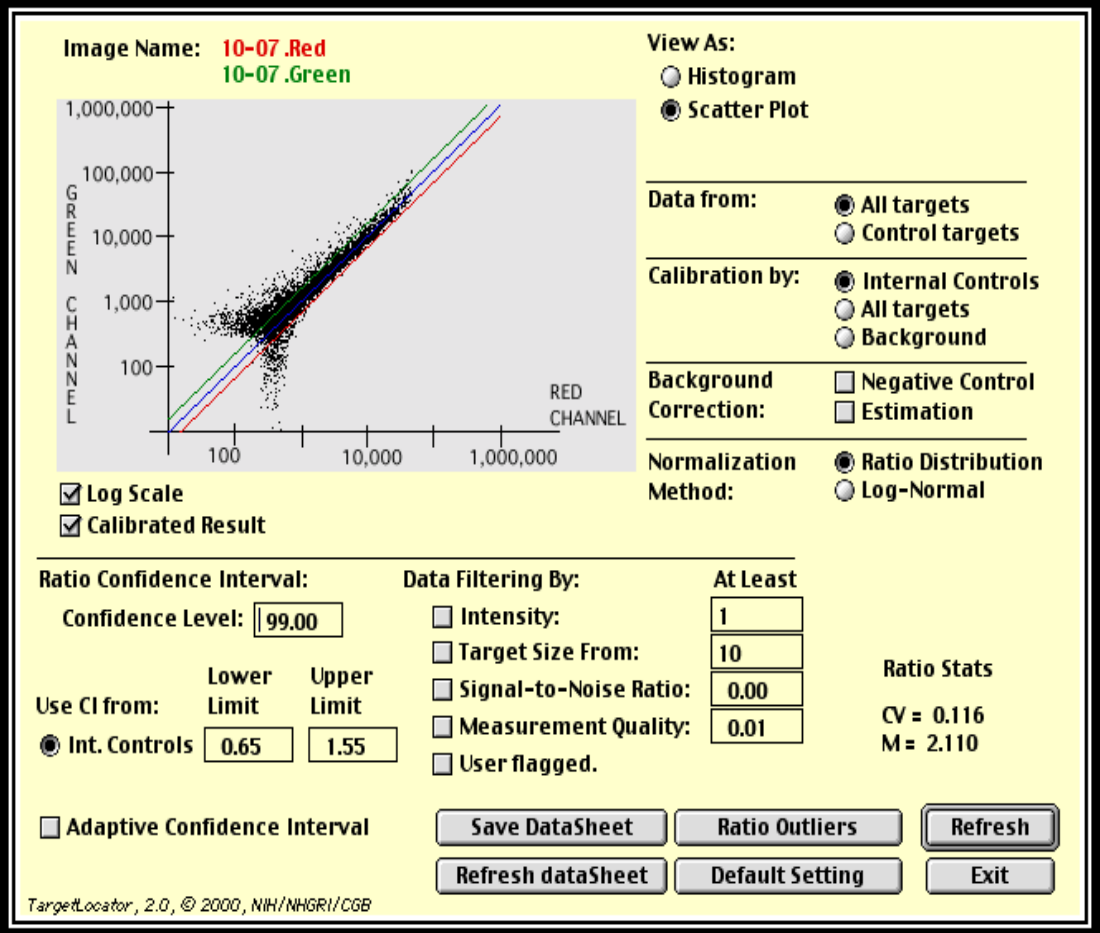
Complicated method:

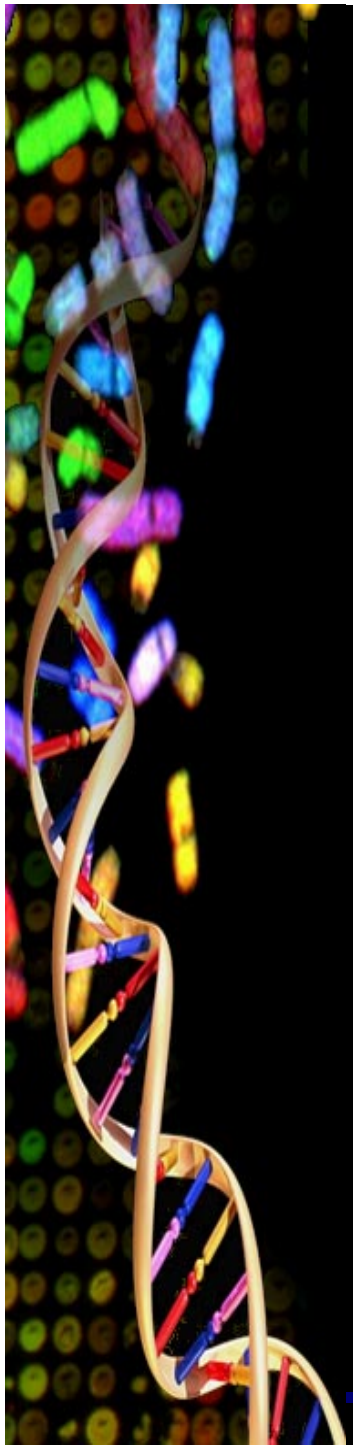
1) Otsu's threshold: *T* is selected such that the variance between-class is maximal

2) Mann-Whitney method

Ratio Statistics: DeArray display

99% Confidence interval, derived from ratio distribution of a set of house-keeping genes, which is used to quickly identify outlier expression ratios.



