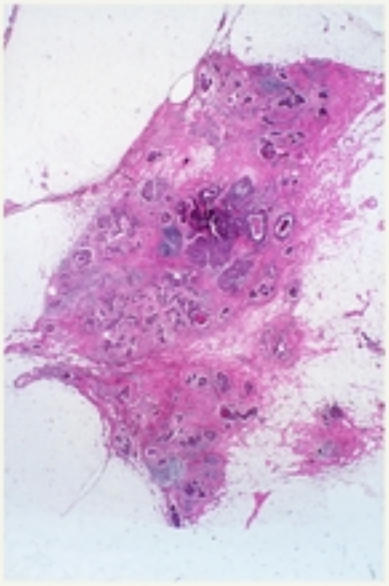
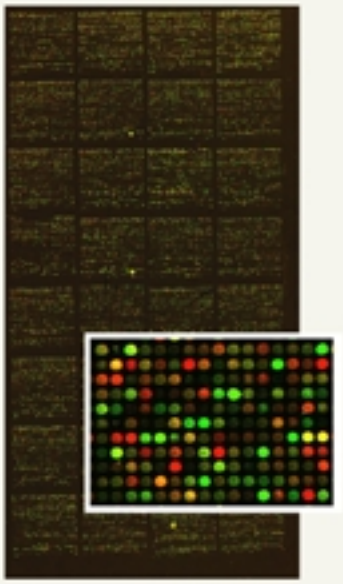
Two views of a tumor





The Connection Between Gene Expression and The State of a Biological System

Genomics and Cancer

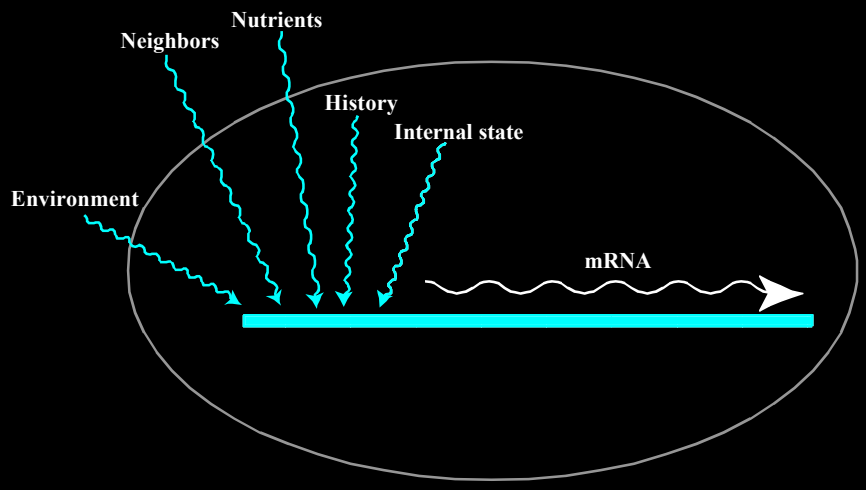
- Human tumors show great clinical heterogeneity, even within well-defined subgroups
- This clinical heterogeneity in tumors likely reflects unrecognized molecular heterogeneity in tumors
- We can characterize this molecular heterogeneity at the gene expression level with DNA arrays
- The logical connection between gene expression patterns and phenotype predicts a direct connection between gene expression patterns and their clinical phenotype

Towards a clinically relevant molecular taxonomy of cancer

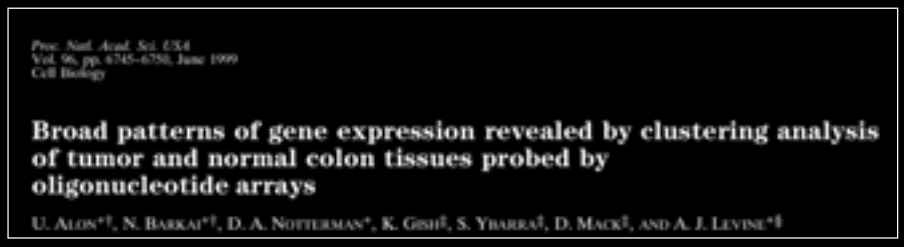
- Access archived clinical tumor samples taken at or near diagnosis from patients with well-characterized subsequent clinical histories
- Use DNA arrays to measure gene expression in these samples
- Look for new molecularly defined groups within or between previously recognized groups of tumors, especially groups with increased clinical homogeneity
- Look for direct associations between molecular and clinical properties of tumors

Gene Expression and the State of Biological Systems

Transcription Control Regions Integrate Complex Information Into Expression Level of Gene



Normal and tumor tissue have very different expression patterns

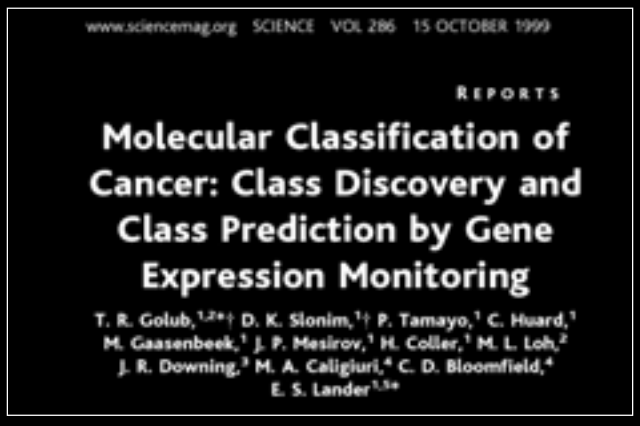


Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 6745–6750, June 1999
Cell Biology

Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays

U. Alon*†, N. Barkai*†, D. A. Notterman*, K. Gish‡, S. Ybarra‡, D. Mack‡, and A. J. Levine*§

Known tumor subclasses can be distinguished by their expression patterns



www.sciencemag.org SCIENCE VOL 286 15 OCTOBER 1999

REPORTS

Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring

T. R. Golub,[1,2*†] D. K. Slonim,[1†] P. Tamayo,[1] C. Huard,[1] M. Gaasenbeek,[1] J. P. Mesirov,[1] H. Coller,[1] M. L. Loh,[2] J. R. Downing,[3] M. A. Caligiuri,[4] C. D. Bloomfield,[4] E. S. Lander[1,5*]

Breast Cancer

Stanford/Norwegian Radium Hospital/NYU

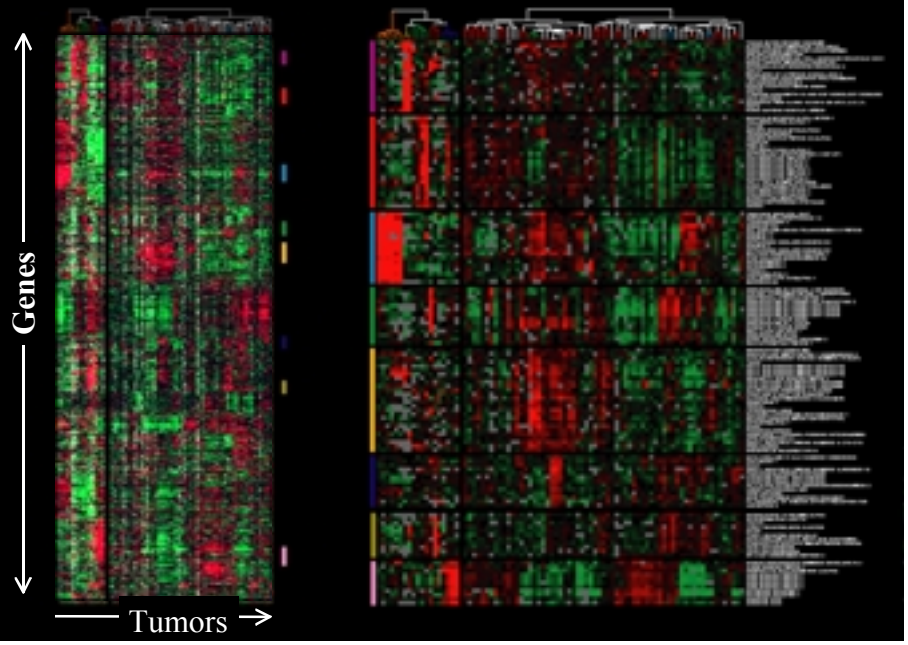
- Norway: Samples taken from same tumor at diagnosis and after 16 weeks of chemo and clinical followup
- Stanford: Tumors, lymph-node metastases and normal breast tissue, large and clinically heterogenous collection of archived samples with clinical followup

