



A Physicist's Approach to Breast Cancer

Gyan Bhanot

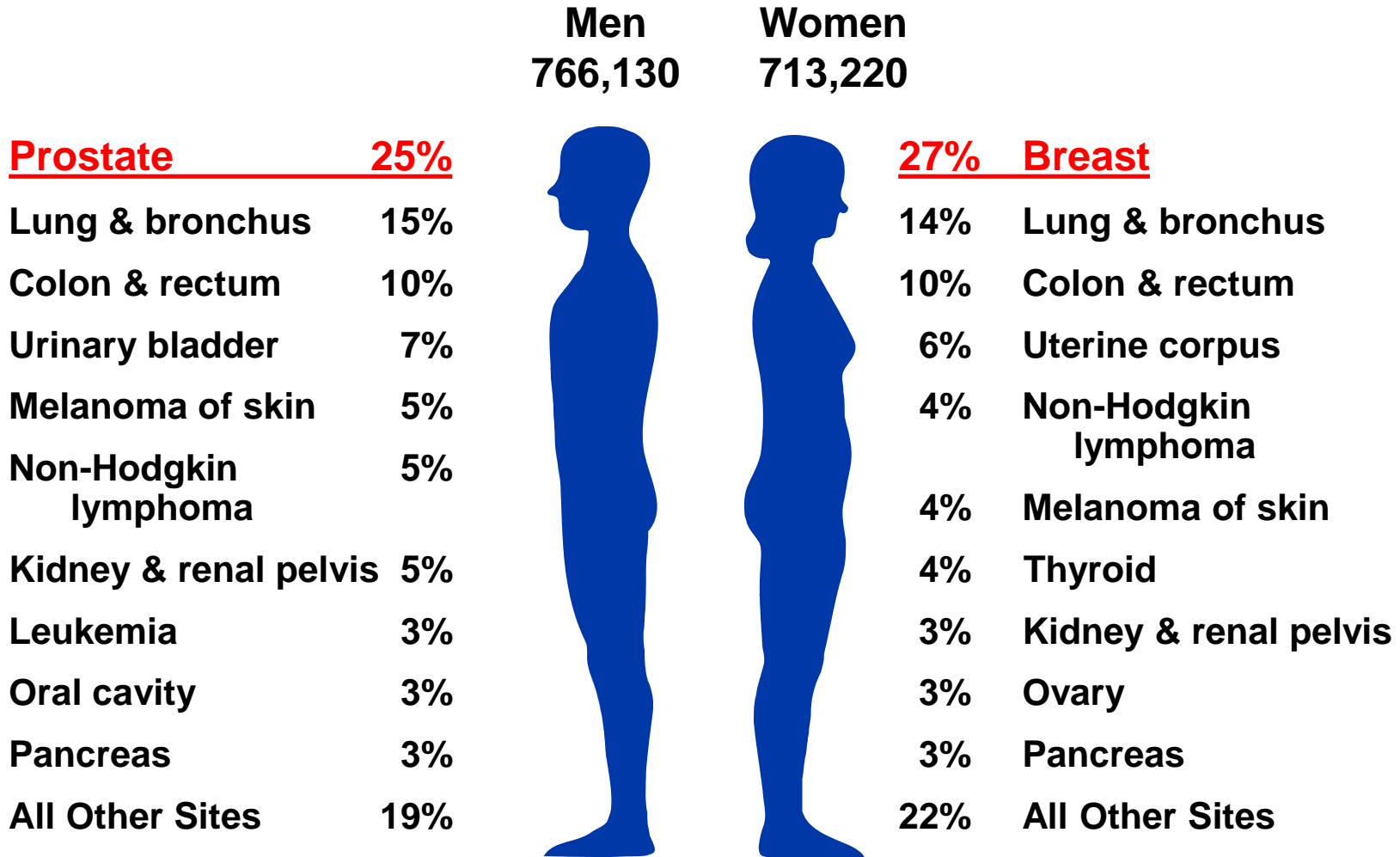
Rutgers/CINJ

The three laws

- Work with Clinicians – they know more than you
- **Cancer = Death.** Focus on improving outcome
- Be humble. Your ego has no place in this

KNOW THE ENEMY

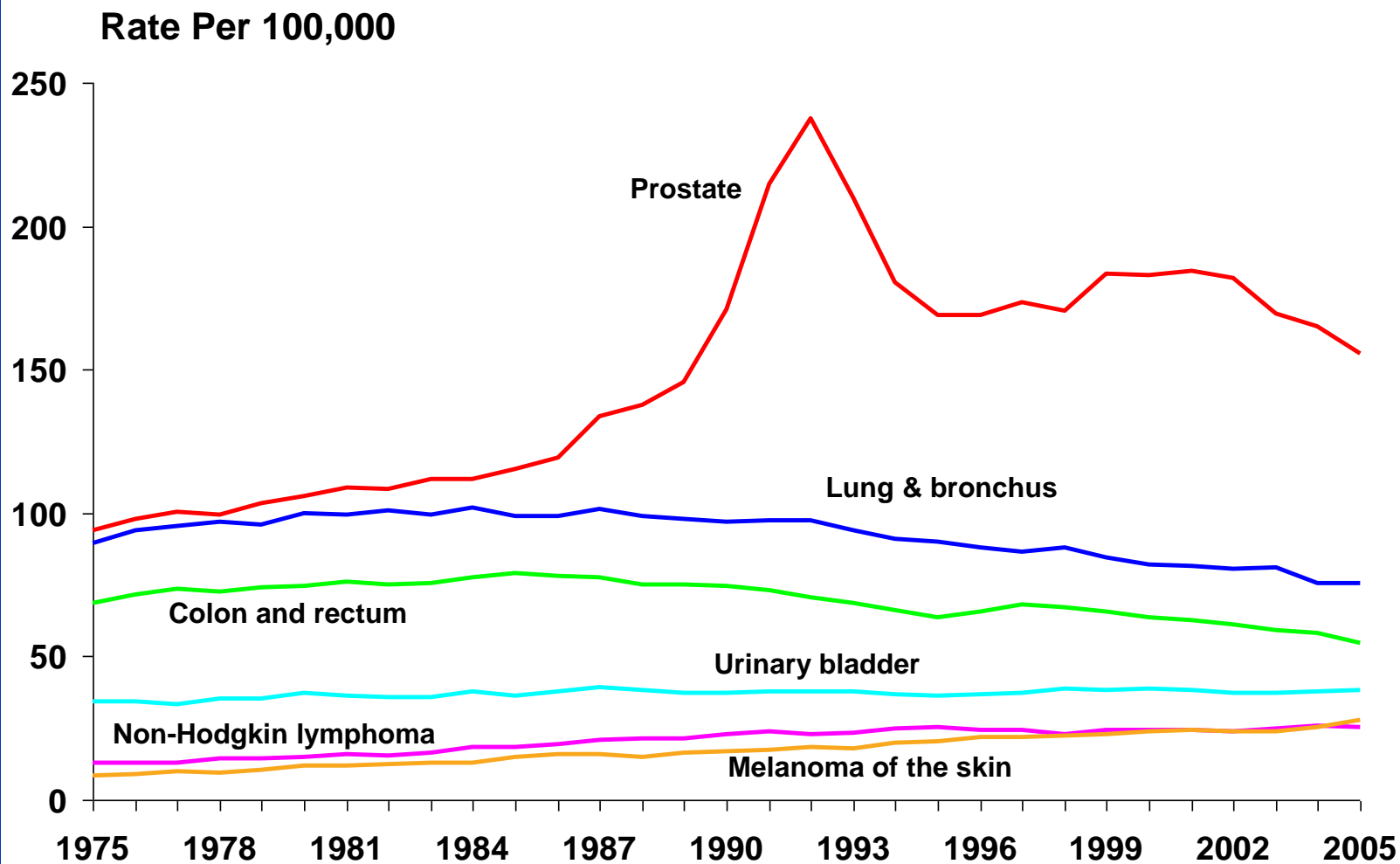
2009 Estimated US Cancer Cases*



*Excludes basal and squamous cell skin cancers and in situ carcinomas except urinary bladder.