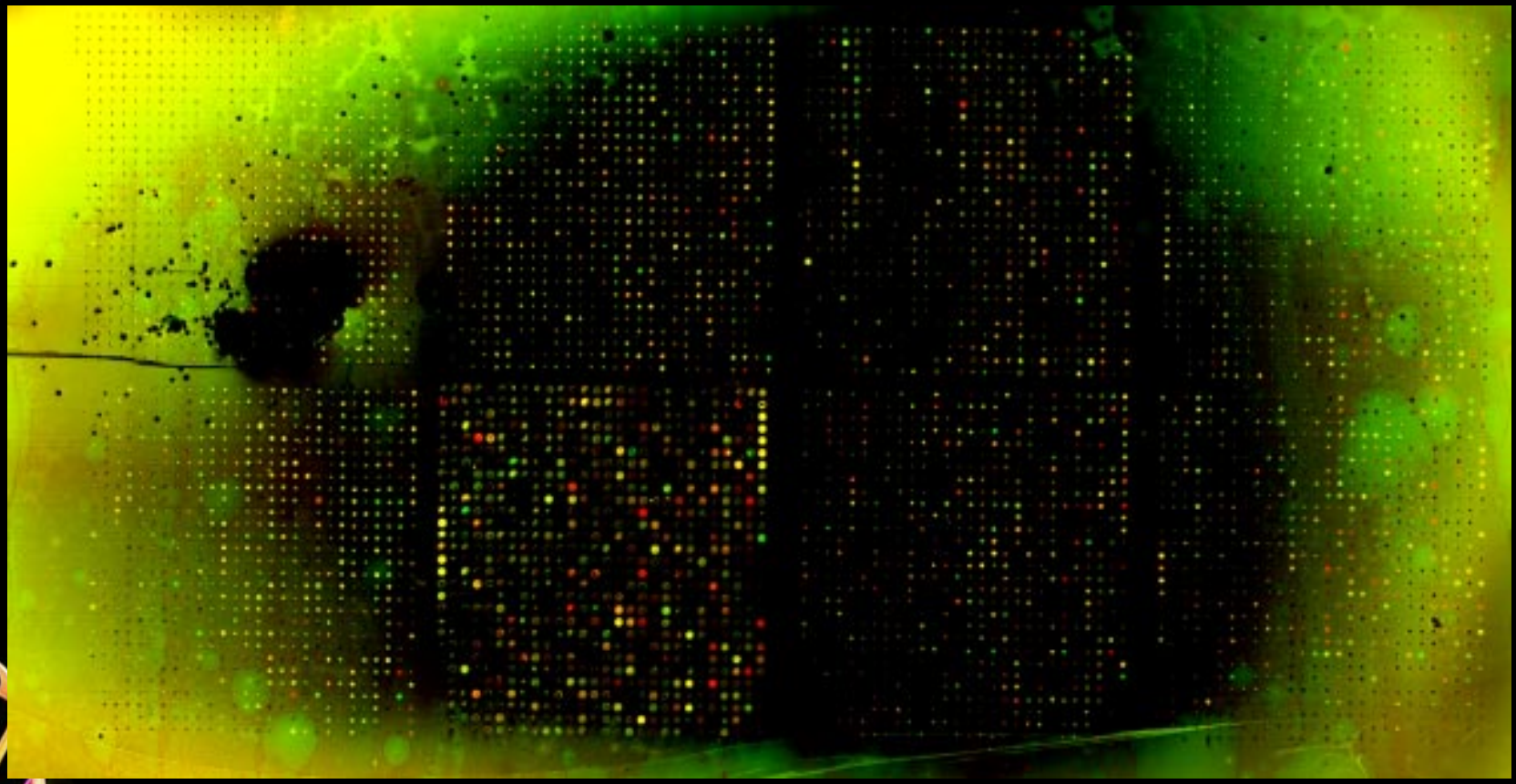


A Random Signal Model for cDNA Microarrays

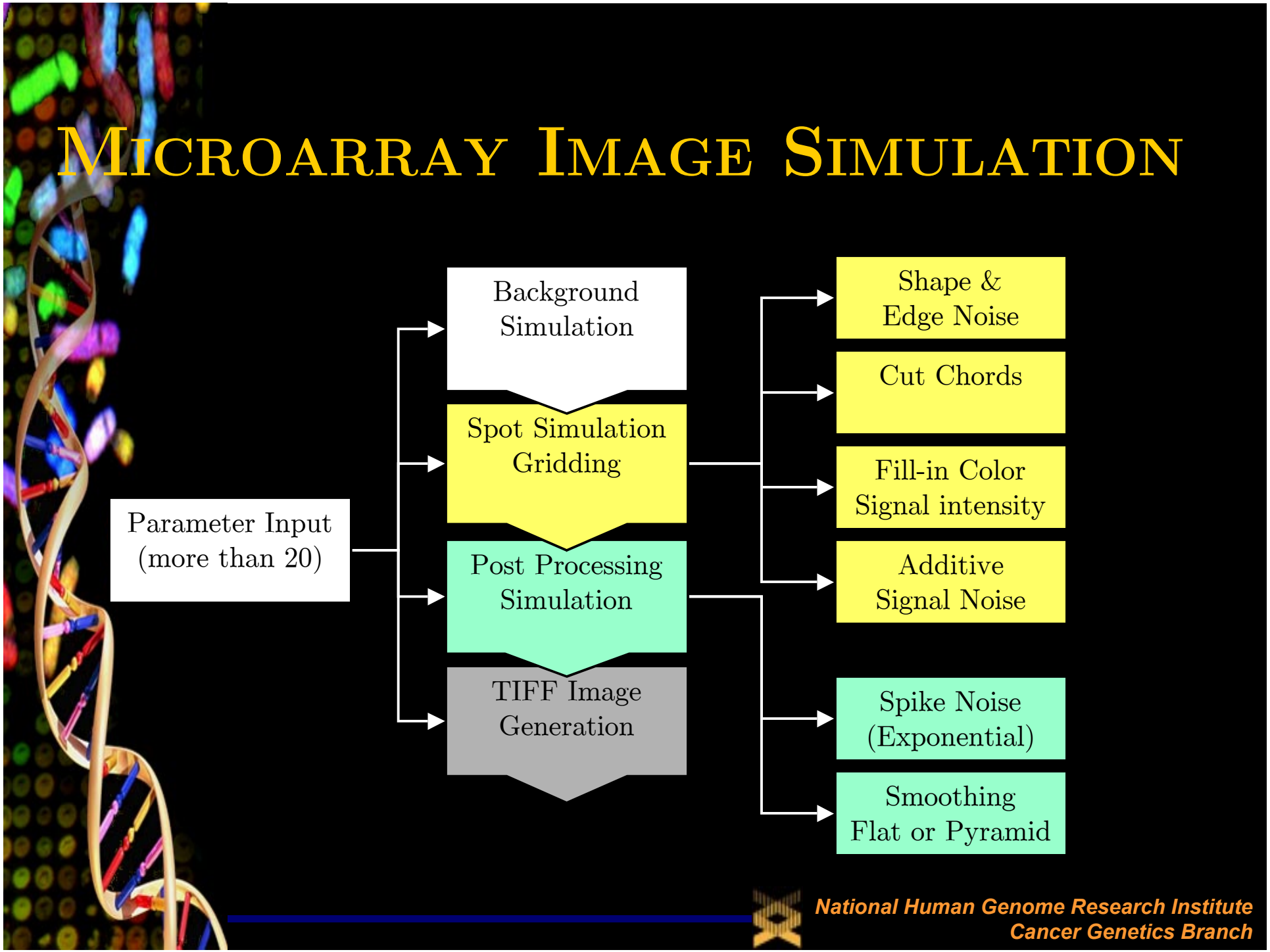
Yoganand Balagurunathan, Edward R. Dougherty
Department of Electrical Engineering
Texas A&M University

Yidong Chen, Michael L. Bittner, and J. M. Trent
National Human Genome Research Institute
National Institutes of Health