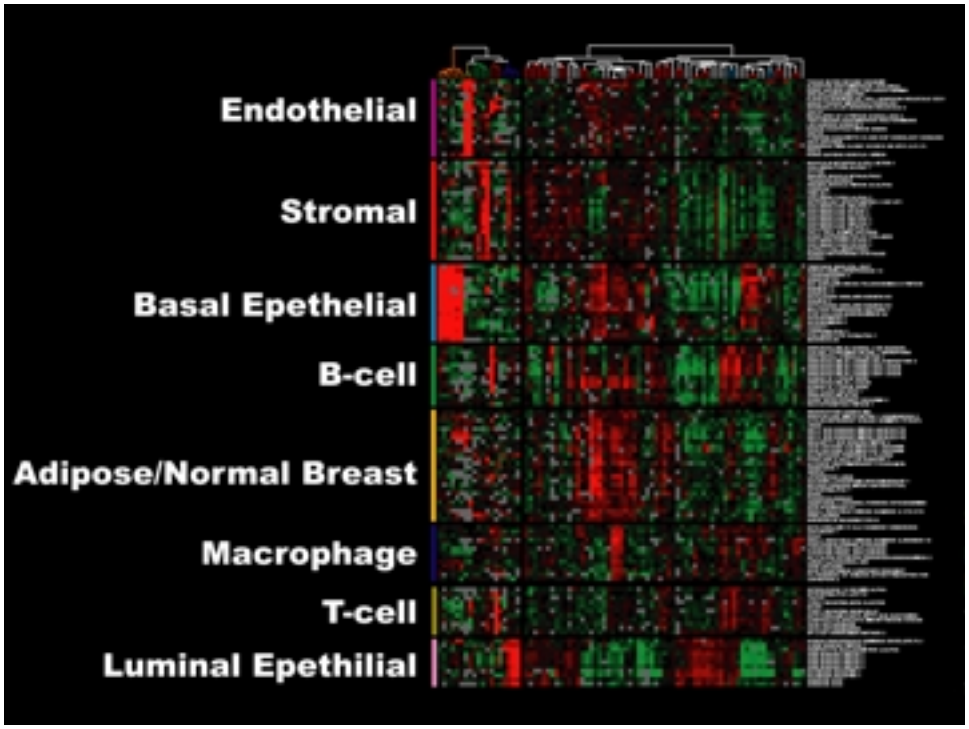
Breast Cancer: Stage I Array Studies

- **65 surgical specimens from 42 individuals (predominantly ductal carcinomas)**
- **20 before/after chemotherapy**
- **2 tumor/lymph node metastases pairs**
- **3 normal breast samples**
- **19 cell lines**

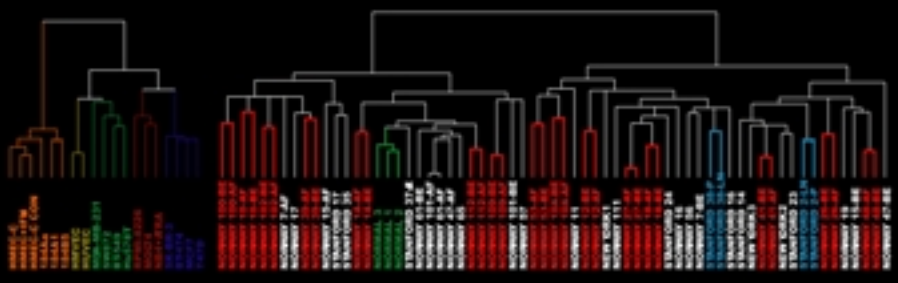
- **Perou et al., *Nature*, 17/8/2000.**

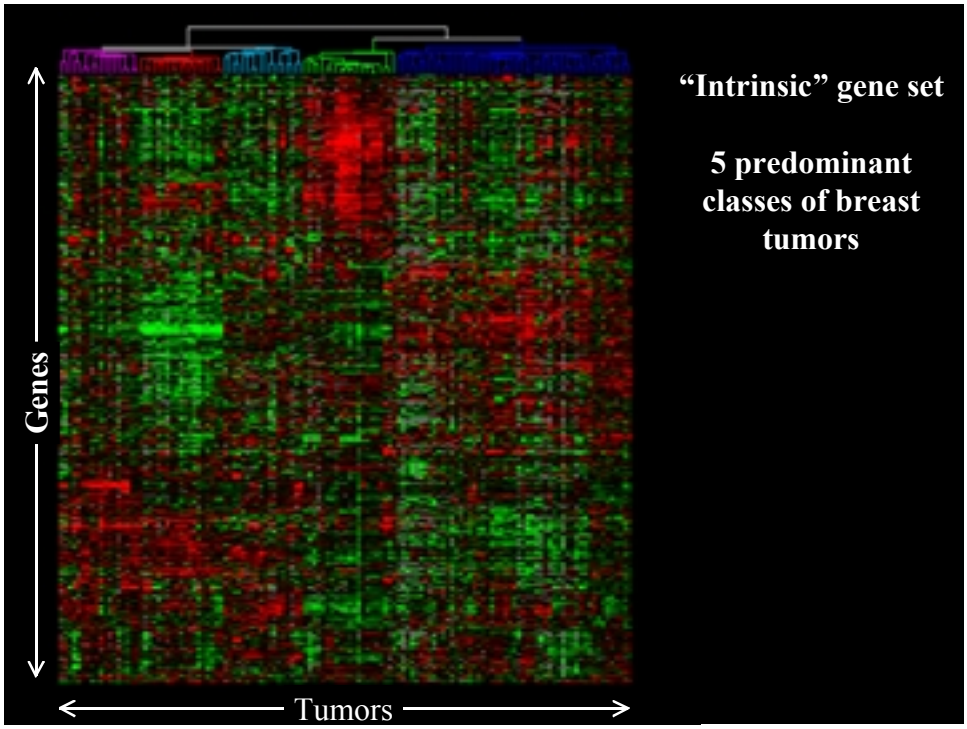
Multivariate Gene Expression Variation in Breast Tumors