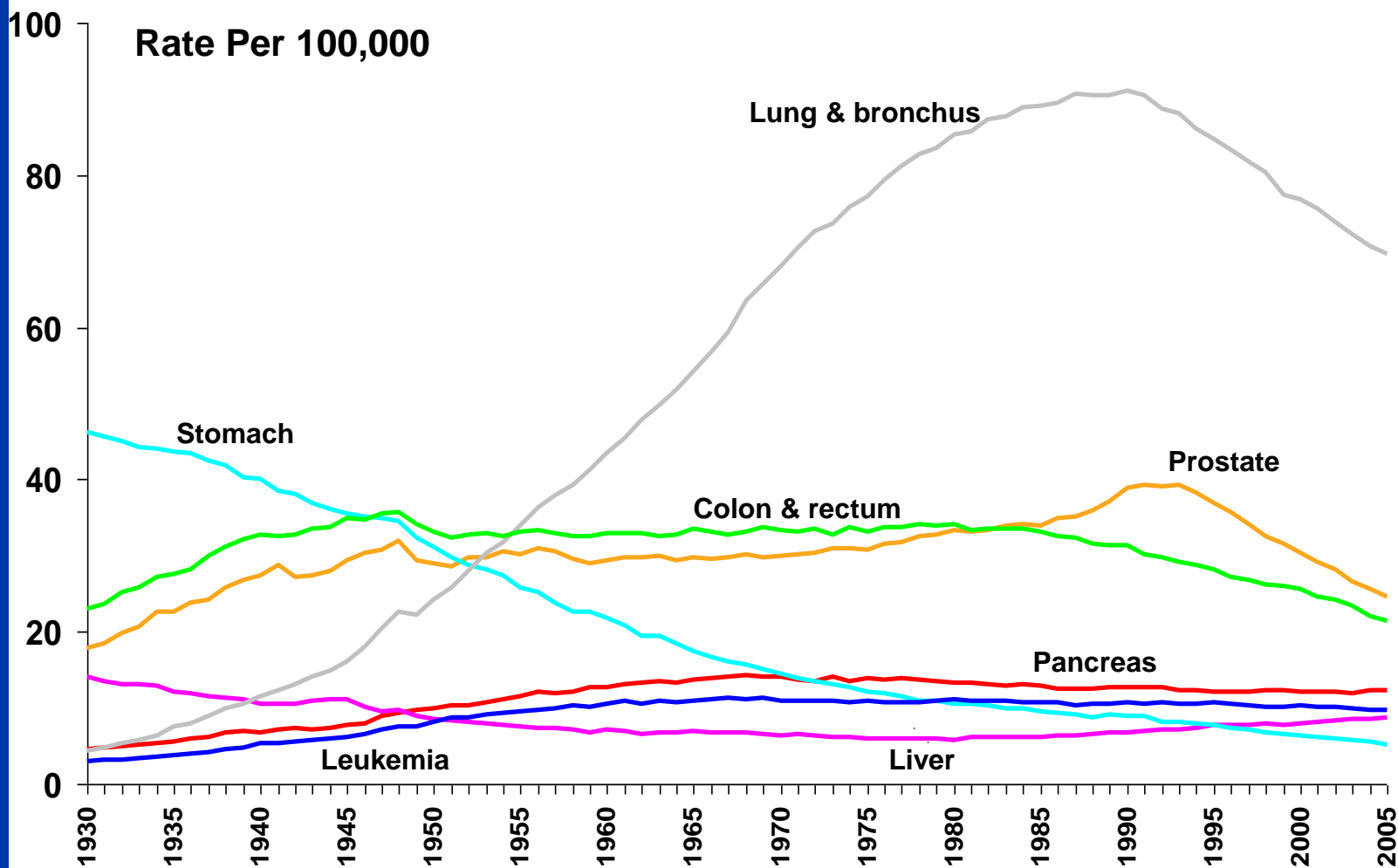
Source American Cancer Society, 2009.

Cancer Incidence Rates* Among Men, US, 1975-2005



*Age-adjusted to the 2000 US standard population and adjusted for delays in reporting.
Source Surveillance, Epidemiology, and End Results Program, Delay-adjusted Incidence database
SEER Incidence Delay-adjusted Rates, 9 Registries, 1975-2005, National Cancer Institute, 2008.

Cancer Death Rates* Among Men, US, 1930-2005

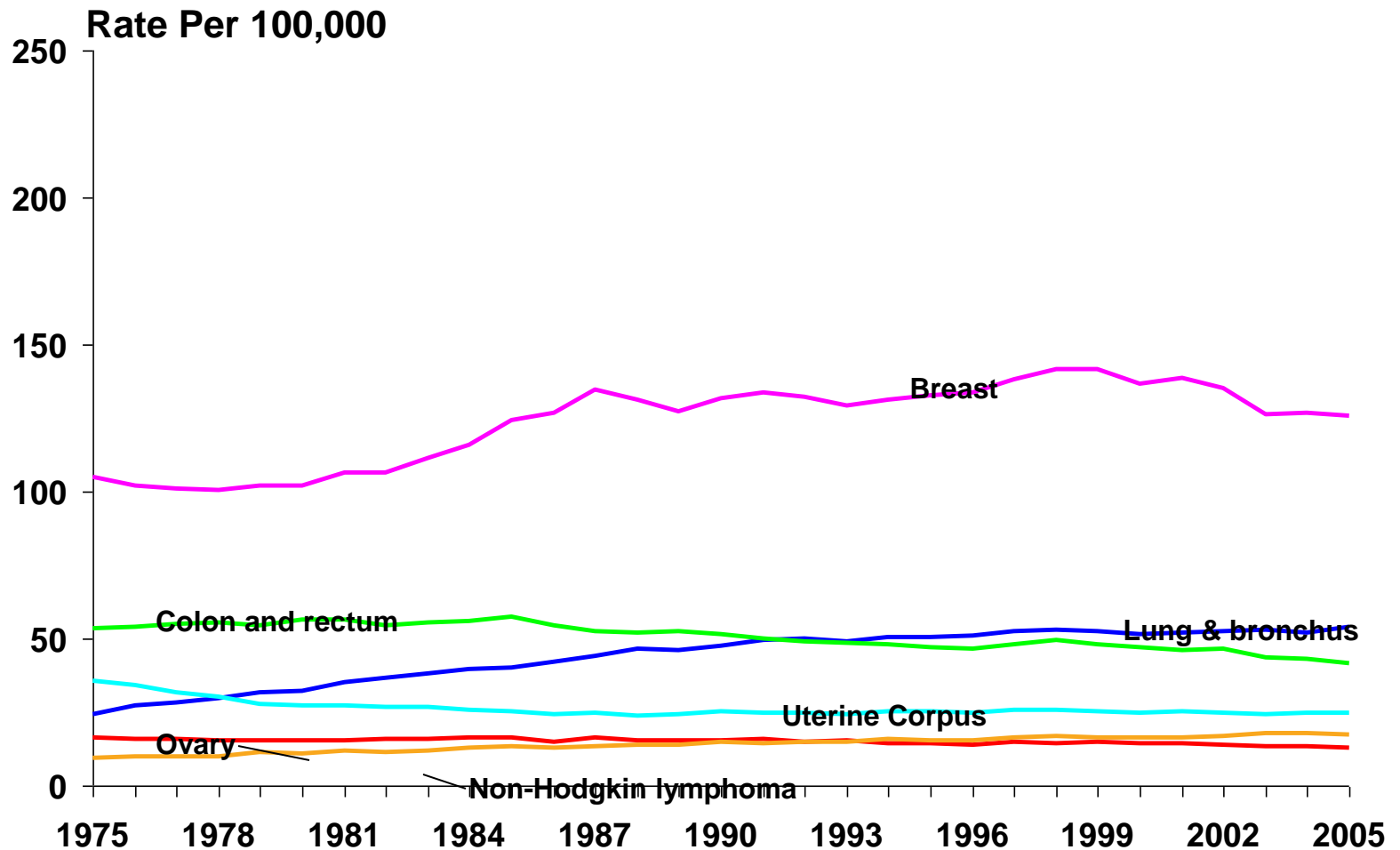


*Age-adjusted to the 2000 US standard population.

Source: US Mortality Data 1960-2005, US Mortality Volumes 1930-1959,

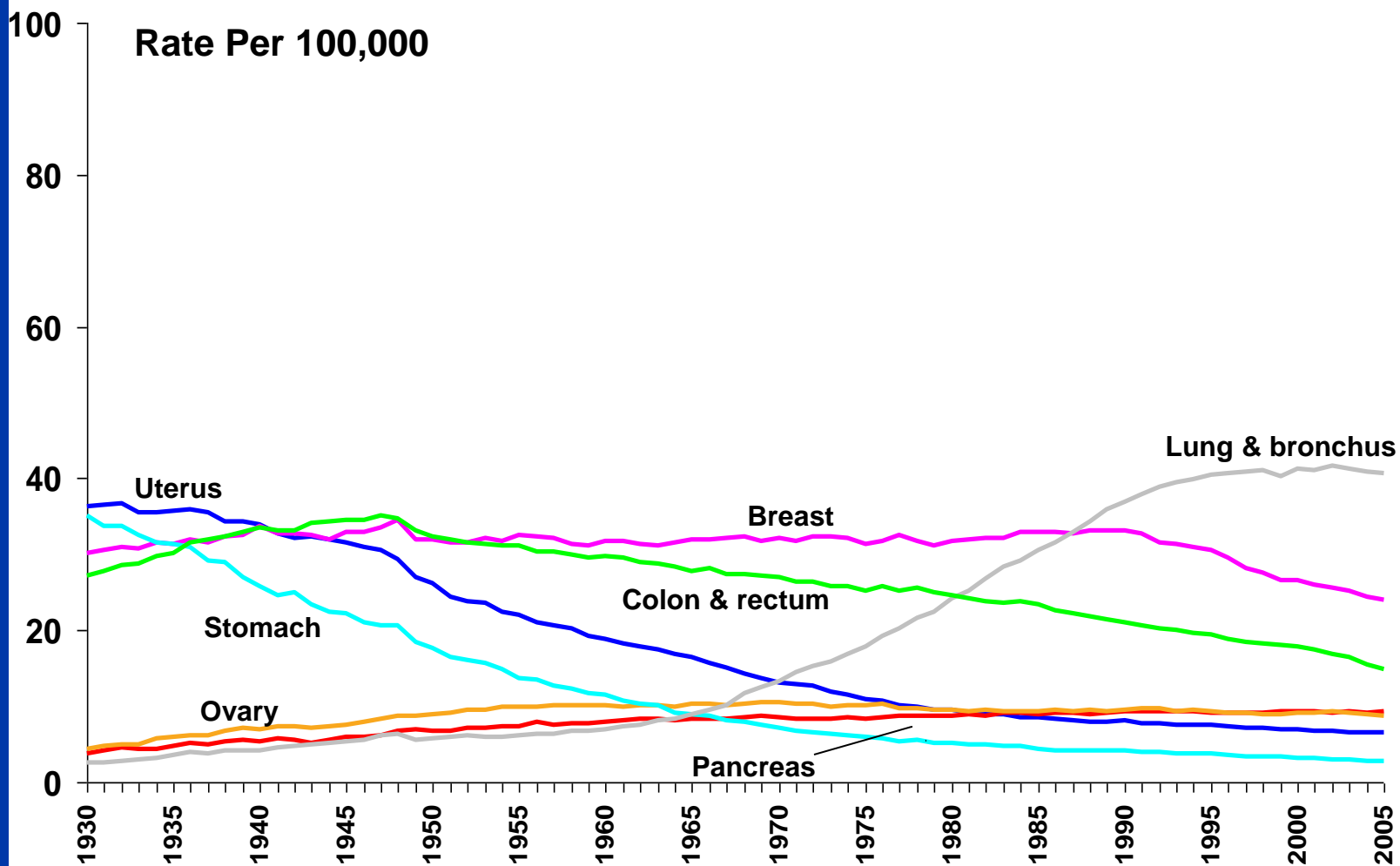
National Center for Health Statistics, Centers for Disease Control and Prevention, 2008.