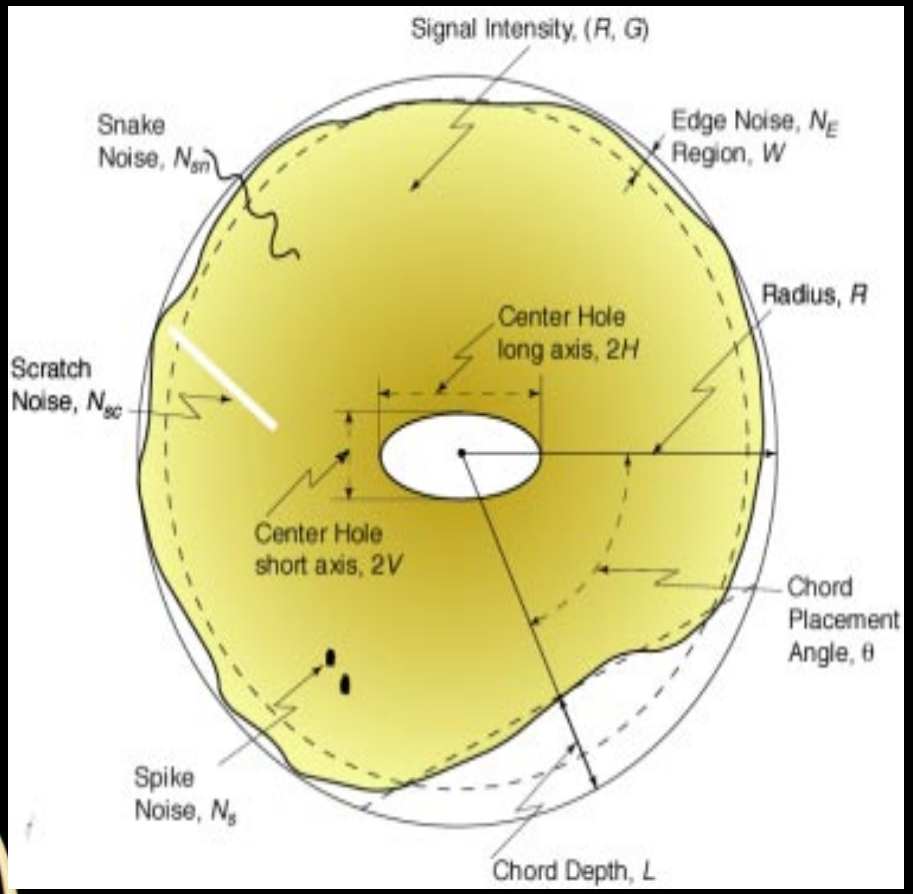
Typical cDNA Microarray Image

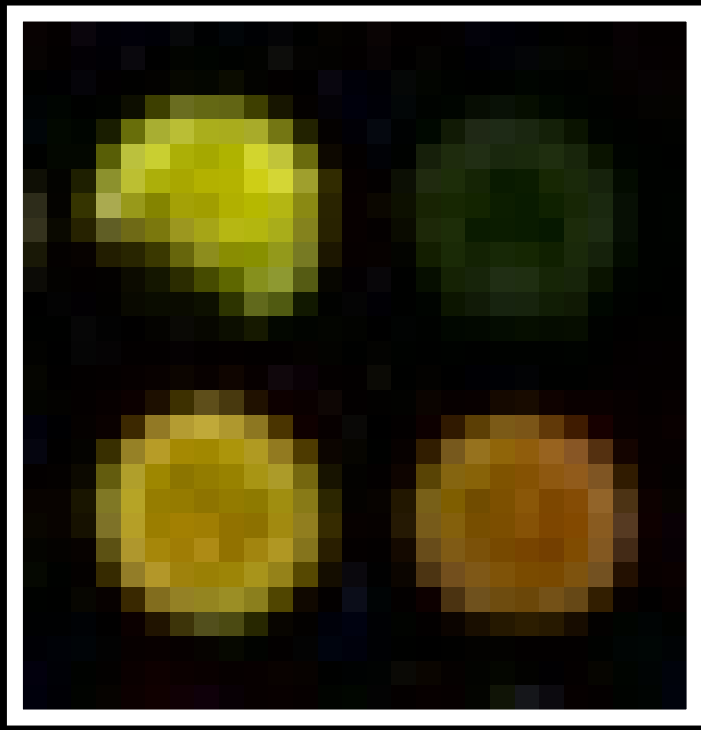


Microarray Image Simulation

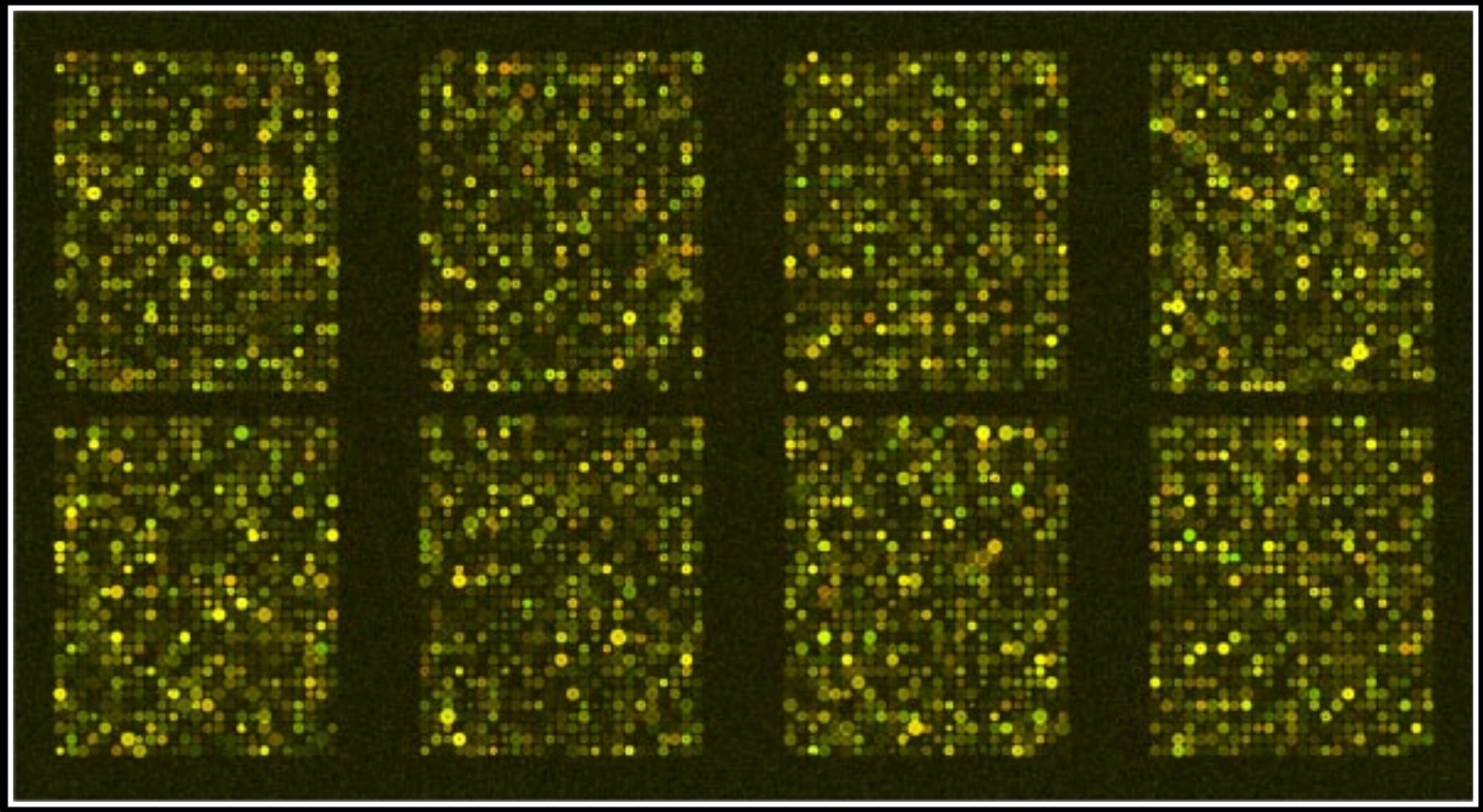


cDNA Microarray Spot Model





Simulated Microarray : Average/Noisy

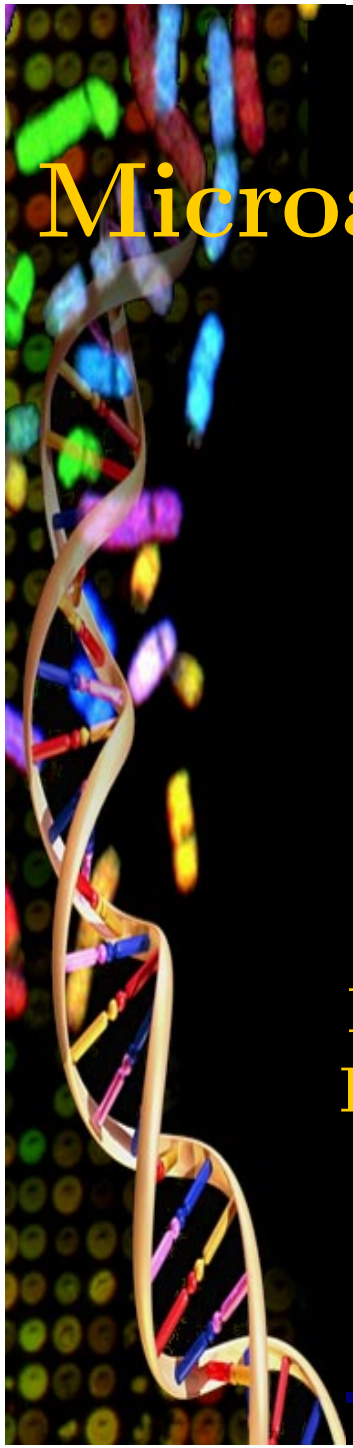


Molecular Classification of Cutaneous Malignant Melanoma by Gene Expression profiling

M. Bittner, P. Meltzer, Y. Chen, Y. Jiang, *et al*

National Human Genome Research Institute
National Institutes of Health

Nature 406, p536-540

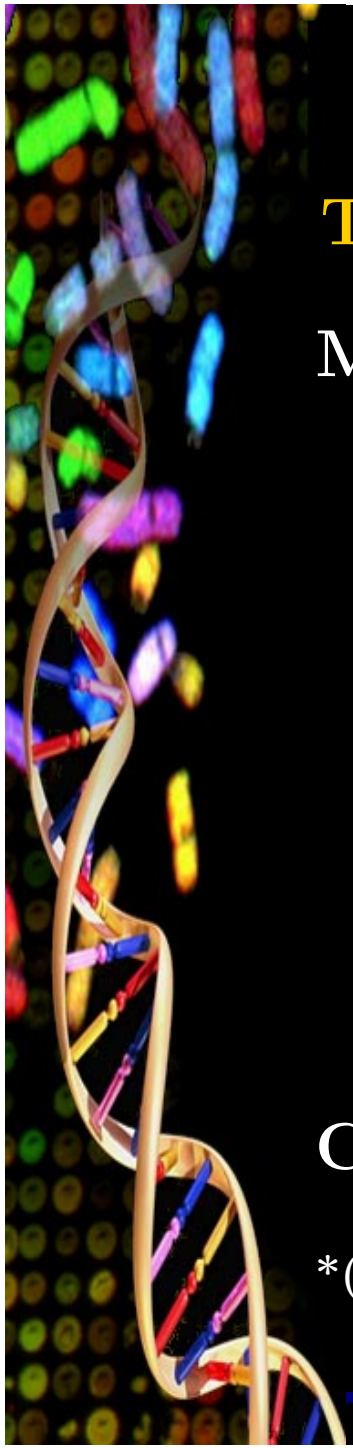


Microarray Analysis Of Melanoma

Finding The Biological Information Encoded In Gene Expression Profiles

Melanoma behaves differently across patients,
But no biological markers have been identified
which allow subset of melanoma.

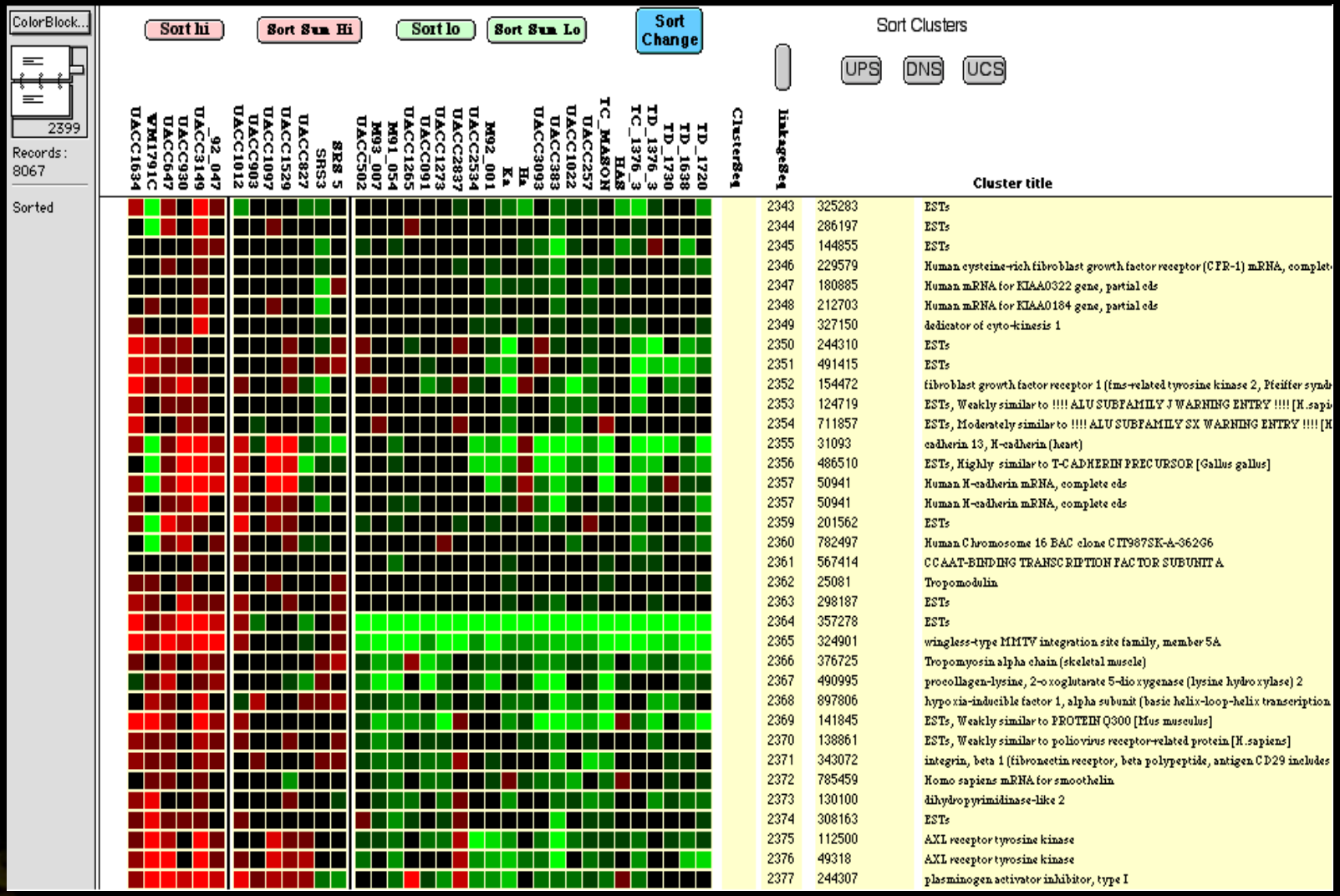
National Human Genome Research Institute
Cancer Genetics Branch