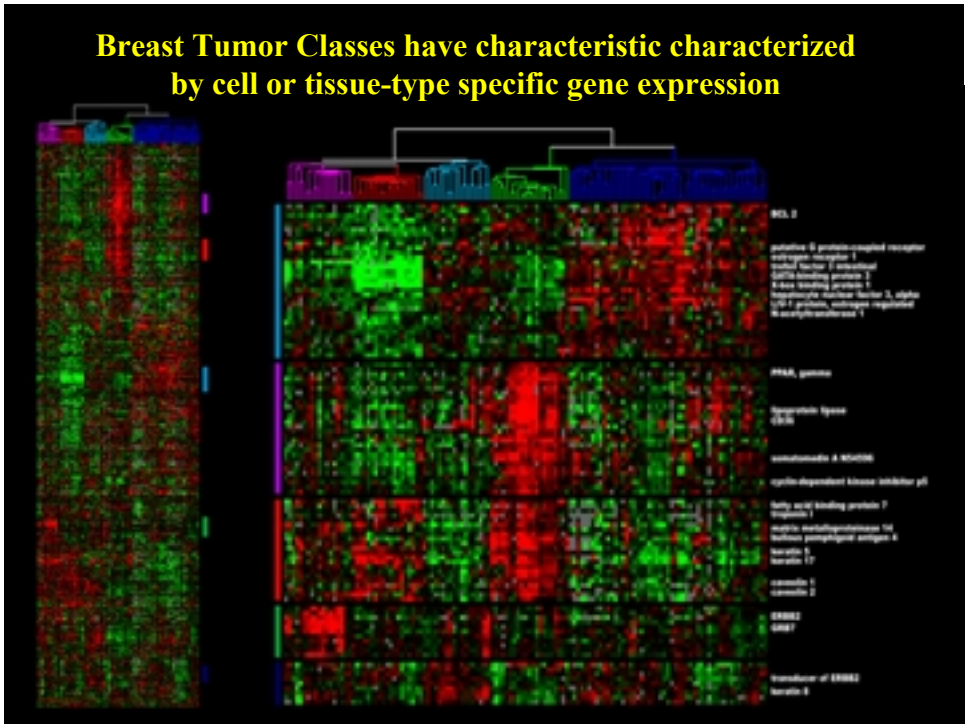


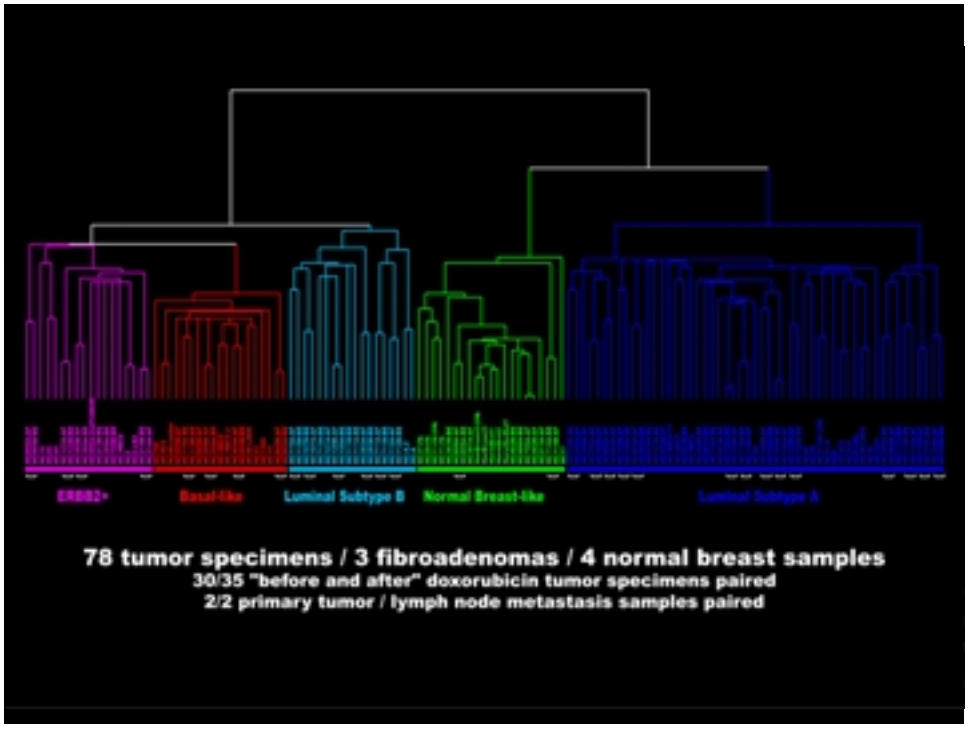


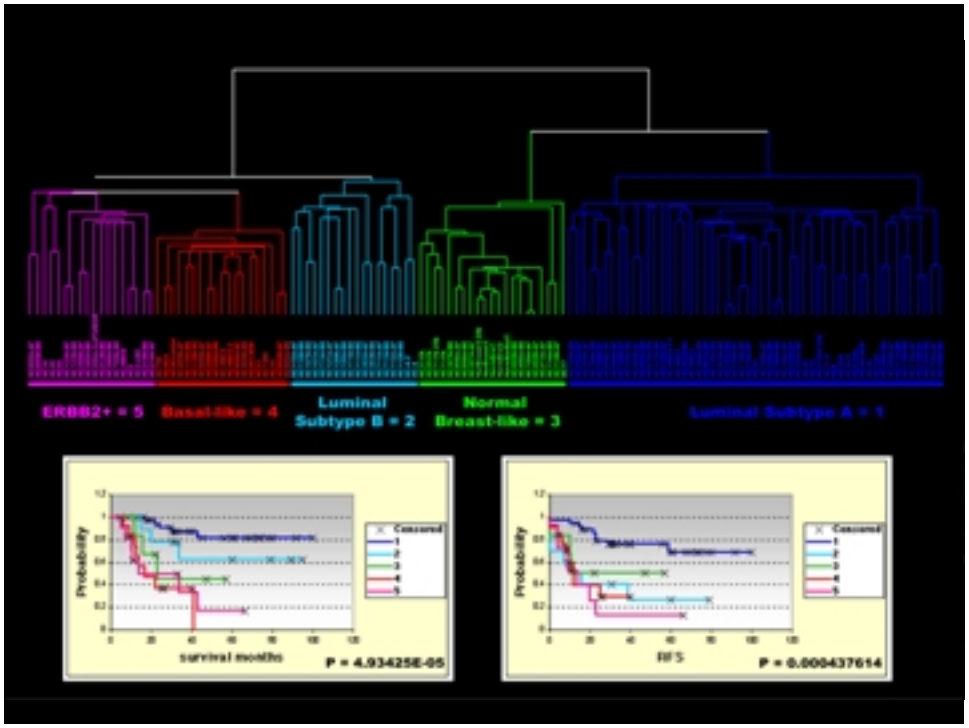
Pervasive Gene Expression Pattern of Tumors are Distinct and Consistent Over Time, Site and Following Chemotherapy



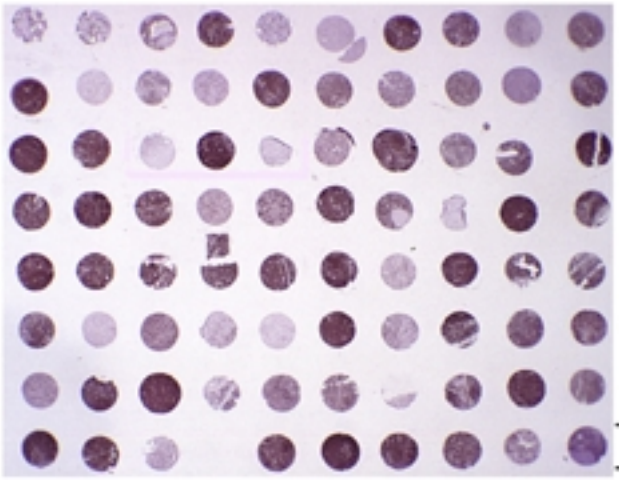

Genes
Tumors
"Intrinsic" gene set
5 predominant classes of breast tumors


Breast Tumor Classes have characteristic characterized by cell or tissue-type specific gene expression


ERBB2+
Basal-like
Luminal Subtype B
Normal Breast-like
Luminal Subtype A
78 tumor specimens / 3 fibroadenomas / 4 normal breast samples
30/35 "before and after" doxorubicin tumor specimens paired
2/2 primary tumor / lymph node metastasis samples paired


ERBB2+ = 5
Basal-like = 4
Luminal Subtype B = 2
Normal Breast-like = 3
Luminal Subtype A = 1
Probability
survival months
P = 4.93425E-05
Probability
RFS
P = 0.000437614