Cancer Incidence Rates* Among Women, US, 1975-2005



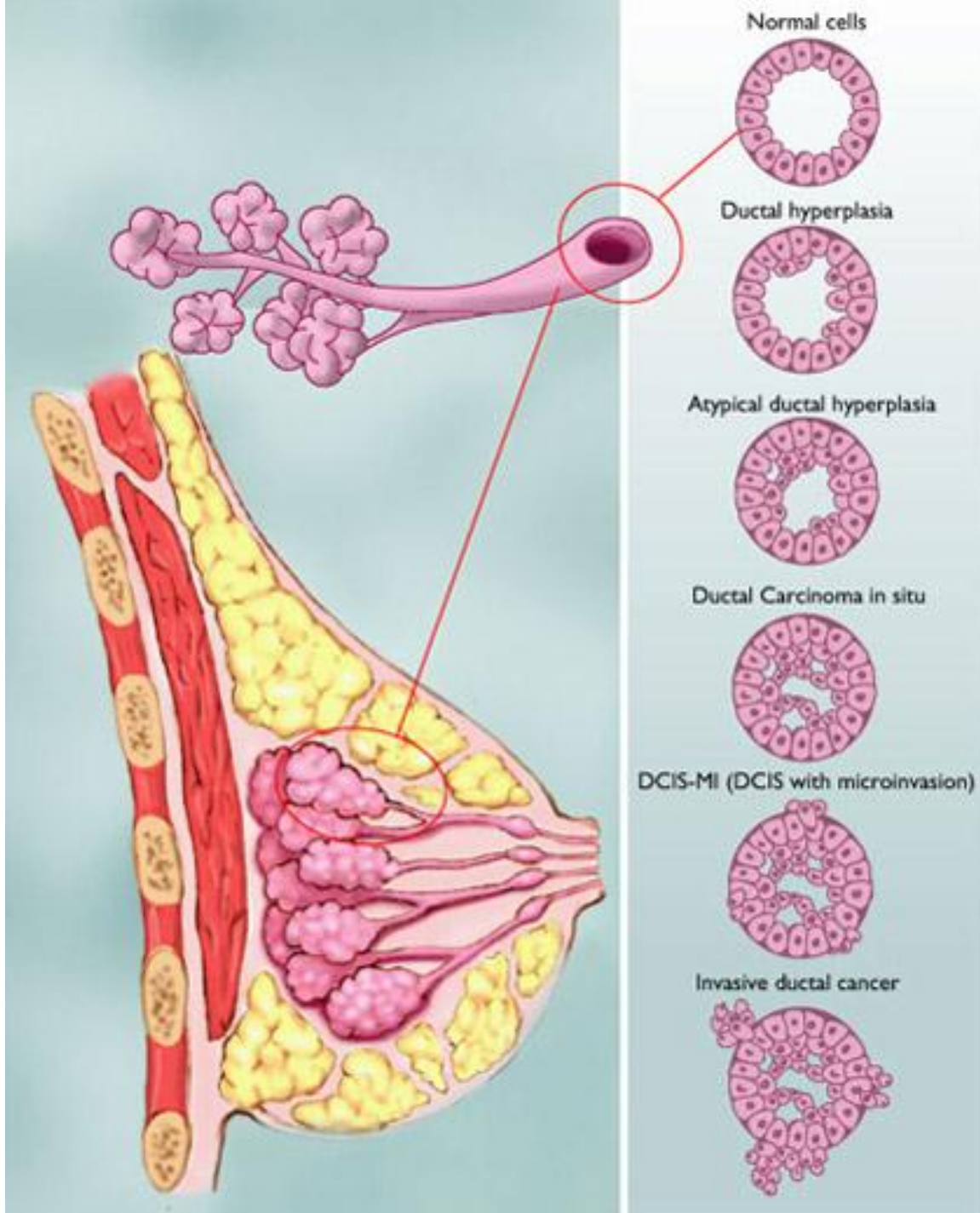
*Age-adjusted to the 2000 US standard population and adjusted for delays in reporting.
Source Surveillance, Epidemiology, and End Results Program, Delay-adjusted Incidence database
SEER Incidence Delay-adjusted Rates, 9 Registries, 1975-2005, National Cancer Institute, 2008.

Cancer Death Rates* Among Women, US, 1930-2005



*Age-adjusted to the 2000 US standard population.

Source: US Mortality Data 1960-2005, US Mortality Volumes 1930-1959, National Center for Health Statistics, Centers for Disease Control and Prevention, 2008.



Normal cells

Normal

Ductal hyperplasia

ADH

Atypical ductal hyperplasia

ADH

Ductal Carcinoma in situ

DCIS

DCIS-MI (DCIS with microinvasion)

DCIS-MI

Invasive ductal cancer

IDC

Risk Factors

1. Age
2. Family History
3. Susceptibility genes (BRCA1 and BRCA2)
4. **Endocrinological Factors**
5. Dietary factors (?)
6. Benign Breast Disease

Endocrinologic Factors

Increased Risk:

Early menarche

Nulliparity

Late first pregnancy

Hormone Replacement
Therapy

Decreased Risk:

late menarche

Early and repeated
pregnancy

Early menopause

Prolonged lactation

Heterogeneous disease

Estrogen Receptor:

1. 60-70% expresses ER
2. ER+ BrCa often responds to anti-estrogen therapy
3. ER- BrCa more aggressive, no hormonal Rx, more common in younger women
4. Tamoxifen treatment decreases risk of new ER+ BrCa but does not affect incidence of ER- BrCa.

HER2/neu:

1. 20-30% have amplification of HER2
2. More aggressive, higher grade
3. Some respond to Rx with trastuzumab/Herceptin

Gene Expression Array Analysis:

Can it be used to impact clinical care?

Define biologically relevant subtypes?

Predict natural history?

Predict response to Rx ?

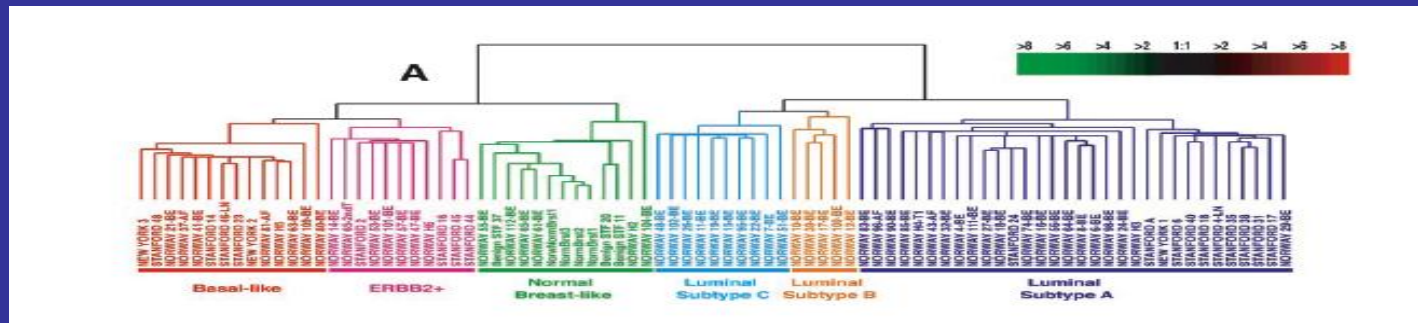
Pros: generates much data on many genes

Cons: generates much data on many genes.