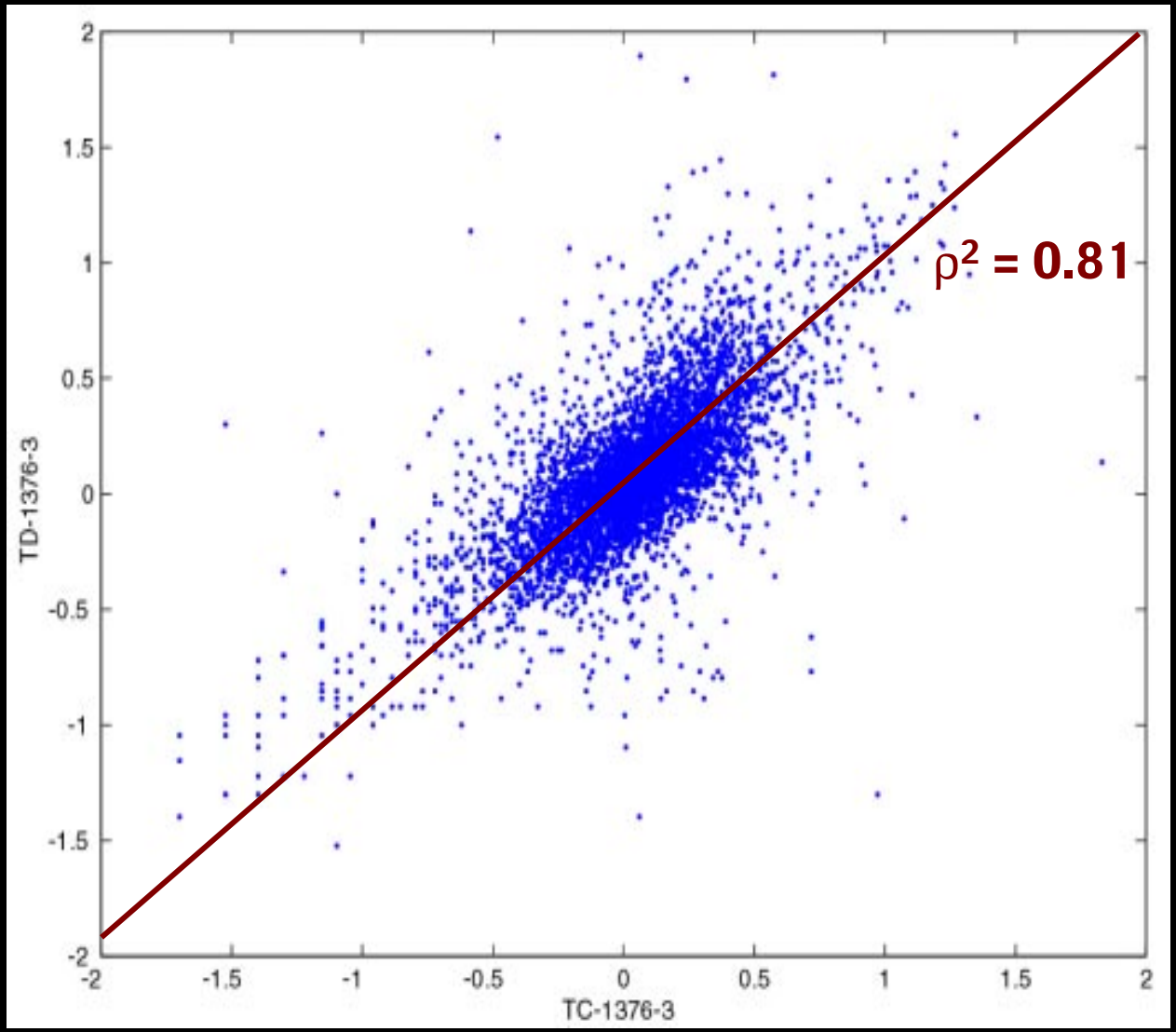
Total Experiments	**38**
Melanomas	31
Pt. Biopsies	5
Cell Cultures	26
(≤5 passage)	11
(6-10 passage)	4
(11-20 passage)	9
(>20 passage)	2
Controls*	7

*(Fibroblasts, tumor cell lines, normal epithelial cells)