Human Tissue Microarrays



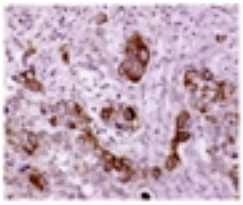
600 mm

BCL2 staining

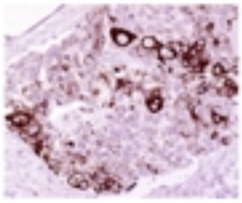
O. Kallioniemi

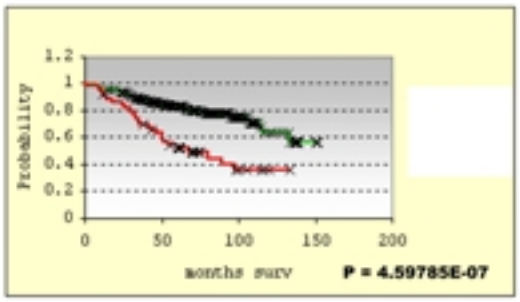
Human Breast Tumor Microarray (600 specimens)

BC790 with Keratin 17



BC790 with Keratin 5





keratin 5 and 17 negative
keratin 5 and/or 17 positive

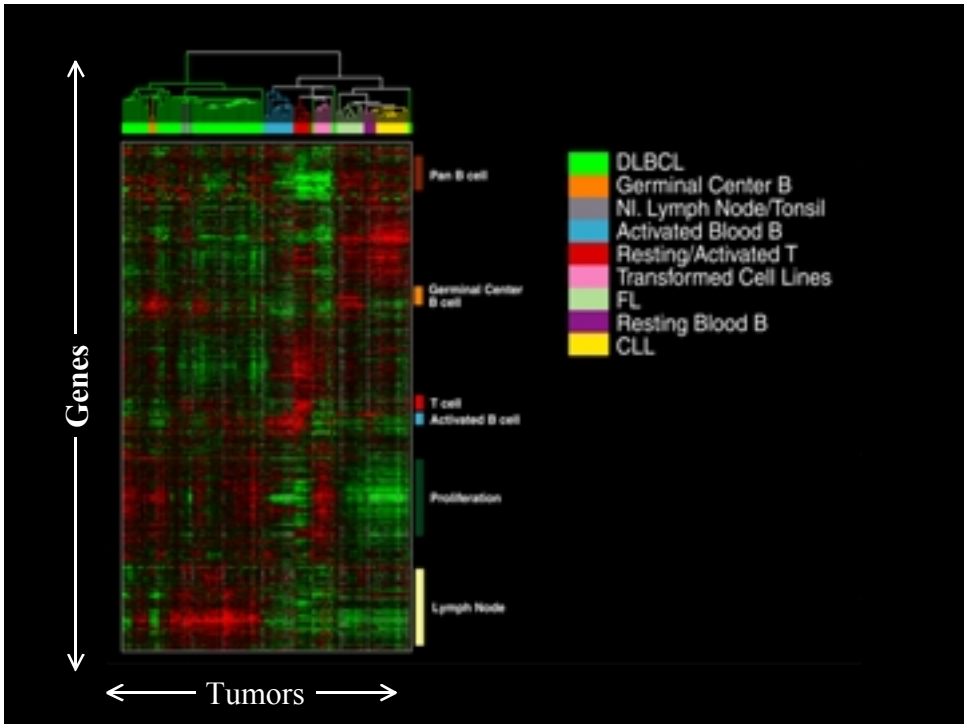
Diffuse Large B-cell Lymphoma
Stanford/NCI/University of Nebraska

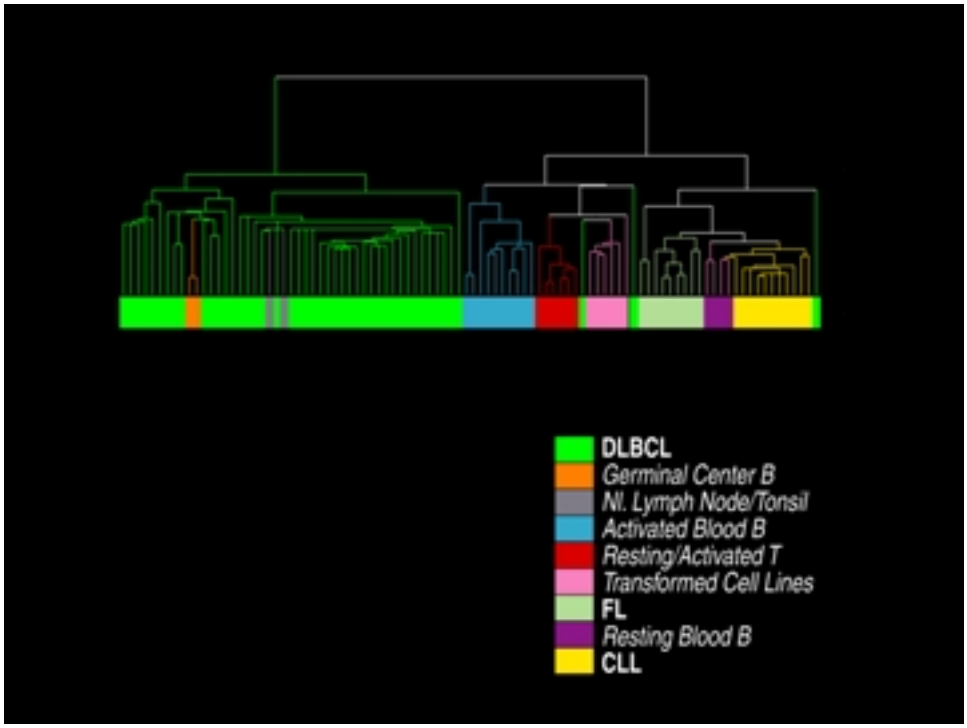
- Most common form of non-Hodgkin's lymphoma (~40%)
- Treated by combination chemotherapy regime
- Although most patients respond initially, only 50% achieve durable remission; the remainder succumb to the disease relatively rapidly
- Sub-classification has been unsuccessful

DLBCL Stage I Array Studies

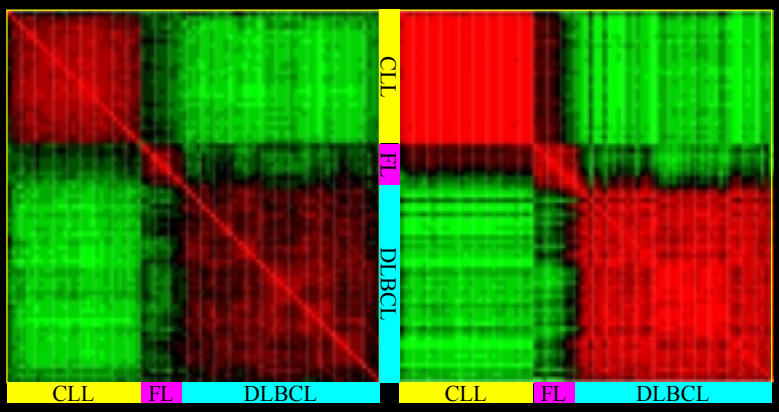
- **Lymphoid targeted microarray (LymphoChip) with 18,000 cDNA's representing > 10,000 genes**
- **42 DLBCL samples**
- **11 CLL and 9 FL samples**
- **GC B-cells, tonsils, resting and activated B and T cells and transformed DLBCL cell-lines**