number of variables \gg number of samples

→ OVERFITTING

BrCa subtypes have distinct molecular signatures



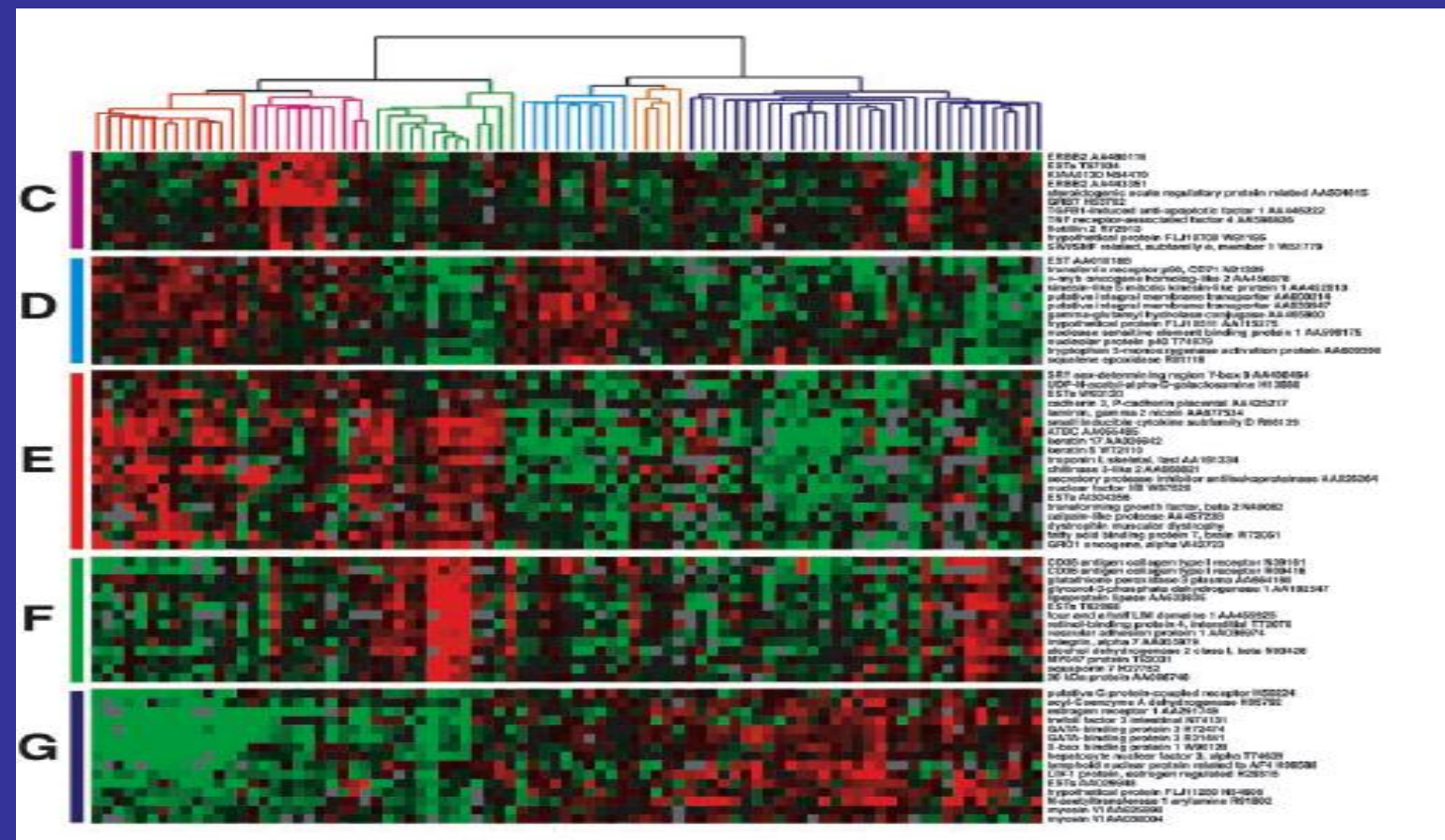
HER2

????

Basal

Normal

Luminal



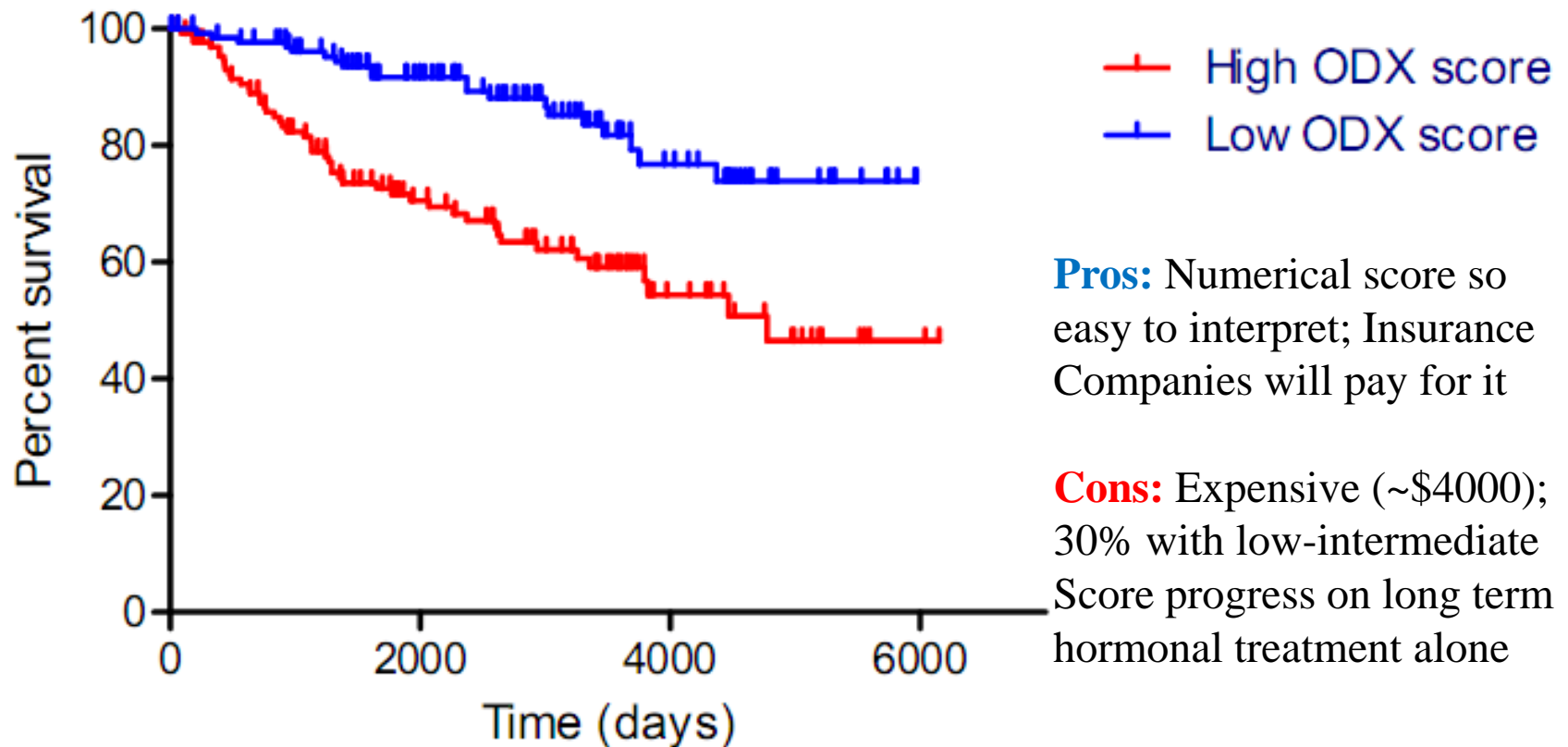
Currently Available Predictive Panels

- Paik et al : *NEJM* 351 (27), 2817 (2004)
21 genes predicting outcome in node negative, ER+ patients
Oncotype DX™ (In clinical use in US)
- van't Veer et al: *Nature medicine* 9 (8), 999 (2003) 70 genes correlated with clinical outcome in large mixed cohort
Mammaprint® (USFDA Approved, Clinical use in Europe – but on decline)
- Wang et al : *Lancet* 365 (9460), 671 (2005)
76 genes correlated with recurrence
Rotterdam signature

Oncotype Dx[®] (Genomic Health)

RT-PCR based assay, measures mRNA levels of 21 genes: HER2, GRB7, GSTM1, CD68, BAG1, invasion markers (MMP11,CTSL2), proliferation markers (Ki67,STK15,Survivin,CCNB1,MYBL2), ER and reference markers.

ODX score is a linear combination of normalized gene expression levels.



Some Simple Bioinformatics Tools to identify patterns in high throughput data:

Clustering

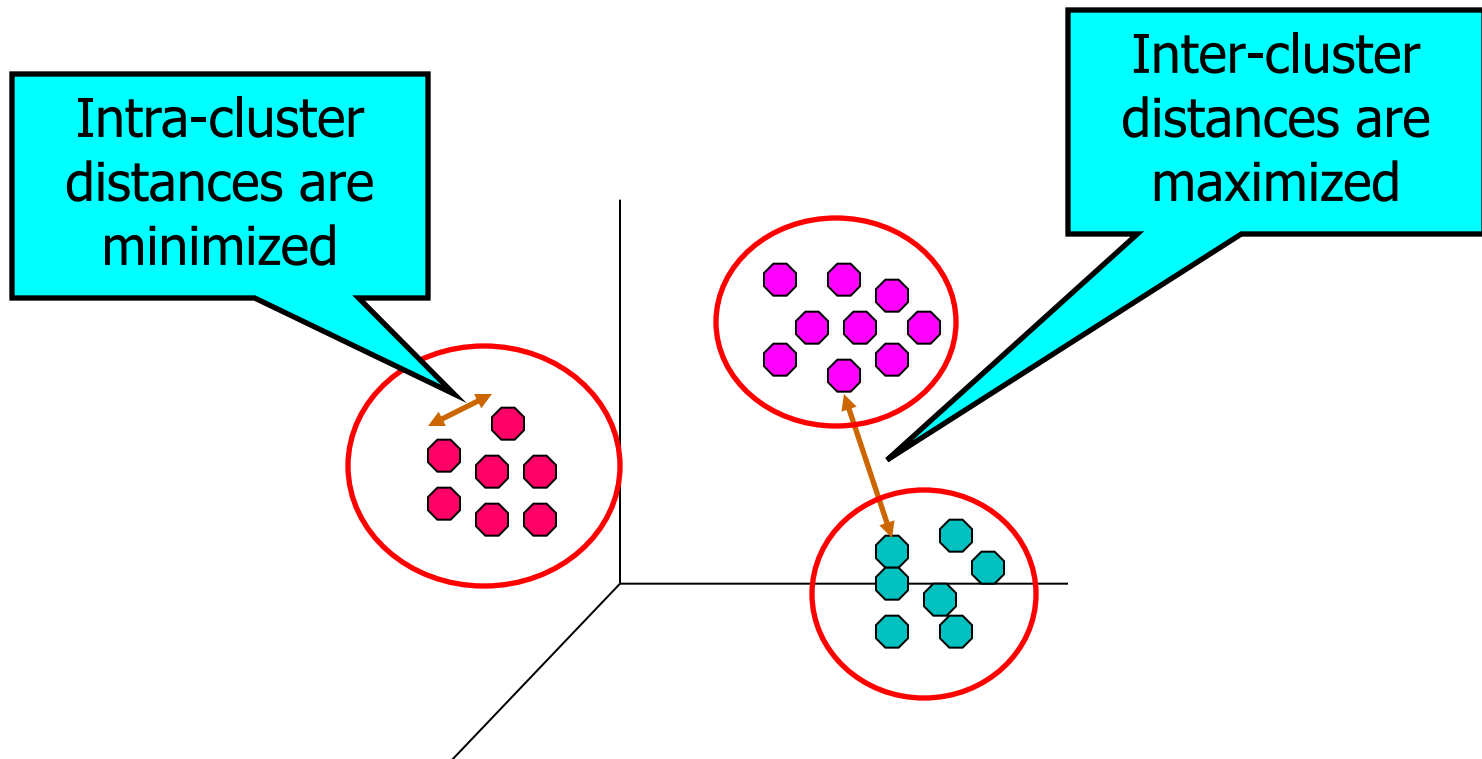
Principal Component Analysis (PCA)

Logical Analysis of Data (LAD)

Network Analysis

What is Cluster Analysis?