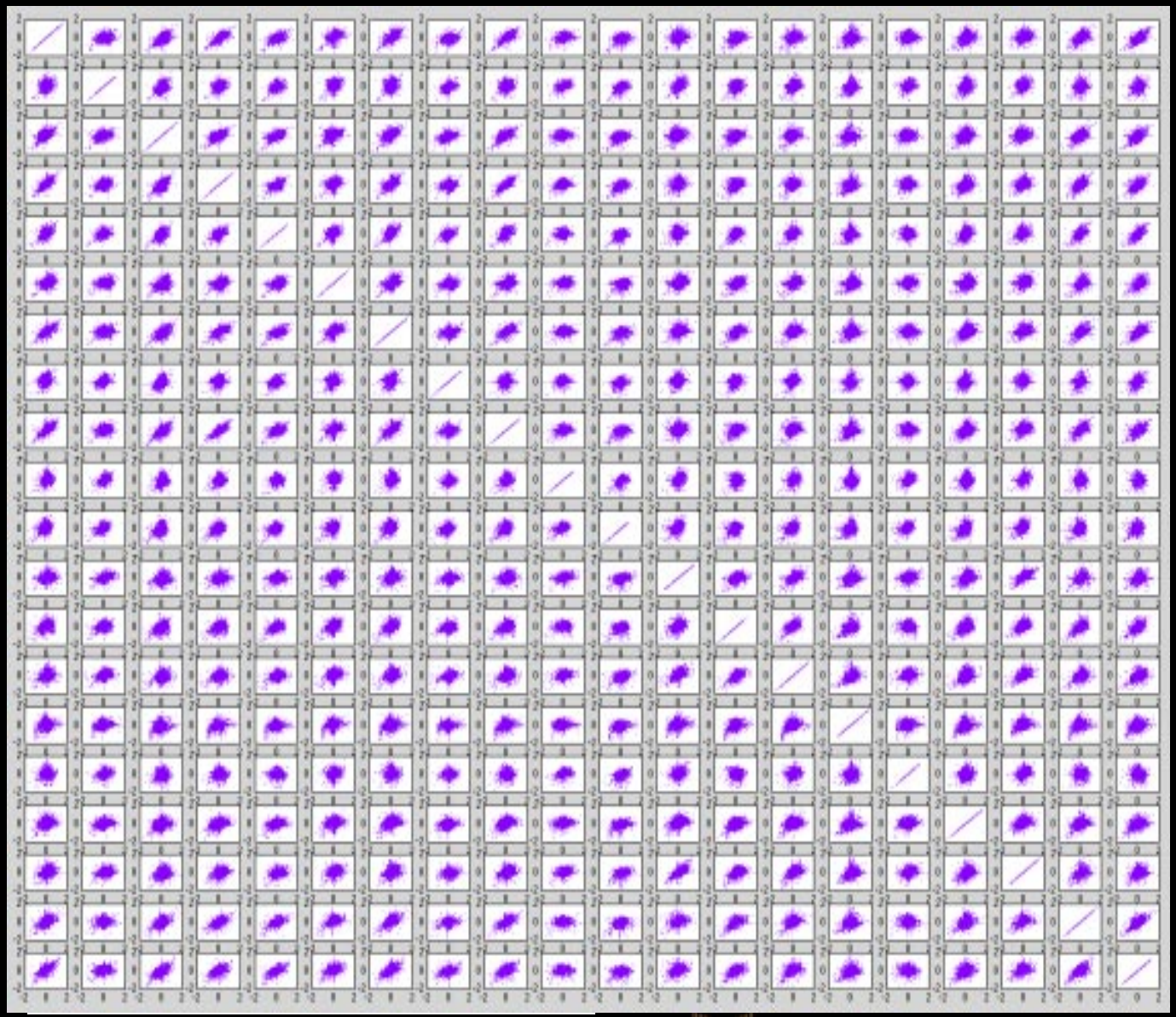
Gene Expression Pattern



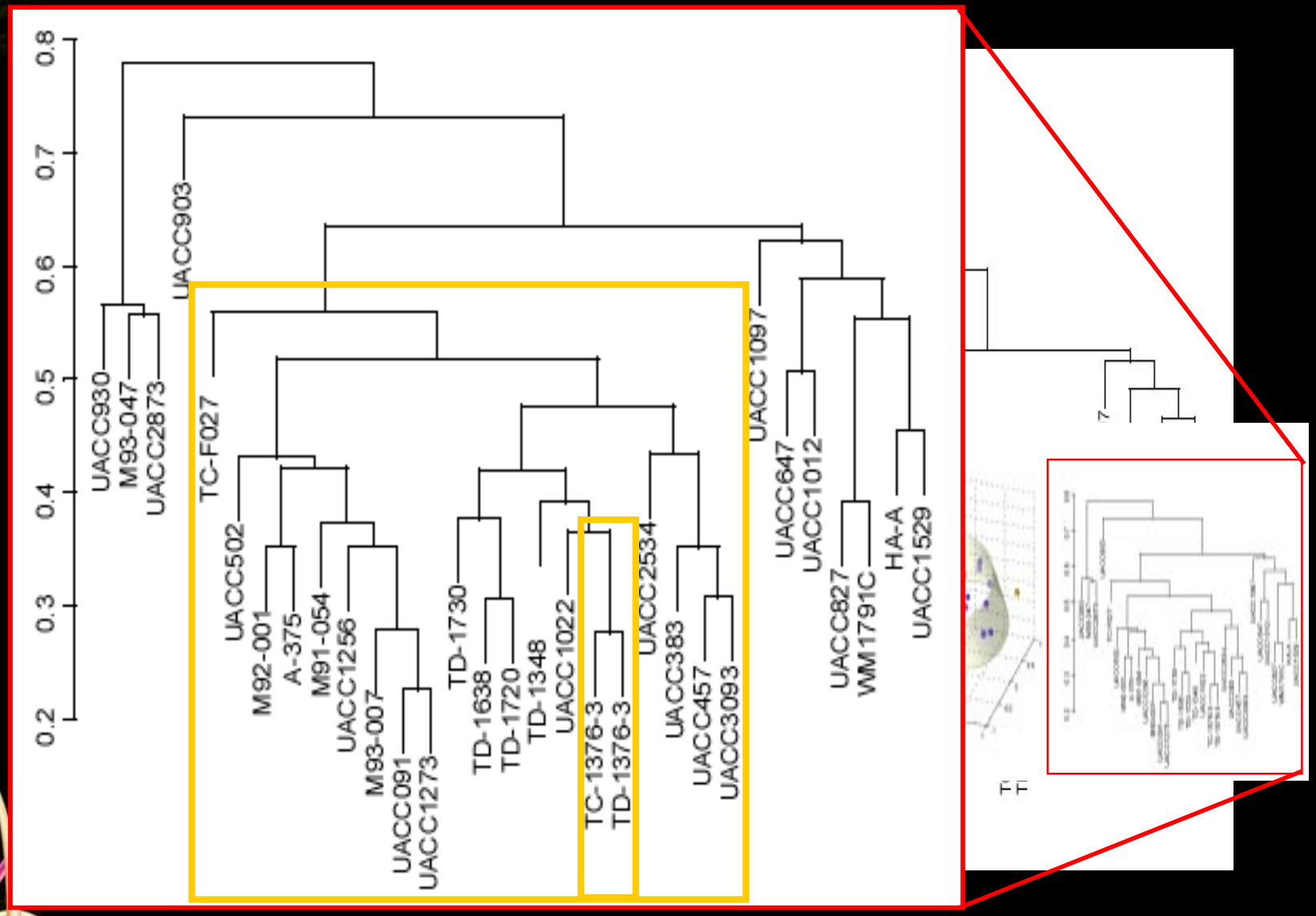
Expression Pattern Similarity



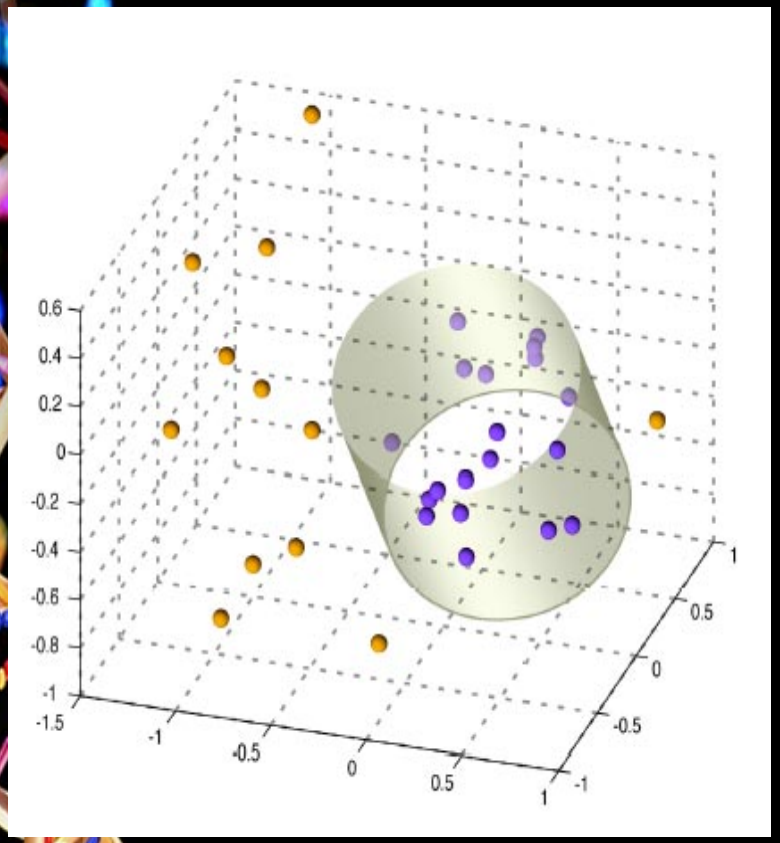
Expression Pattern Similarity

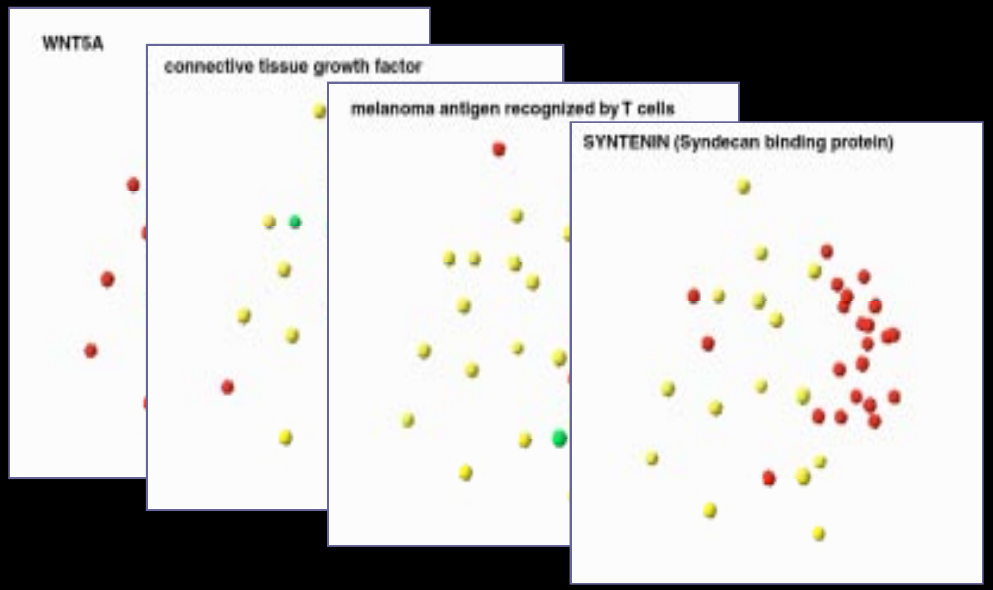


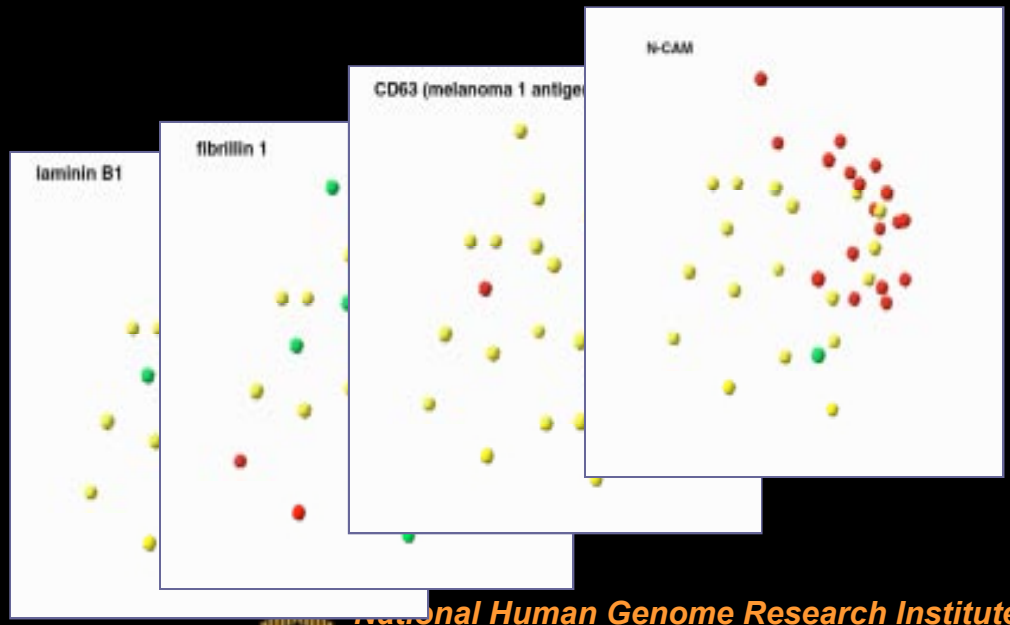
Expression Pattern Similarity

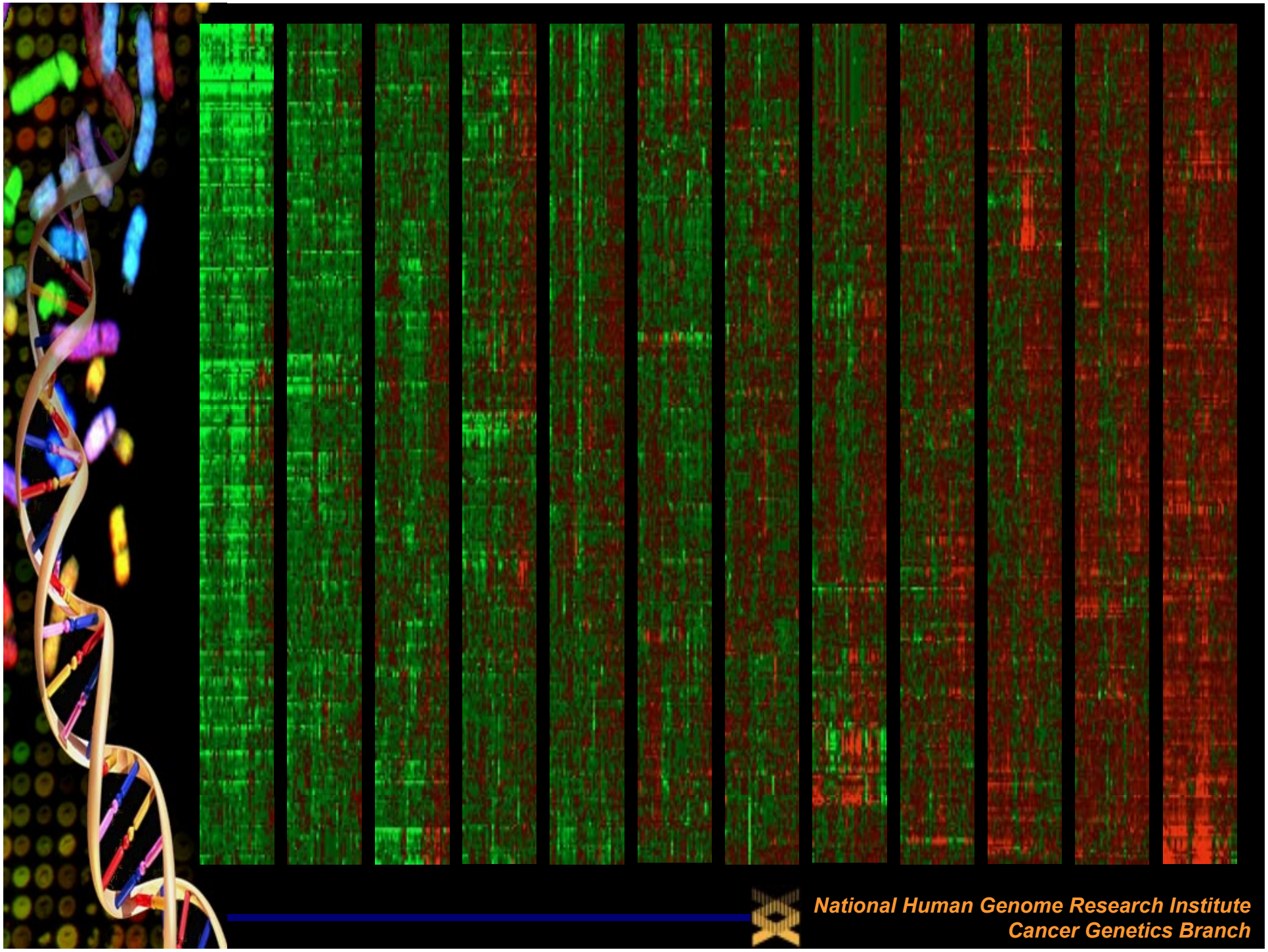


Visualization of Phenogram



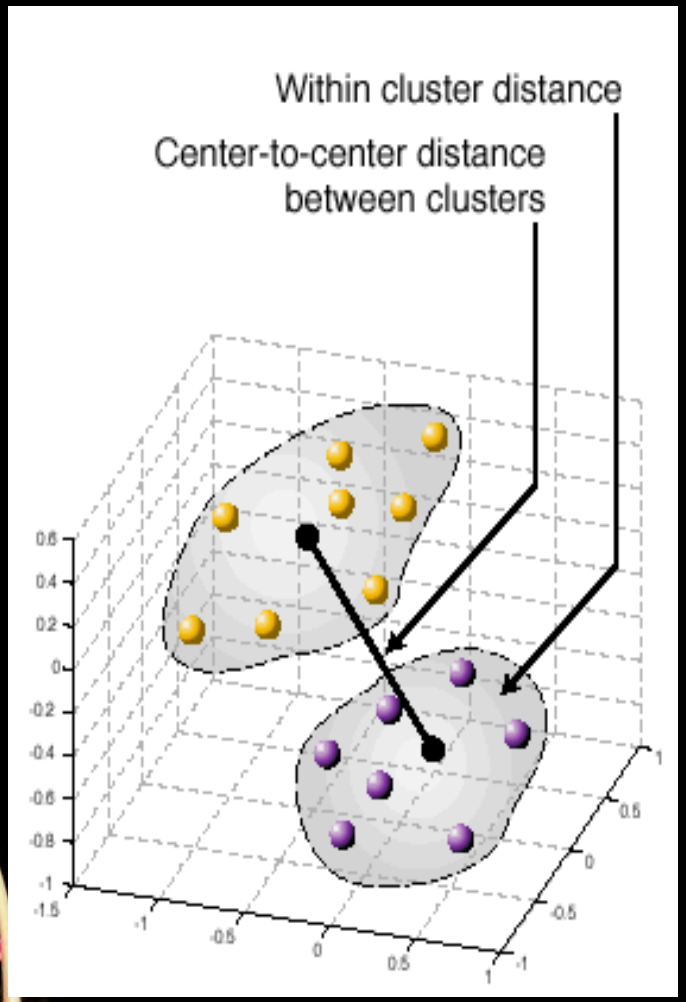






National Human Genome Research Institute
Cancer Genetics Branch

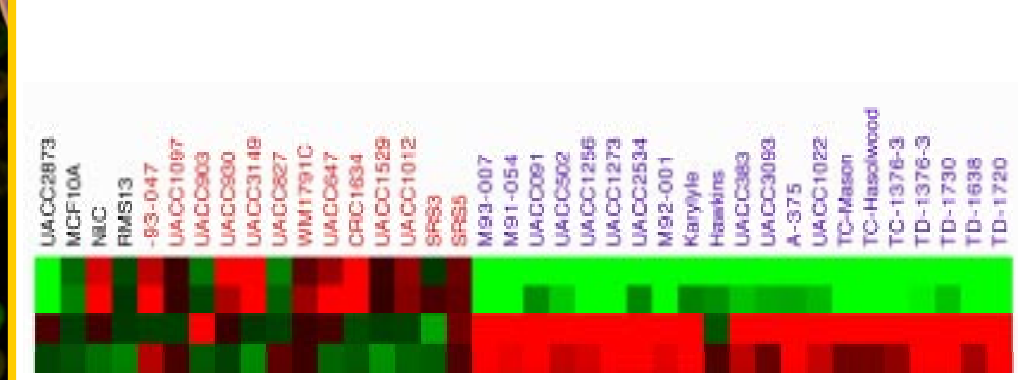
Sample Separation by Genes



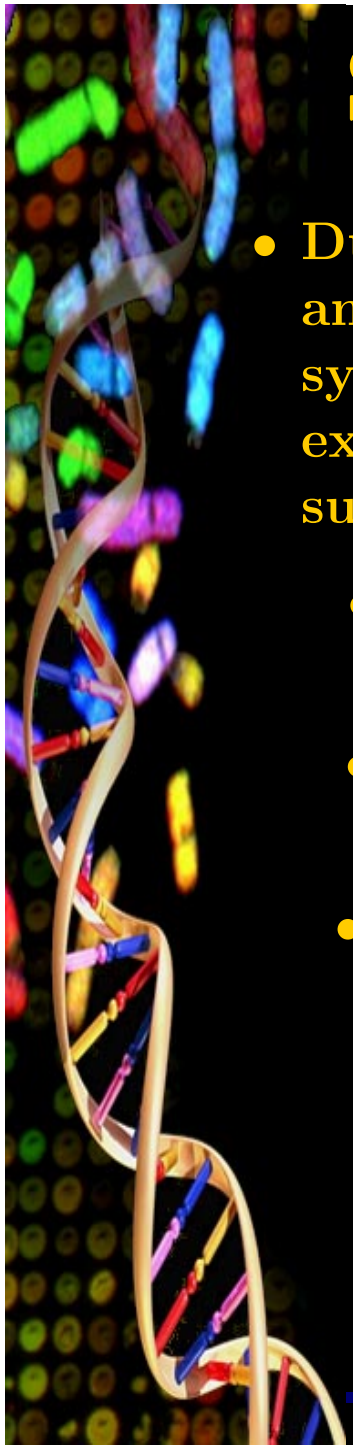
To evaluate each gene contribution for the given clustering result or a given classification defined by pathologist, several methods were developed,

- Weighted Gene Analysis - based on the ability of cluster separation.
- TNoM score - the number of misclassification by each gene