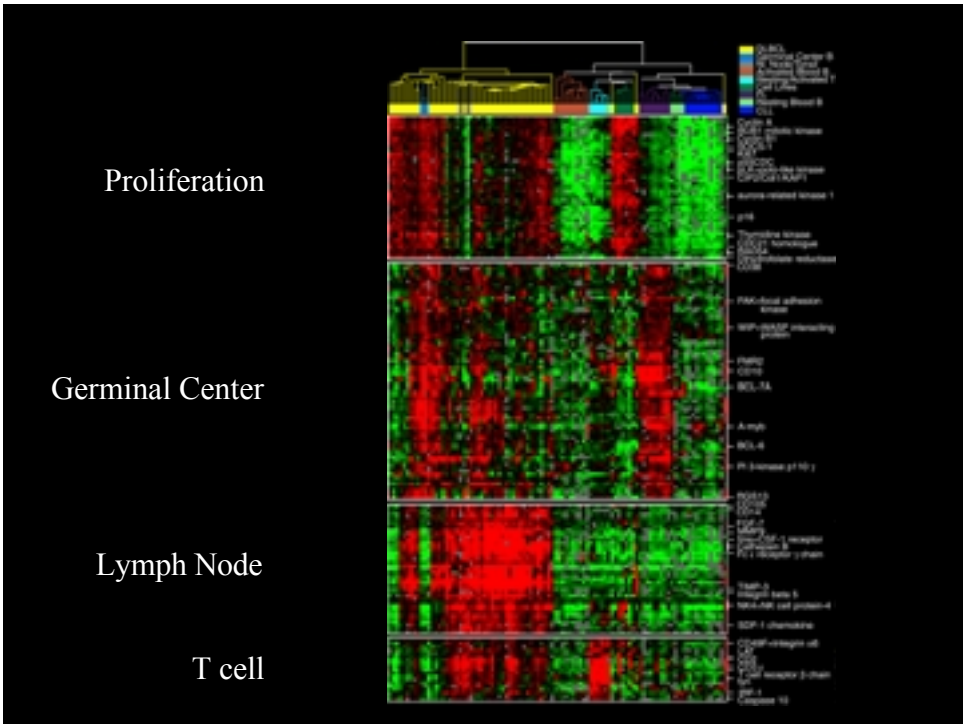
- **Alizadeh et al., *Nature*, 13/2/2000.**

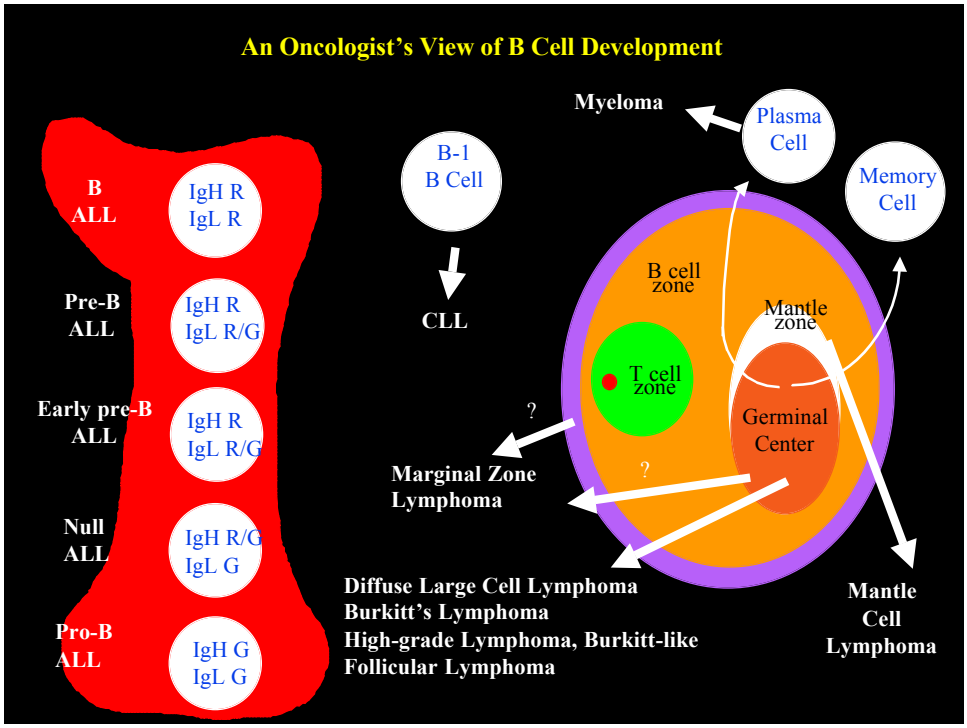

Genes
Tumors
Pan B cell
Germinal Center B cell
T cell
Activated B cell
Proliferation
Lymph Node
DLBCL
Germinal Center B
NI. Lymph Node/Tonsil
Activated Blood B
Resting/Activated T
Transformed Cell Lines
FL
Resting Blood B
CLL

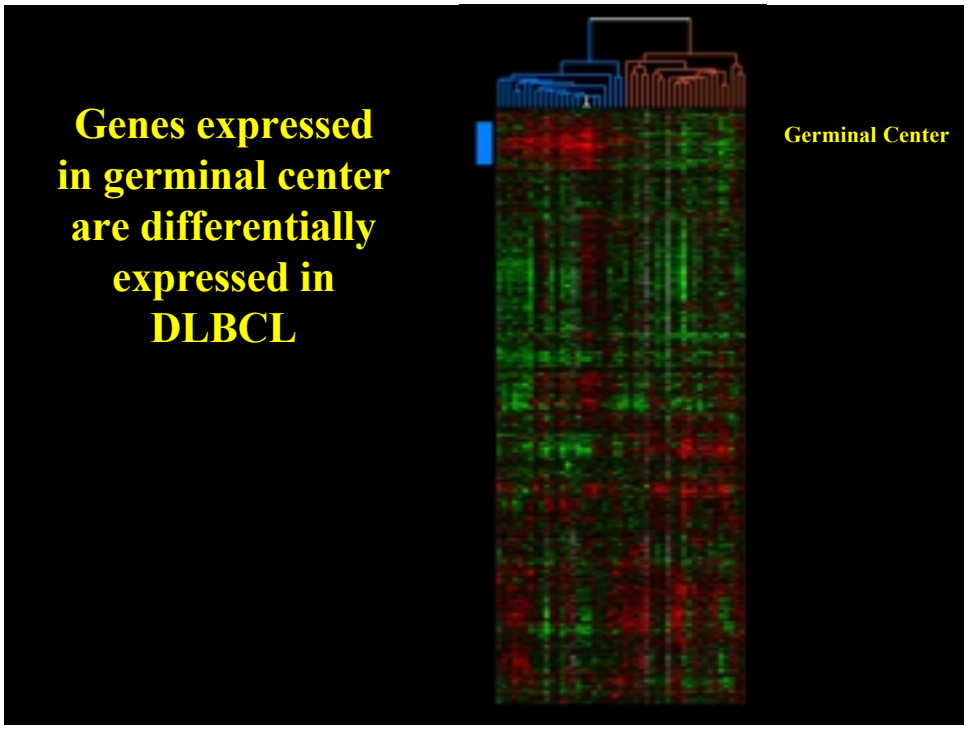

DLBCL
Germinal Center B
NI. Lymph Node/Tonsil
Activated Blood B
Resting/Activated T
Transformed Cell Lines
FL
Resting Blood B
CLL

CLL, FL and DLBCL are Readily Distinguished on the Basis of Gene Expression Patterns