- Finding groups of objects such that the objects in a group will be similar (or related) to one another and different from (or unrelated to) the objects in other groups



Unsupervised Consensus Ensemble Clustering

- Unsupervised Clustering
 - Group expression data into clusters
 - Maximize intra-cluster similarity
 - Minimize inter-cluster similarity
- Ensemble:
 - Apply many Clustering Techniques
 - Apply many Data Perturbations (bootstrap)
 - Combine Results into Agreement Matrix

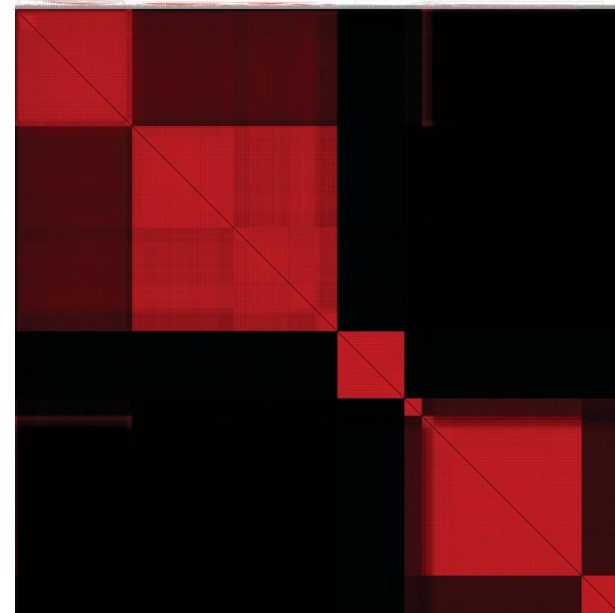
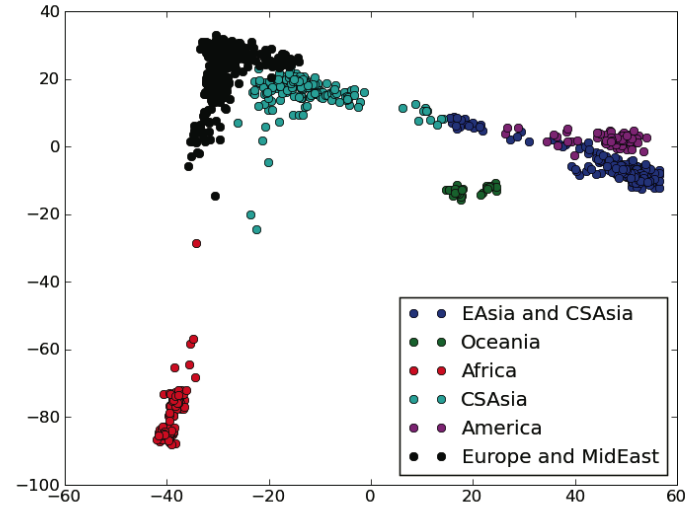
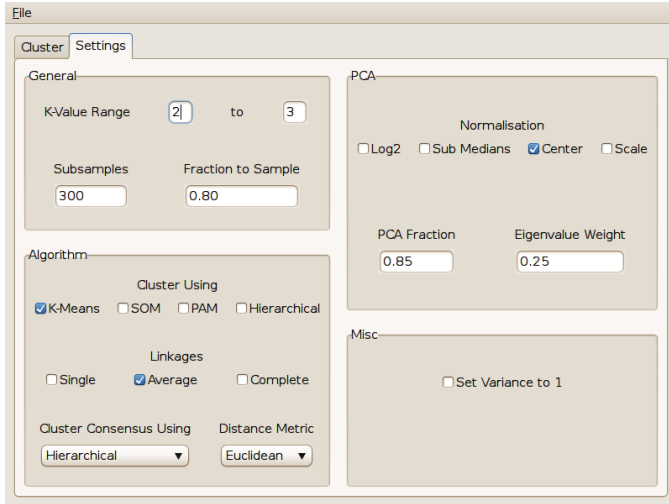
Clustering Methods Used

- Partition Relocation Methods
 - PAM, CLARA, k-Means, Graph-Partitioning
- Agglomerative Methods
 - Average Linkage, Complete Linkage, Mcquitty, Ward, Centroid metrics, bagglo
- Probabilistic Methods
 - Expectation Maximization (EM), Entropy Based Clustering (ENCLUST), SOM,

Agreement Matrix

- Combine bootstrap results per method into $N_{\text{sample}} \times N_{\text{sample}}$ matrix $M(i,j)$.
- $M(i,j)$ = probability that i, j are in same cluster.
- Sort rows to get block diagonal form
- Combine matrices across clustering methods
- Re-sort to get final Agreement Matrix

ConsensusCluster: A tool for unsupervised cluster discovery in numerical data



Seiler M, Huang CC, Szalma S, Bhanot G.
ConsensusCluster: a software tool for unsupervised
cluster discovery in numerical data.
OMICS 2010, 14(1):109.

Study 1: Immune Infiltrate and HER2+ disease: Data from Wang et al, Lancet, 2005

Tumor Bank at the Erasmus Medical Center (Rotterdam, Netherlands)

286 patients (1980-1995)

no systemic therapy

219 patients: conservative surgery

67 patients: mastectomy

248 patients: radiotherapy

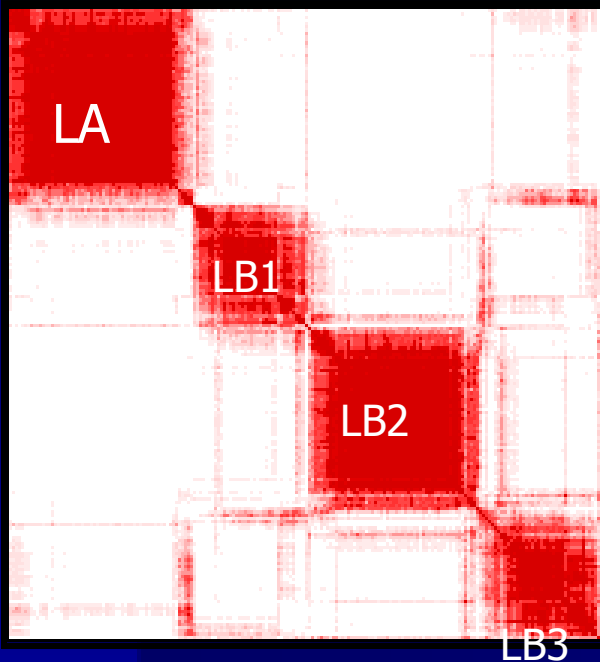
Affymetrix U133a mRNA microarrays

Clinical information

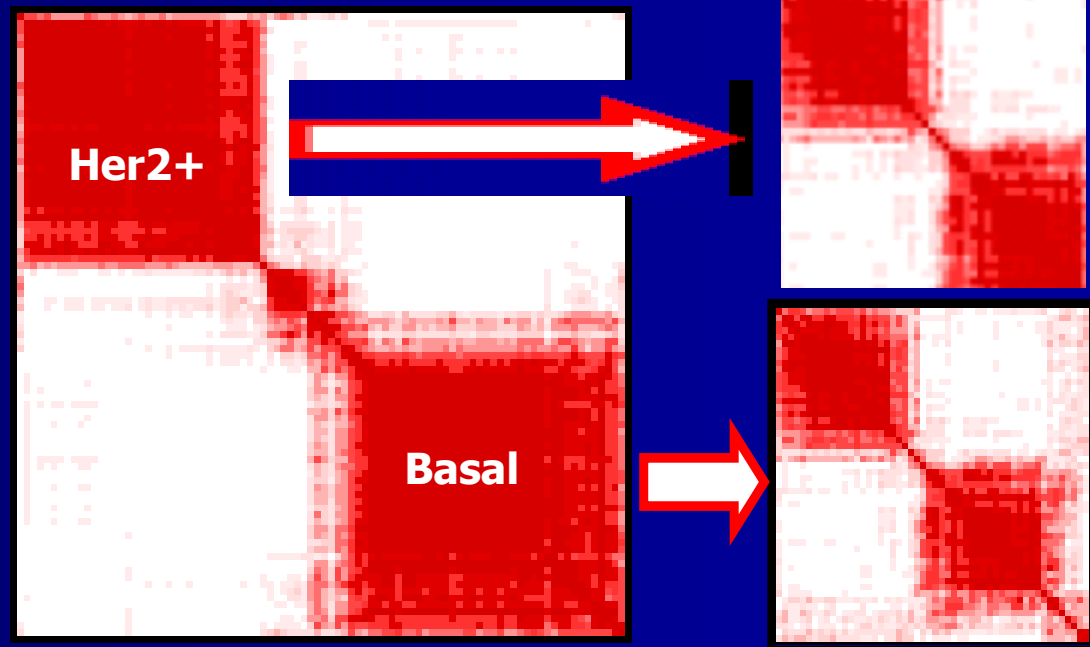
- ER status: 209 ER+, 77 ER-
- time to follow up: median 101 months
- relapse status within 5 years: 93 yes, 183 no
- median age 52 (range 26-83)