- Mutual Information, or Covariance between gene's expression level and its class assignment.

Weighted List



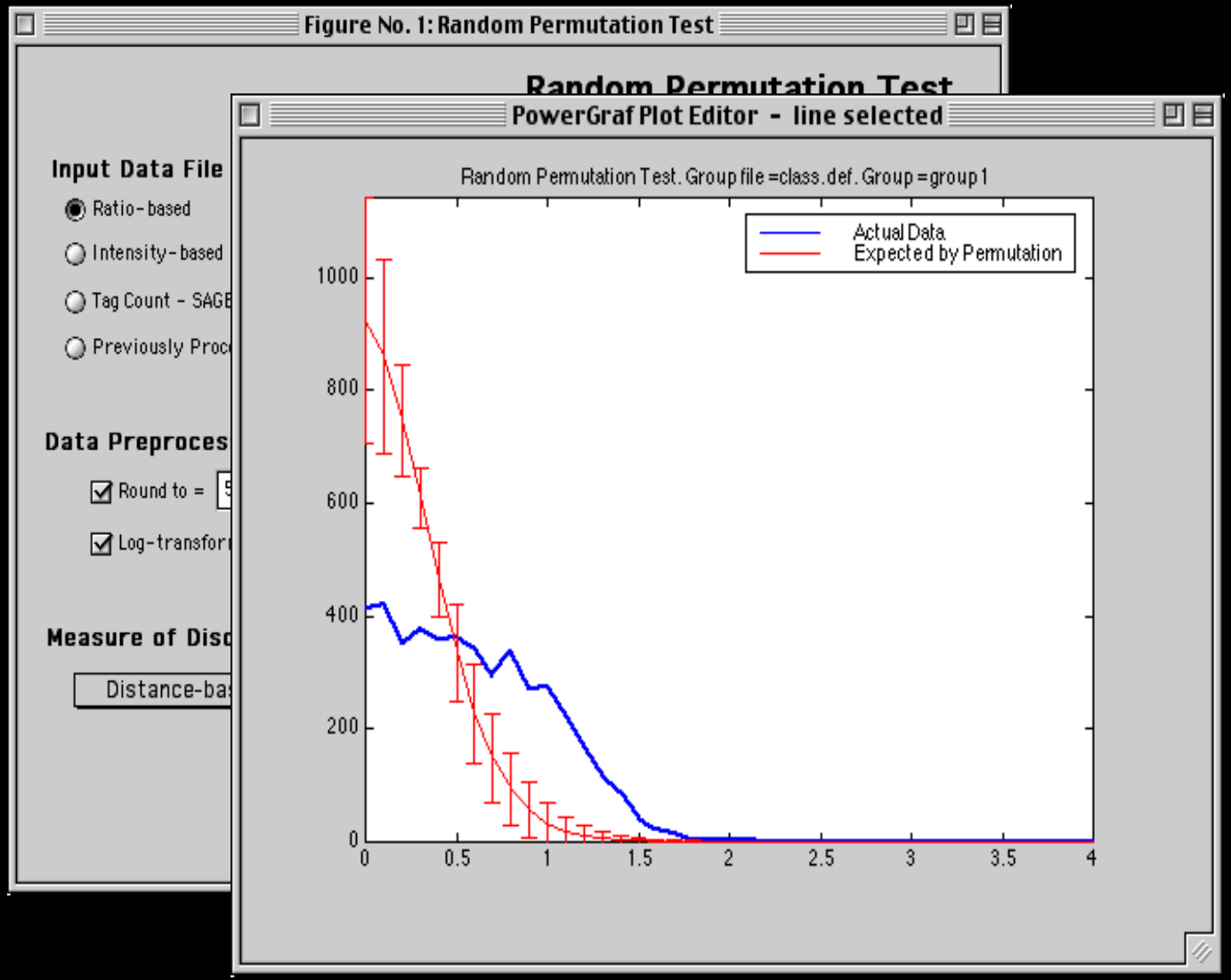
Weight	CloneID	Title
9.058926	357278	ESTs
6.20304	324901	wingless-type MMTV integration site family, member 5A
5.963354	266361	Human melanoma antigen recognized by T-cells (MART-1) mRNA
4.05084	234237	pirin



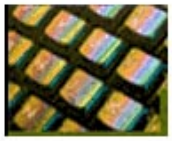
Statistical Validation

- Due to the intrinsic variability of gene expression level in any biological system, and due to imperfect detection system of gene expression level, any claim derived from expression similarity or dissimilarity has to be verified such that
  - It is far better than random chance - random Permutation.
  - It is insensitive to some noise levels - Noise perturbation
  - the claim is robust, and can be validated by other methods