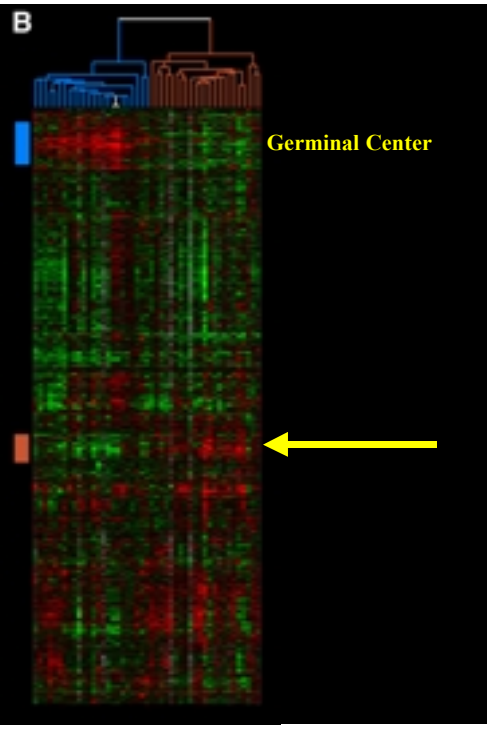




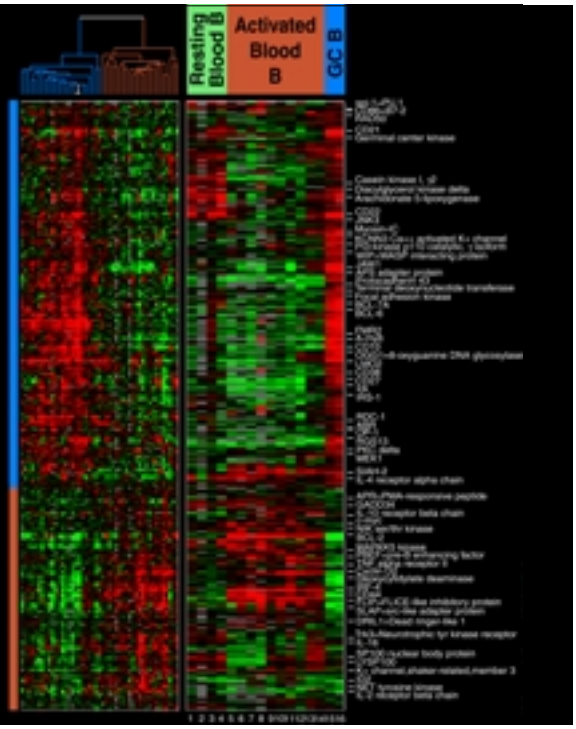

An Oncologist's View of B Cell Development
B ALL
IgH R IgL R
Pre-B ALL
IgH R IgL R/G
Early pre-B ALL
IgH R IgL R/G
Null ALL
IgH R/G IgL G
Pro-B ALL
IgH G IgL G
B-1 B Cell
CLL
Myeloma
Plasma Cell
Memory Cell
B cell zone
Mantle zone
T cell zone
Germinal Center
?
?
Marginal Zone Lymphoma
Diffuse Large Cell Lymphoma
Burkitt's Lymphoma
High-grade Lymphoma, Burkitt-like
Follicular Lymphoma
Mantle Cell Lymphoma


Genes expressed in germinal center are differentially expressed in DLBCL
Germinal Center

A set of genes is poorly expressed in samples with high-level expression of germinal center genes and highly expressed in the complement

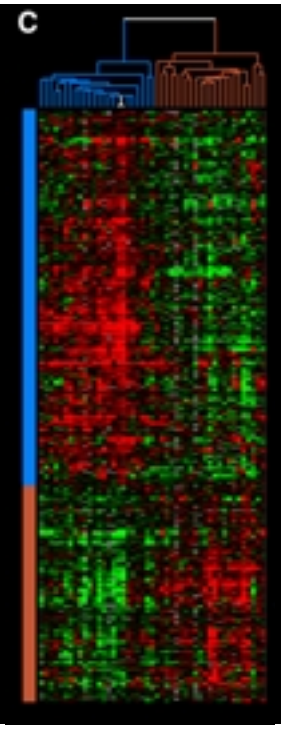


GCB complement genes are activated during in vitro activation of blood B cells

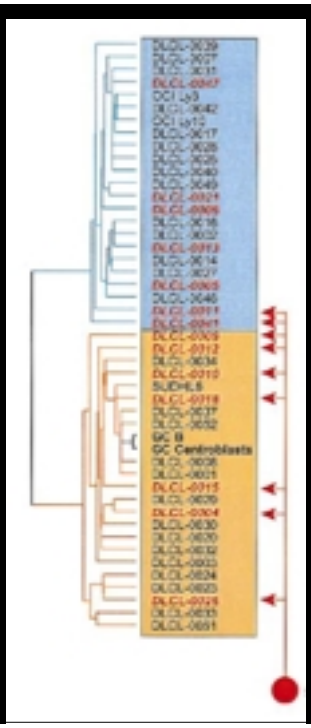


Classification of DLBCL based on expression of germinal center and activated blood B-cell genes

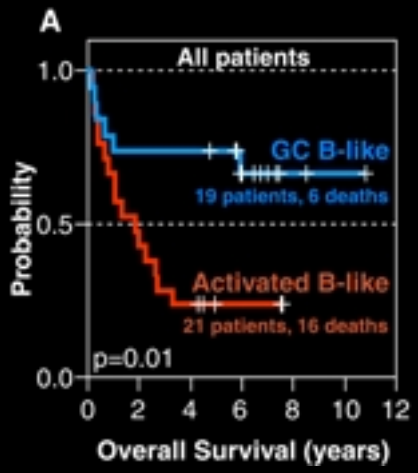
GCB-like DLBCL
ABC-like DLBCL



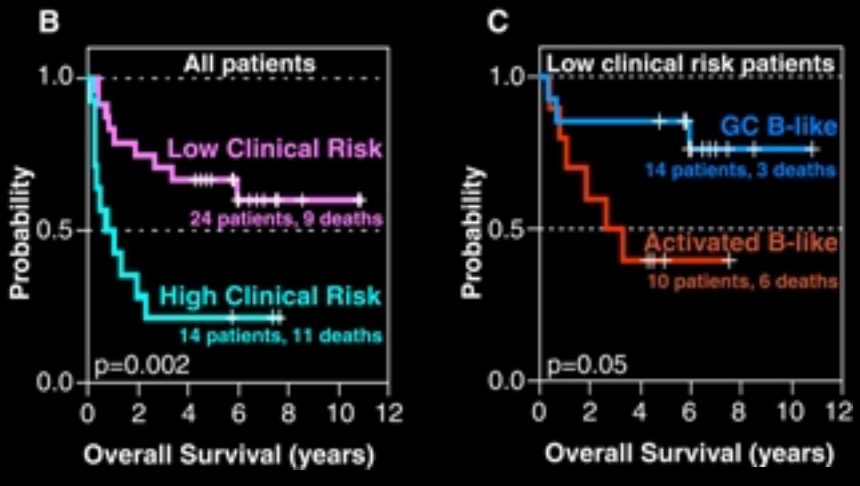
Ongoing somatic hypermutation in DLBCL



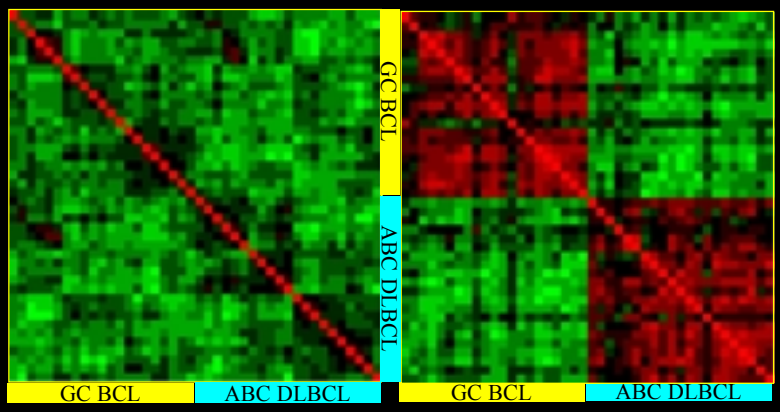
GC/ABC Distinction Associated with Significant Differences in Outcome



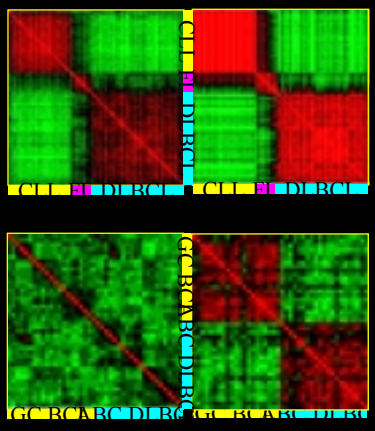
GC/ABC distinction is distinct and complementary to best previously available prognosticators