Clusters in dataset from Wang et al, 2005, Lancet

Luminal,

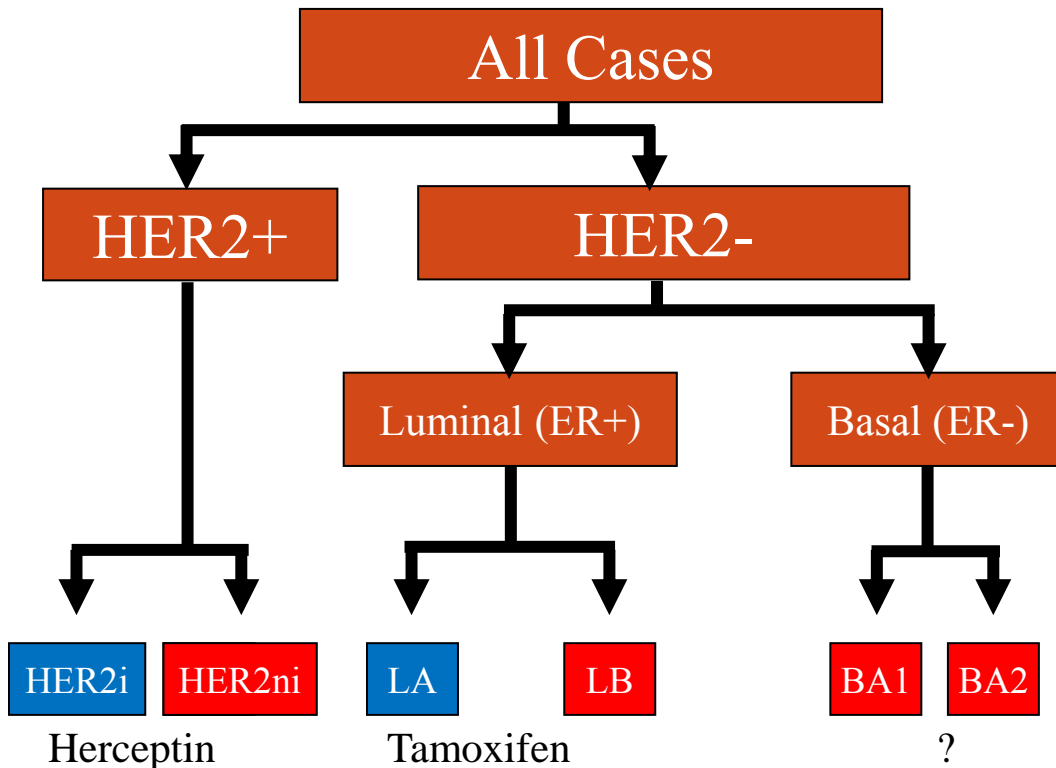


Non – Luminal



G. Alexe, G.S. Dalgin, D. Scanfled, P. Tamayo, J. Mesirov, C. DeLisi, L. Harris, N. Barnard, M. Martel, A.J. Levine, S. Ganesan, G. Bhanot, 'High Expression of Lymphocyte Associated Genes in Node Negative HER2+ Breast Cancer correlates with lower Recurrence rates.' **Cancer Research**, 67, 10669-10676, 2007.

Breast cancer subtypes



Discovery Data Set: (Wang et al. 2005) - 286 early stage (LN) breast cancers treated with surgery + RT. Long term follow-up (7.1 years median)

Validation Data Set: 249 primary invasive breast tumors (Ivshina et al. 2006)- Long term follow-up ~ 9 years.

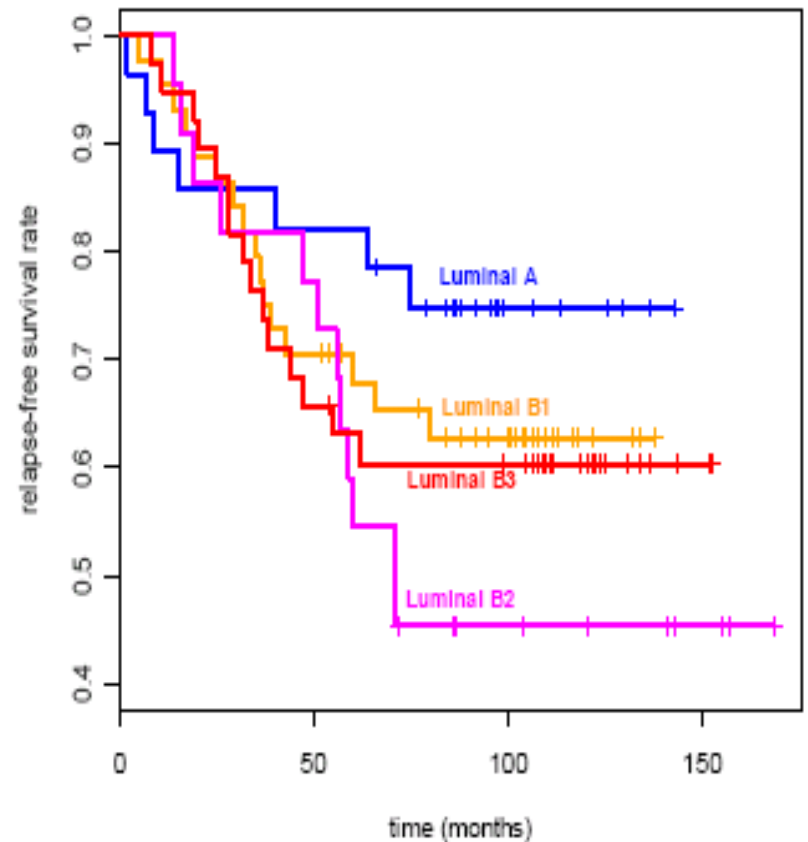
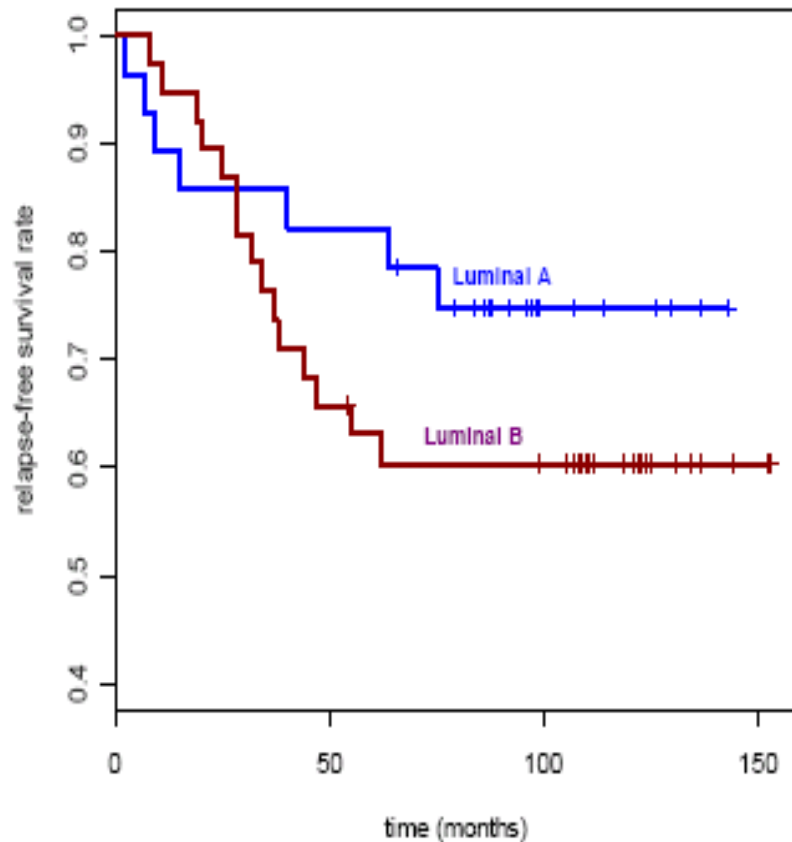
Alexe et al, (2007) Cancer Research, 67, 10669