Validation by Random Permutation

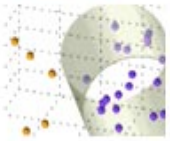


Access



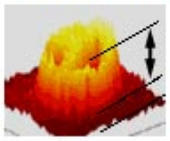
- Overview
- Login
- User Registration
- Upload Input Data
- Download Output
- Data Integration Resources
- Help Index
- Access Policy
- Disclaimers
- Logout

Analysis



- Overview
- Visualization
- Clustering
- Time Course Analysis
- Gene Selection
- Classification
- Image Analysis

HPC Analysis

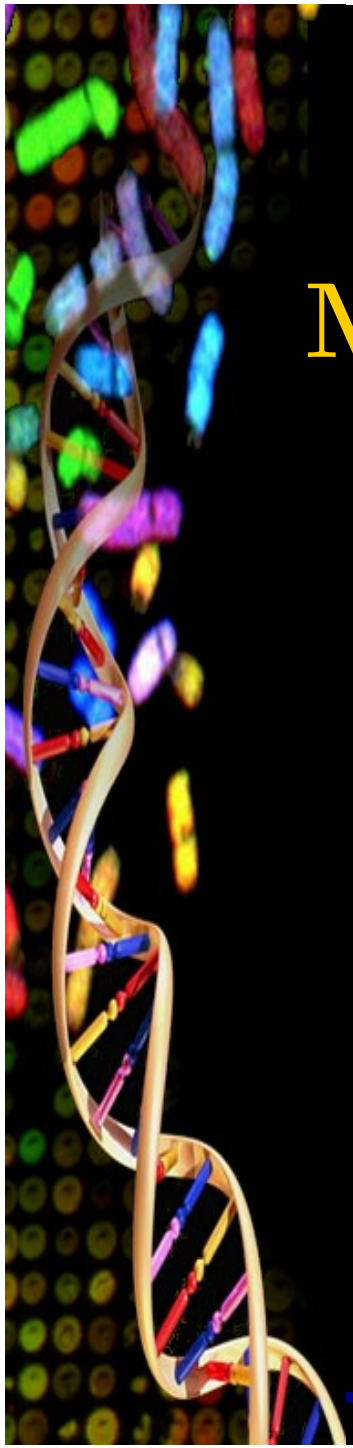


- Overview
- Predictive Network Analysis
- Robust Classification

Resources

- Overview
- Selected Publications
- Download Programs
- Related Links
- Site Map
- Index
- Search
- Credits / Contacts
- Comments
- QA / Forum

CIT
Center for Information Technology



Multivariate Measurement Of Gene Expression Relationship

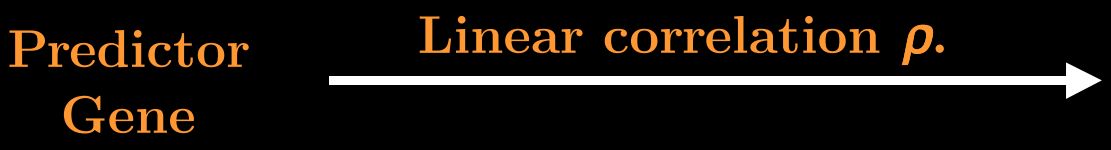
Seungchan Kim, E. Dougherty, Y. Chen, *et al*

Department of E.E., Texas A&M University

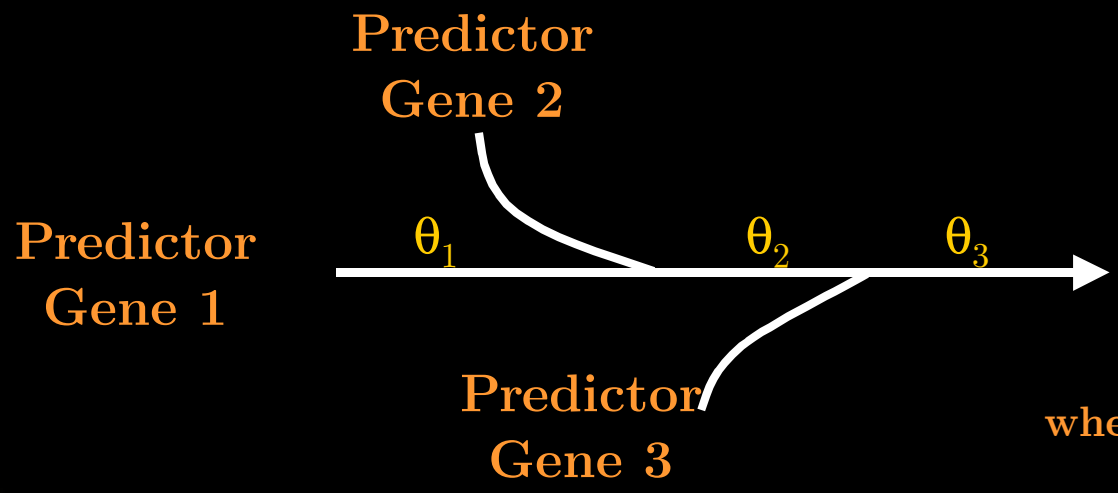
National Human Genome Research Institute
National Institutes of Health

Genomics, 67, 201-209 (2000)

Linear 1-to-1 relationship: Correlation (ρ).

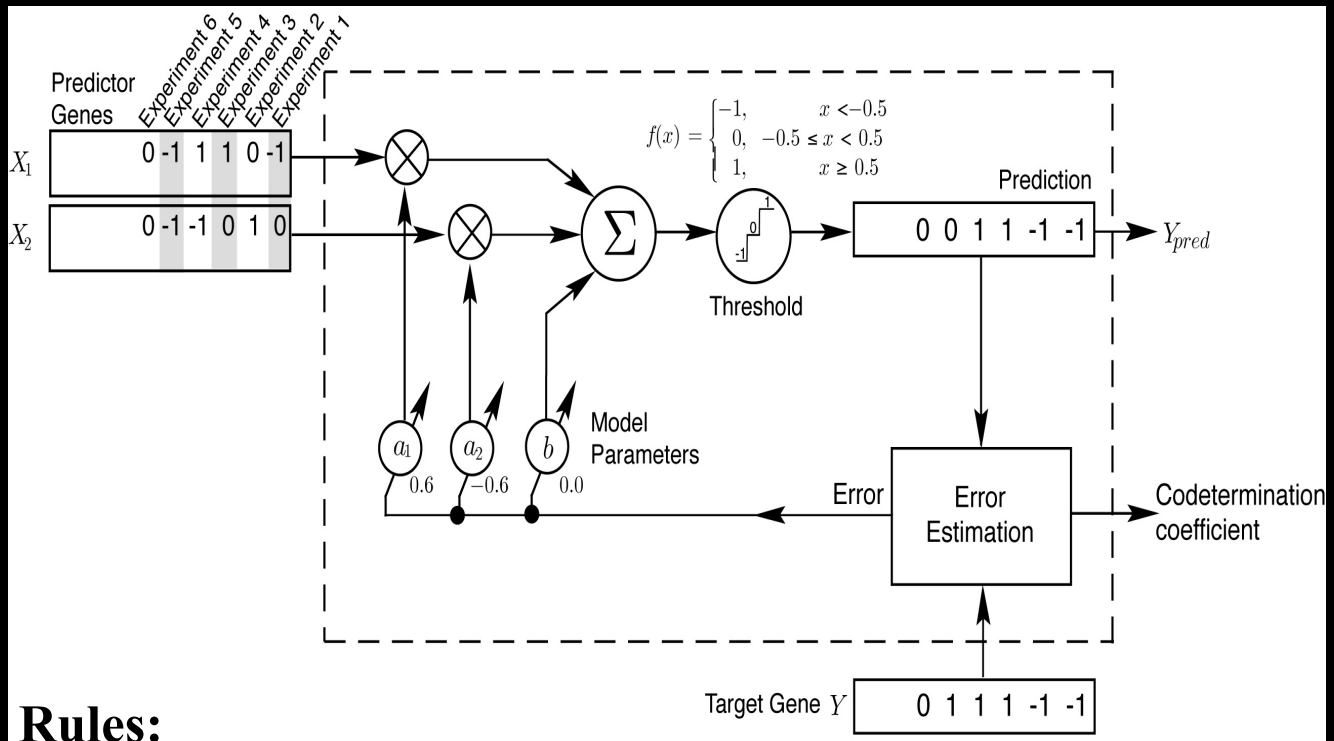


Linear n-to-1 relationship: Codetermination (CoD).



where $\theta_3 > \theta_2 > \theta_1$.