GC/ABC Distinction NOT Dominant in Data
But Well Supported By Data



Some problems are easy
Some are hard

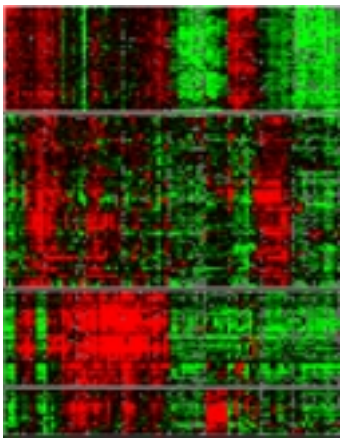


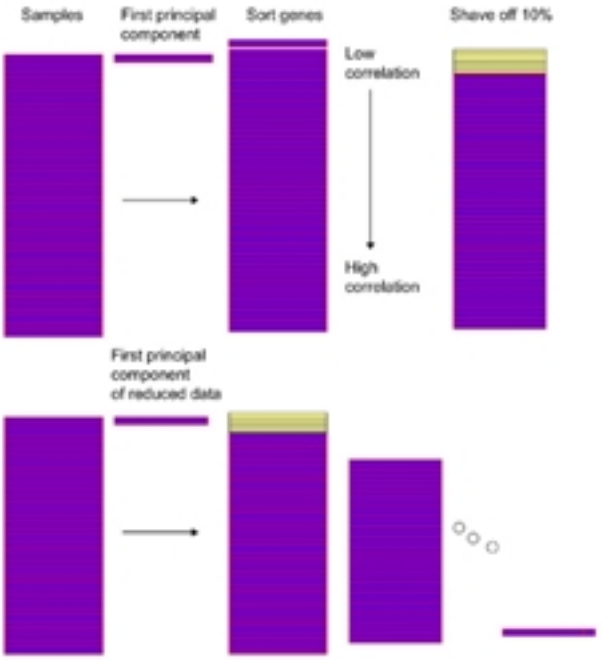
What Next

- Collect much more data
- Better integration with clinical databases
- Further analysis of the relationship between gene expression and phenotype. How valid is the concept of a tumor taxonomy? Is every tumor a unique entity best understood as a function of its own expression pattern.

Gene Shaving

- Goal is to find groups of genes that are coherent and have high variance across dataset





$$S_N \supset S_k \supset S_{k_1} \supset S_{k_2} \supset \cdots \supset S_1$$

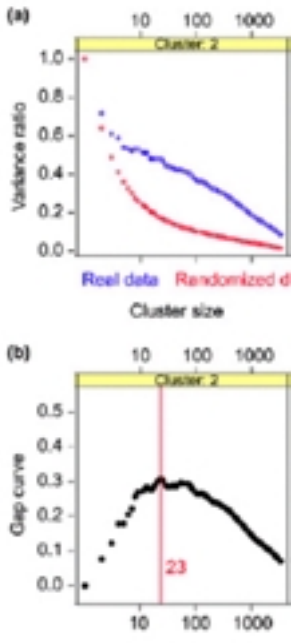
Cluster Quality Measure

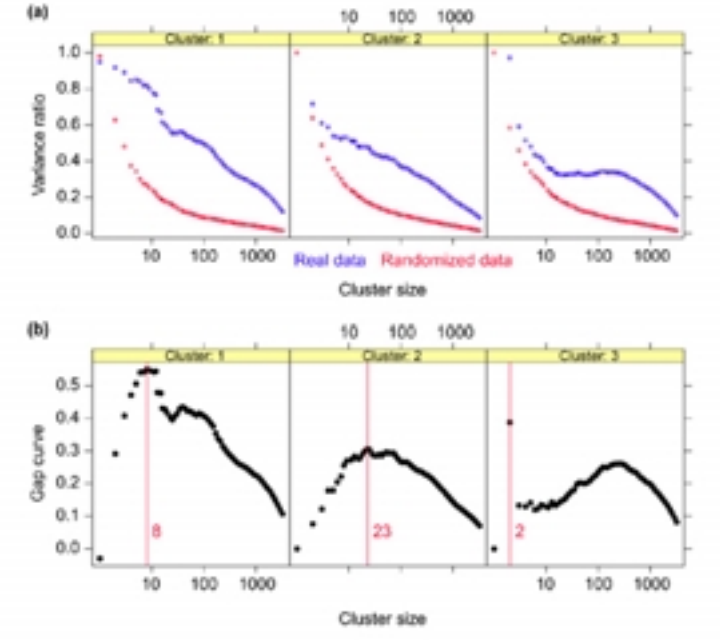
$$R^2 = 100\frac{V_B}{V_T} = \frac{\frac{V_B}{V_W}}{1 + \frac{V_B}{V_W}}$$

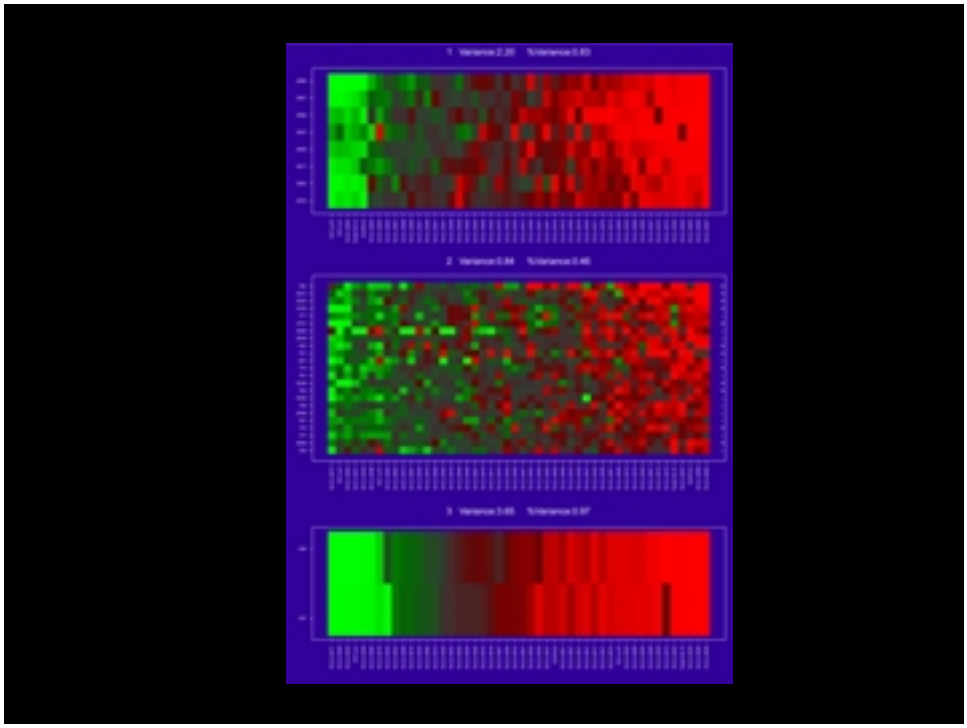
$$S_N \supset S_k \supset S_{k_1} \supset S_{k_2} \supset \cdots \supset S_1$$