Dalgin et al, (2007) BMC Bioinformatics 8:291

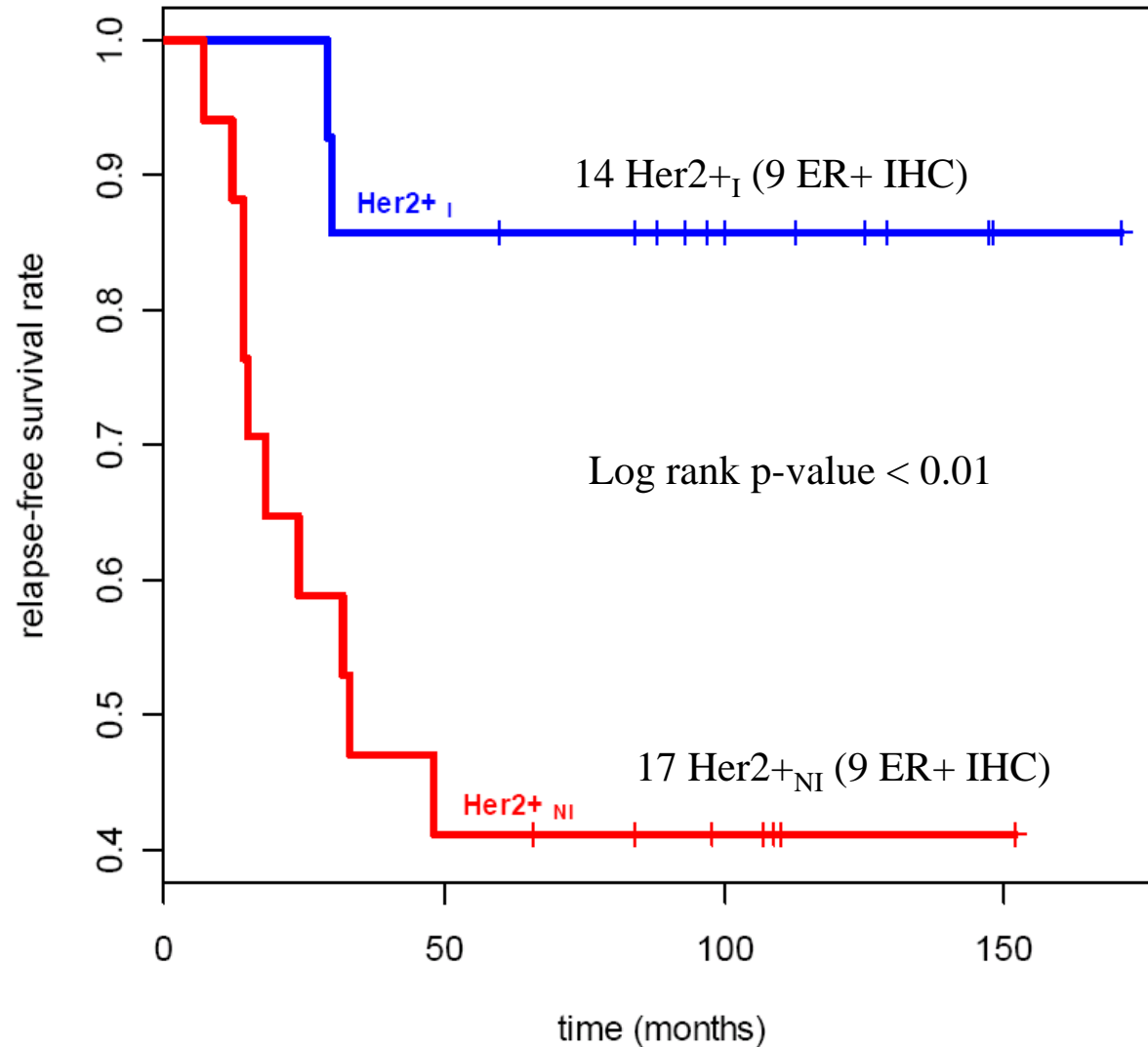
Recurrence free survival in the Luminals

44 core Luminal A, 88 core Luminal B

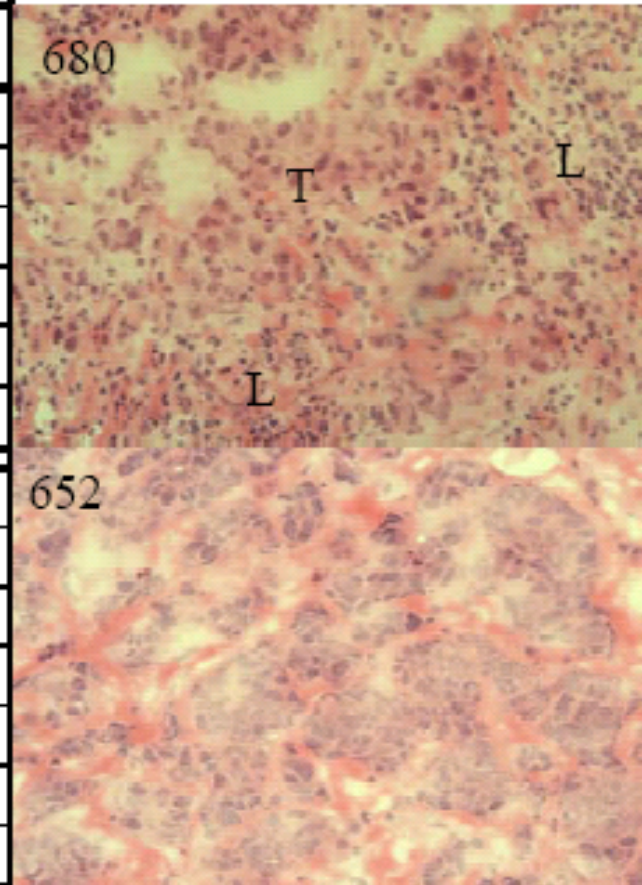
44 LA, 22 LB1, 38 LB2, 28 LB3



Recurrence free survival in HER2+_I and HER2+_{NI}



Tumor	Path CR	Subtype by gene expression	Lymphocytic infiltration	
			Pathologist 1	Pathologist 2
680	No	HER2+ _I	2	2
568	No	HER2+ _I	2	3
334	Yes	HER2+ _I	3	3
438	No	HER2+ _I	NE	NE
422	Yes	HER2+ _I	3	3
611	No	HER2+ _I	NE	NE
<hr/>				
514	No	HER2+ _{NI}	1	1
652	No	HER2+ _{NI}	1	1
405	No	HER2+ _{NI}	1	2
512	No	HER2+ _{NI}	NE	NE
278253	No	HER2+ _{NI}	1	1
641	No	HER2+ _{NI}	1	1
637	No	HER2+ _{NI}	1	1



Lymphocytic Infiltration in HER2+_I subtype from a neoadjuvant phase II trial of trastuzumab and vinorelbine.. H&E sections were independently scored for lymphocytes by two pathologists. NE specimen quality was insufficient for proper evaluation. The score difference had $p < 0.0001$ by the Fisher exact test.

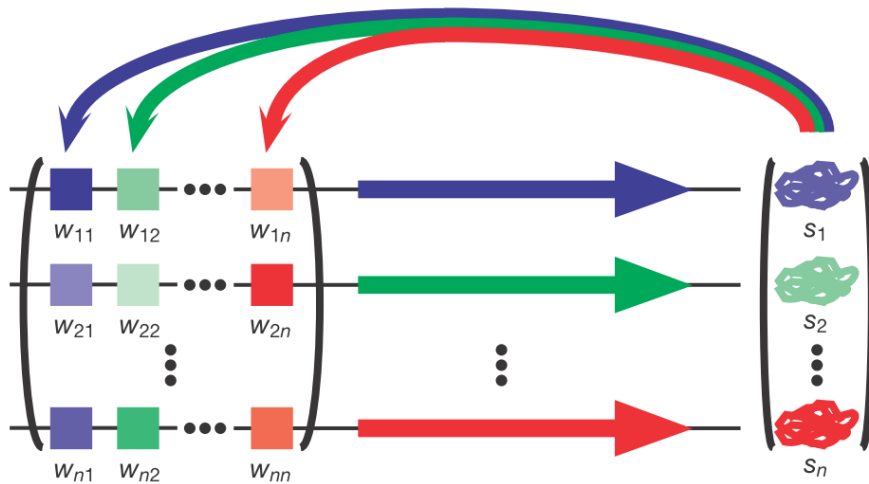
Therapeutic Targets in Triple Negative Breast Cancer



Erhan Bilal (Rutgers/IBM)

Bilal E, et al, Genes and Cancer (2010), 1(10): 1063-73.

Artificial gene networks



$$S(t+1) = f(W * S(t)), f(x) = 1/(1 + e^{-ax})$$

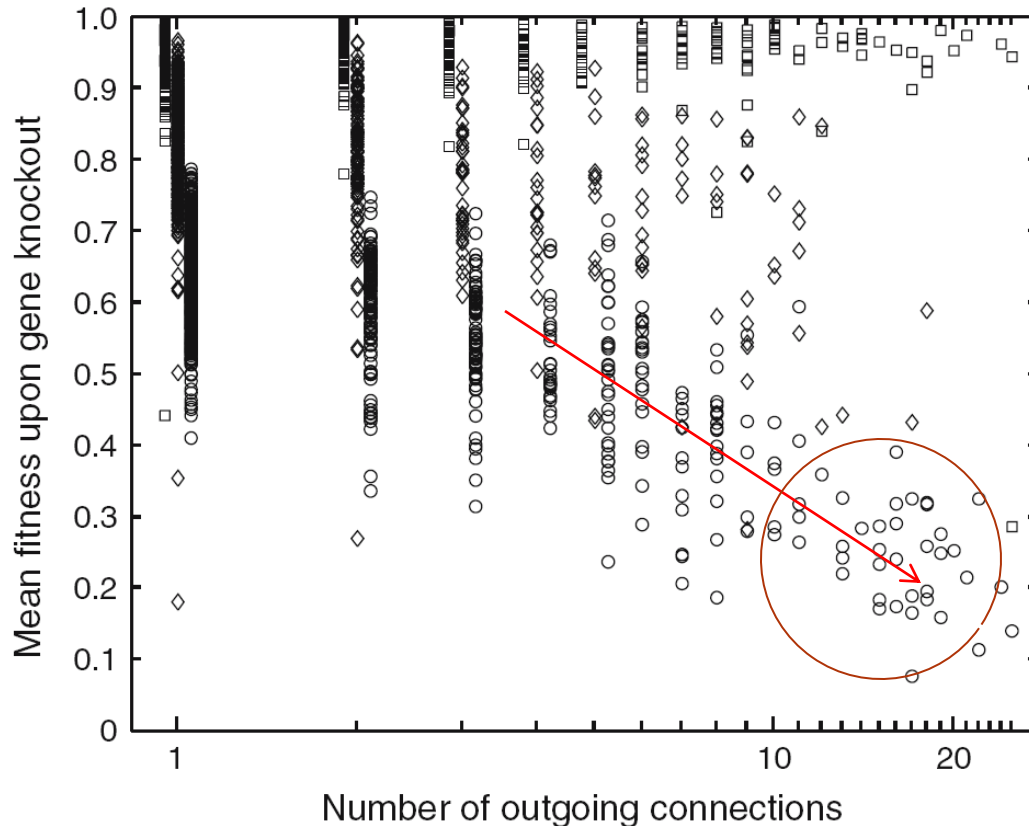
$$F(\hat{S}) = e^{-D/\sigma}$$

$$D = \frac{1}{N} \sum (\hat{S}_k - S_k^{OPT})^2$$

Simulation:

- Evolve 100 networks of 25 genes for 1000 generations under strong selection $\sigma = 0.1$
- Evolution occurs by mutation and recombination of W matrices
- Relevance of a gene is change in fitness on knocking it down.