PREDICTIVE MODEL: TERNARY PERCEPTRON

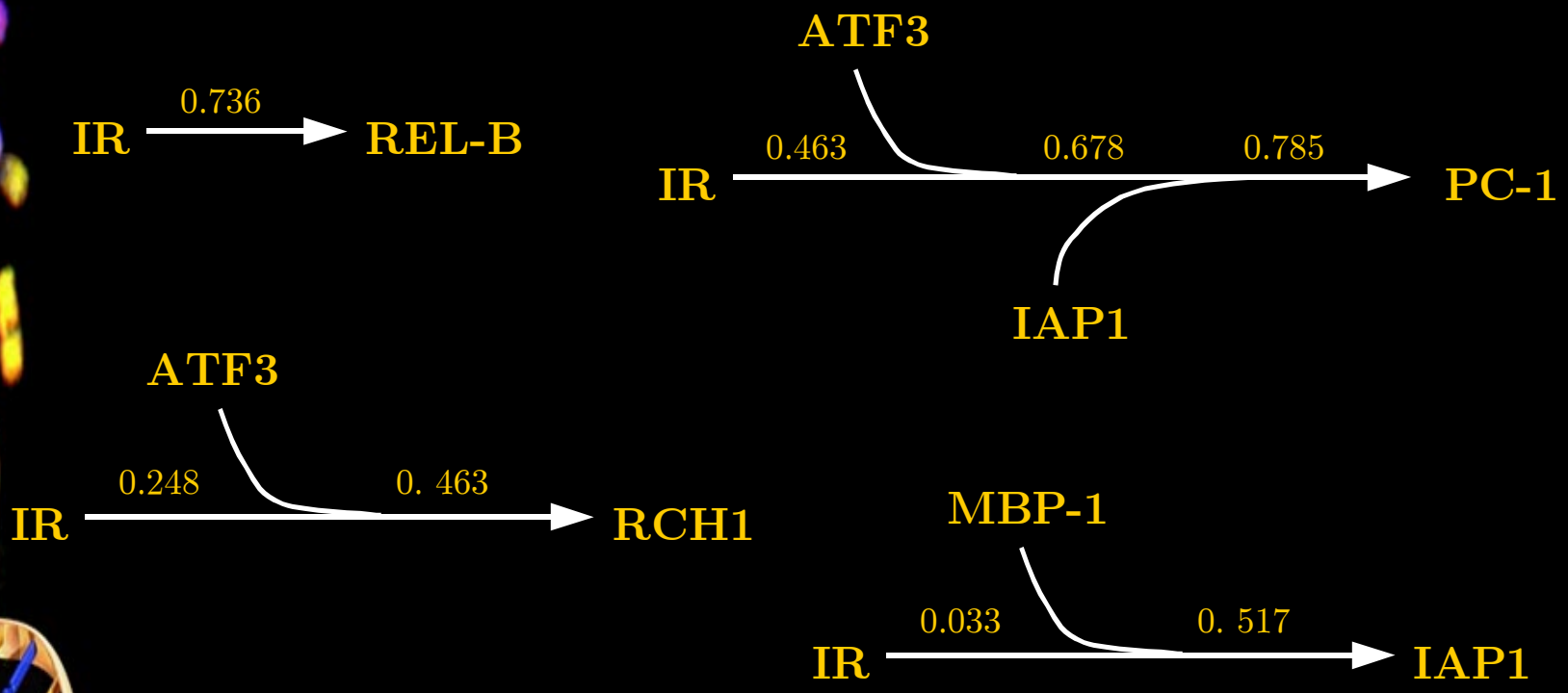


Rules:

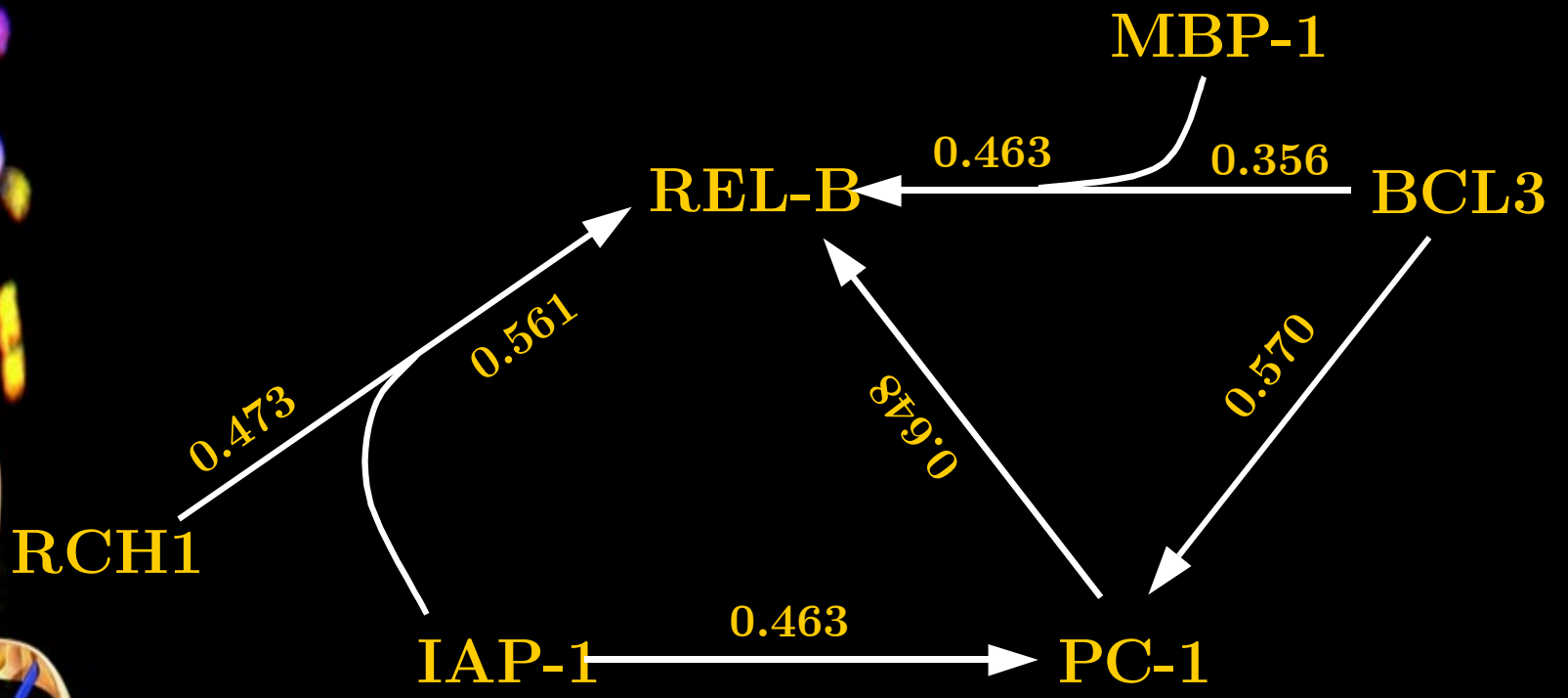
Ternary Logic table

Predictor		*Prediction*		*Square Error*
X_1	X_2	Y_{pred}	Y	$e^2 = (Y_{pred} - Y)^2$
−1	−1	0	1	1
−1	0	−1	−1	0
−1	1	−1	−1	0
0	−1	1	1	0
0	0	0	0	0
0	1	−1	−1	0
1	−1	1	1	0
1	0	1	1	0
1	1	0	0	0

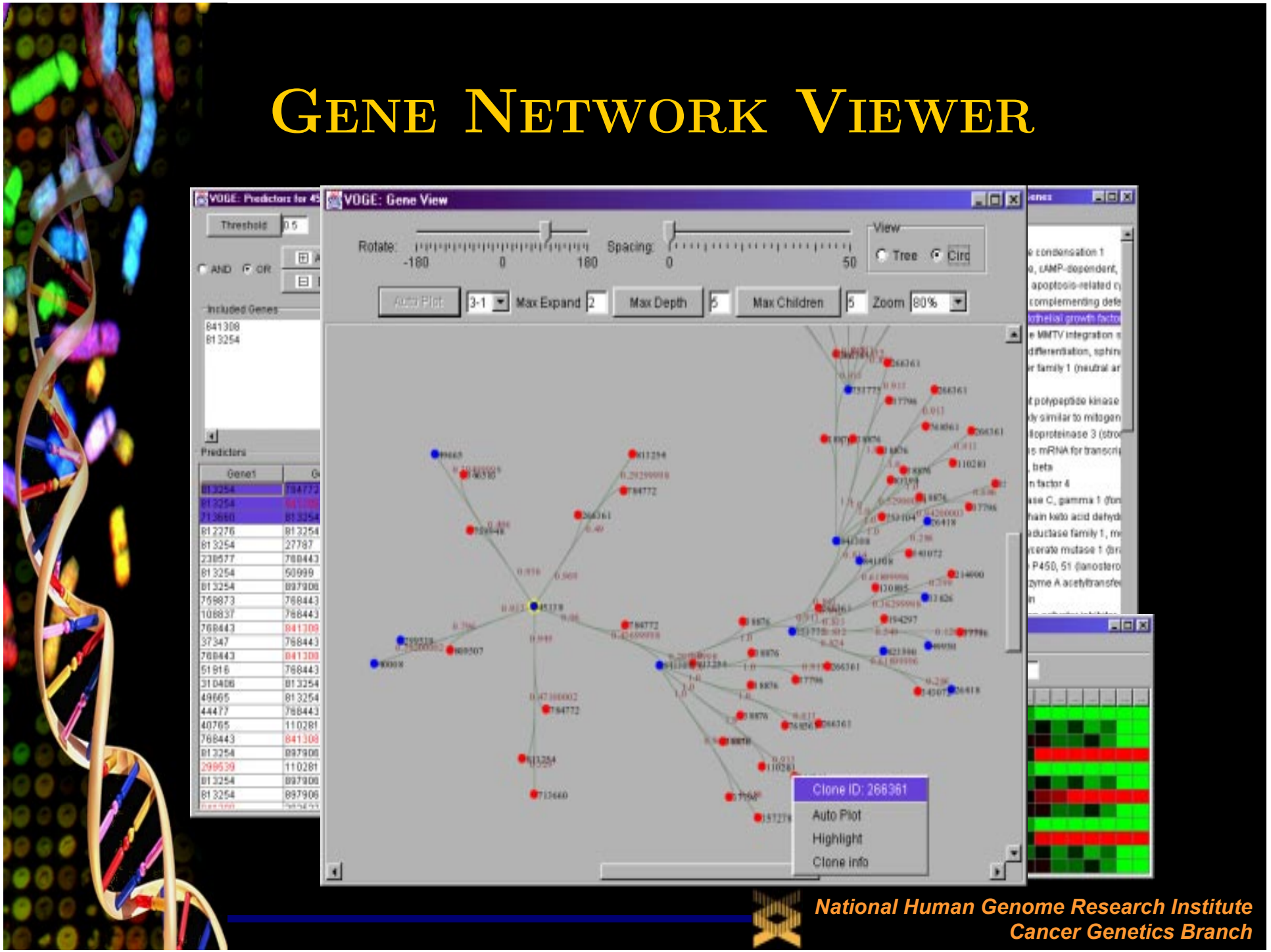
IR as a predictor



Mutual predictability



Gene Network Viewer



Summary

- **Microarray data analysis is a complicated process**
  - Visualization of expression data can reveal underlying order in large datasets.
  - The expression pattern could be related to a subset of cutaneous melanomas and we have successfully predicted aspects of their biological properties (Cancer Cell, Vol. 1, p.1).
  - Future challenges include links to genomic sequence and functional databases; ultimately predictive modeling.