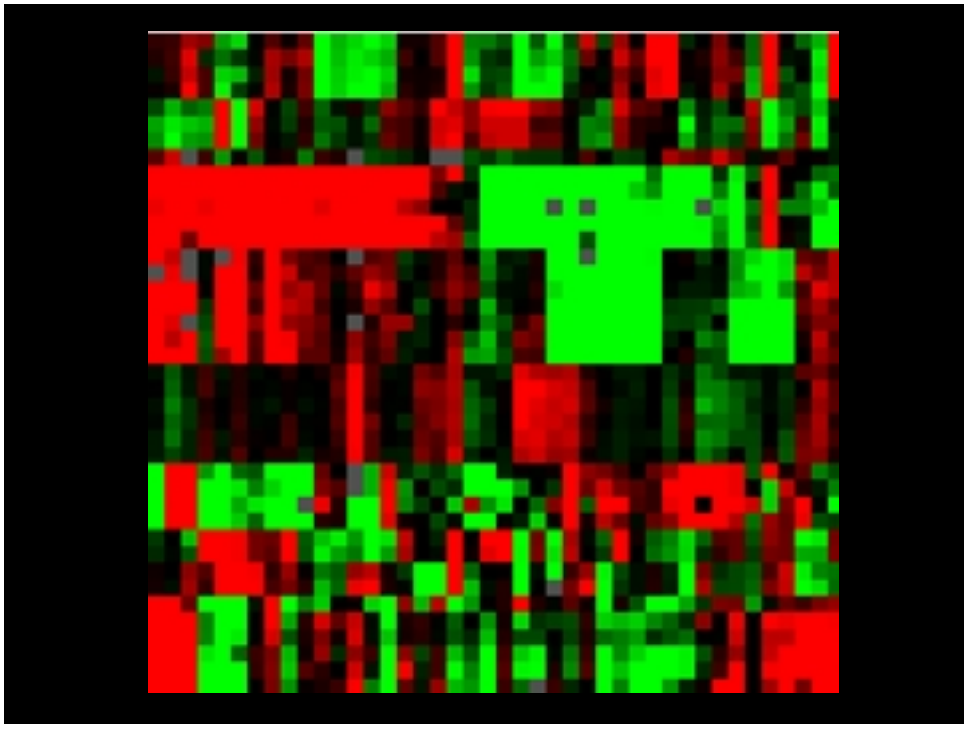
$$\text{Gap}\,(k) = D_k - \bar{D}_k^*$$

$$\hat{k} = \text{argmax}_k\,\text{Gap}(k)$$




(a)
Cluster: 1
Cluster: 2
Cluster: 3
Variance ratio
Real data
Randomized data
Cluster size
(b)
Gap curve
8
23
2
Cluster size



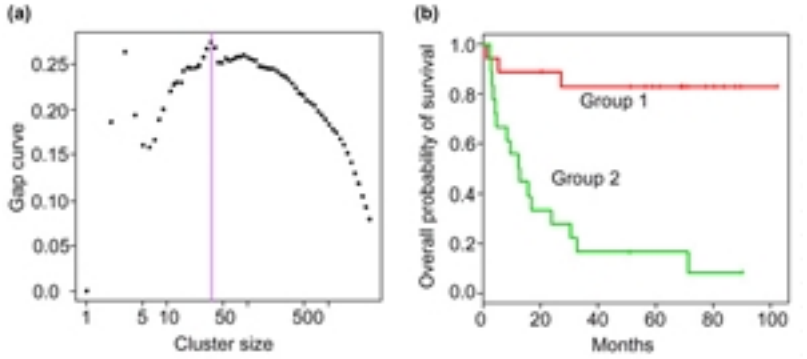


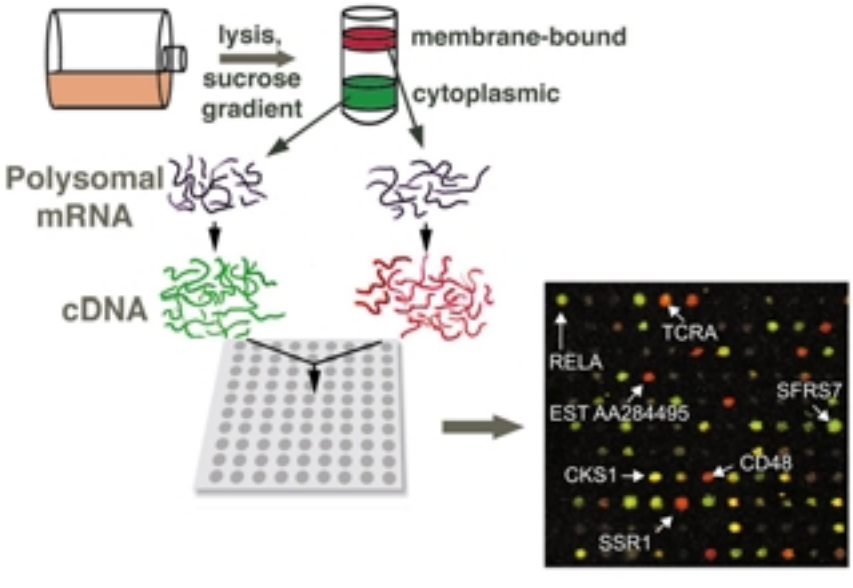
Supervised Shaving

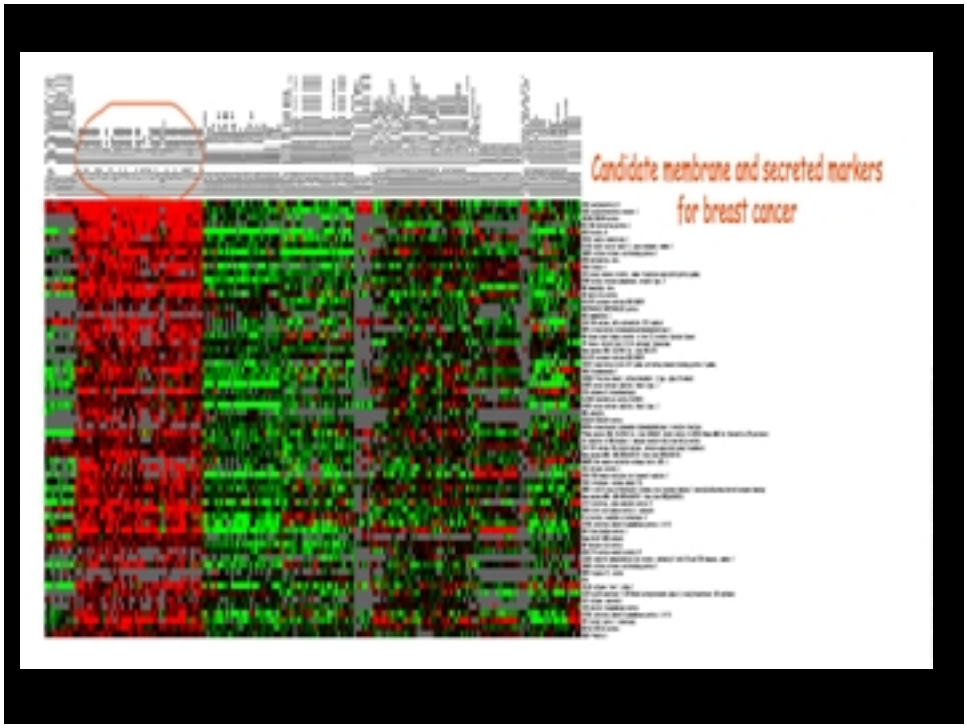
Auxiliary information $y = (y_1, y_2, \ldots y_p)$

$$\max_{S_k}[(1-\alpha) \bullet Var(\bar{x}_{S_k}) + \alpha \bullet J(\bar{x}_{S_k}, y)]$$

$$0 \leq \alpha \leq 1$$


(a)
Gap curve
0.25
0.20
0.15
0.10
0.05
0.0
1
5
10
50
500
Cluster size
(b)
Overall probability of survival
1.0
0.8
0.6
0.4
0.2
0.0
Group 1
Group 2
0
20
40
60
80
100
Months


lysis,
sucrose
gradient
membrane-bound
cytoplasmic
Polysomal
mRNA
cDNA
RELA
TCRA
EST AA284495
SFRS7
CKS1
CD48
SSR1


Candidate membrane and secreted markers
for breast cancer