Node degree and essentiality

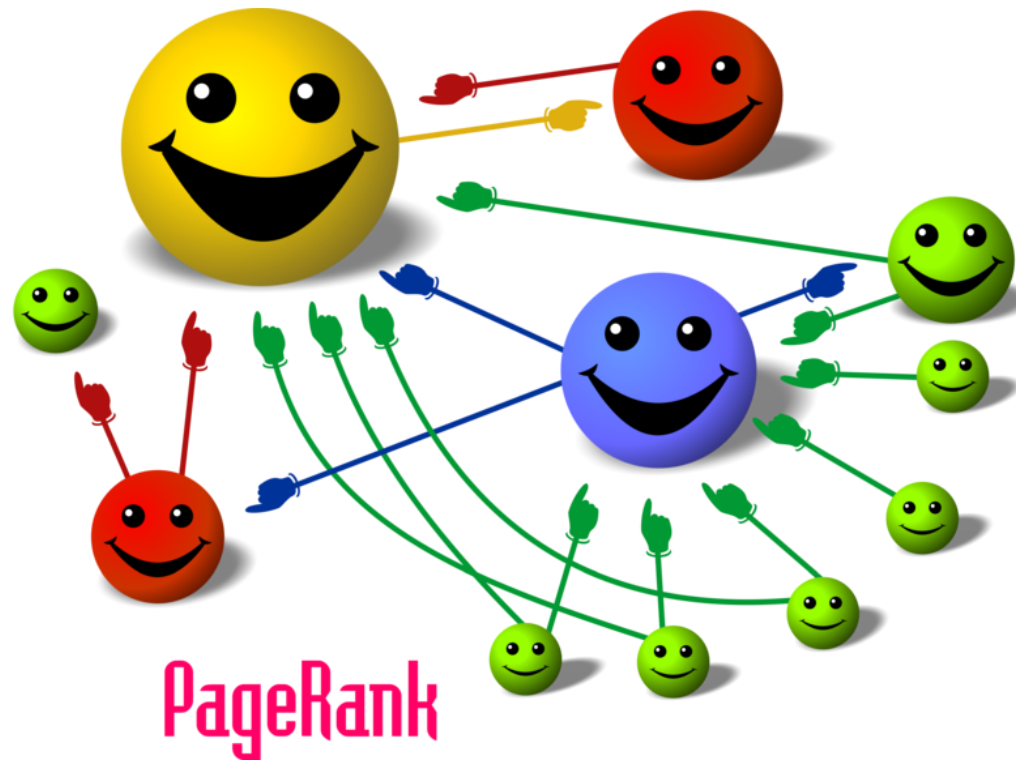


Discovered Hypothesis:

Knocking out genes that are over-expressed and correlated with a large number of other genes should have a big impact on the fitness of the cell

Data are divided into three classes based on the equilibrium expression level of the knocked out gene: $s > 0.75$ (circles), $0.25 < s < 0.75$ (diamonds), and $s < 0.25$ (squares).

Google PageRank Algorithm

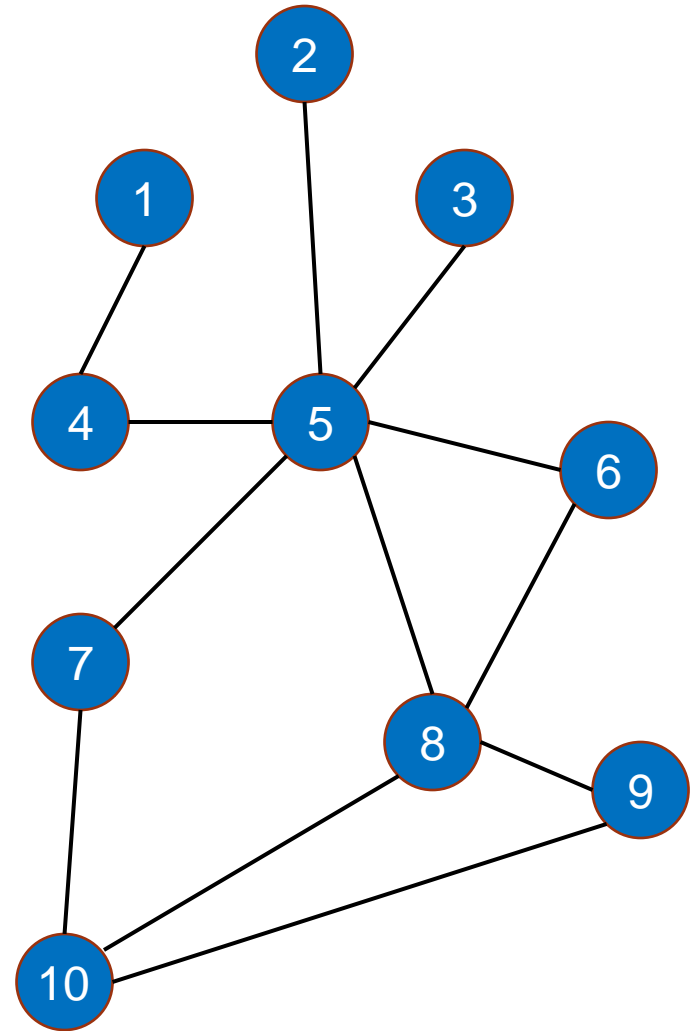


Eigenvector centrality

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

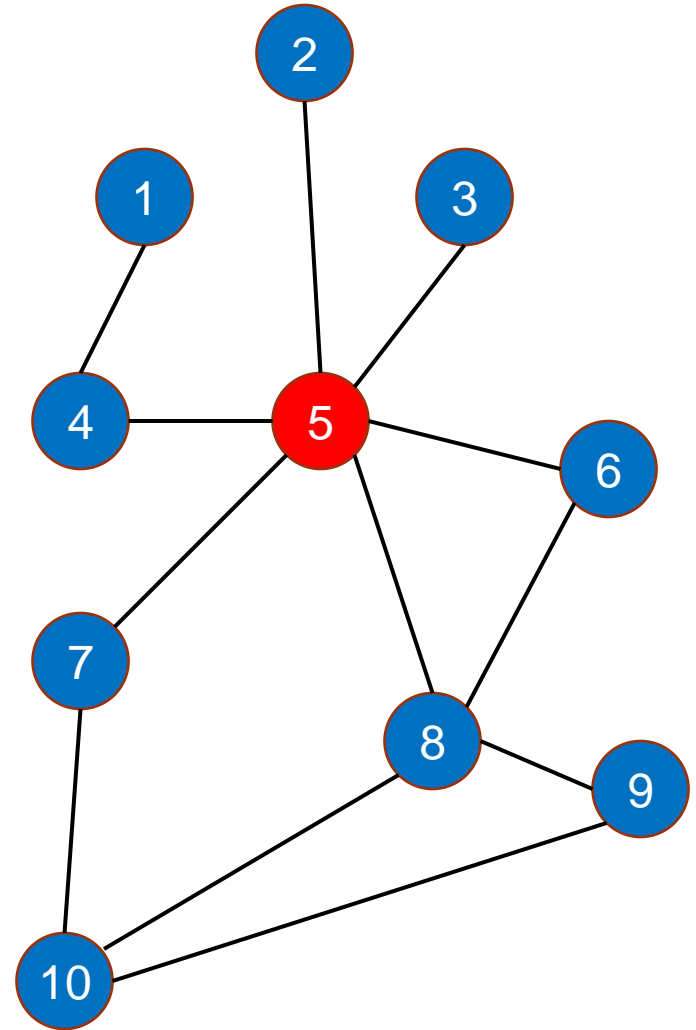
$$s_i = \frac{1}{\lambda} \sum_j a_{ij} s_j, \quad s \text{ centrality score}$$

$$\mathbf{As} = \lambda \mathbf{s}, \quad \text{eigenvector equation}$$



Gene centrality measure

Node	Centrality
1	0.065
2	0.175
3	0.175
4	0.196
5	0.537
6	0.332
7	0.285
8	0.481
9	0.267
10	0.337



Centrality and Outlier scores for top oncogenes in Breast Cancer subtypes:

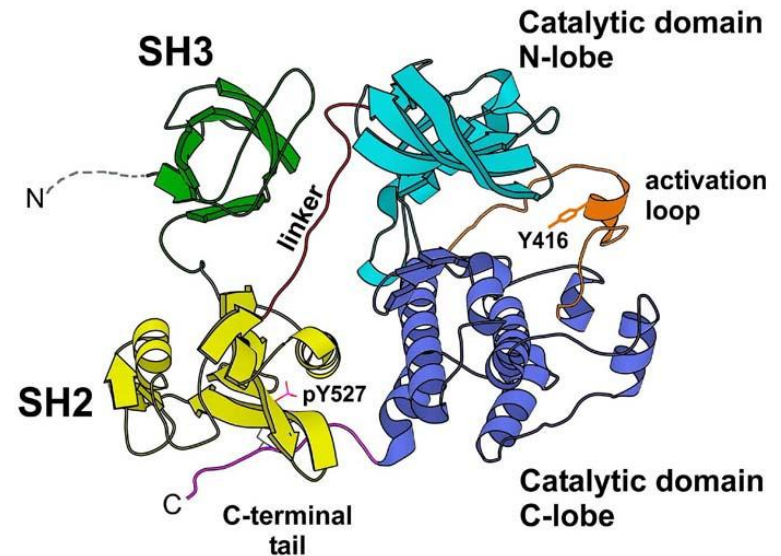
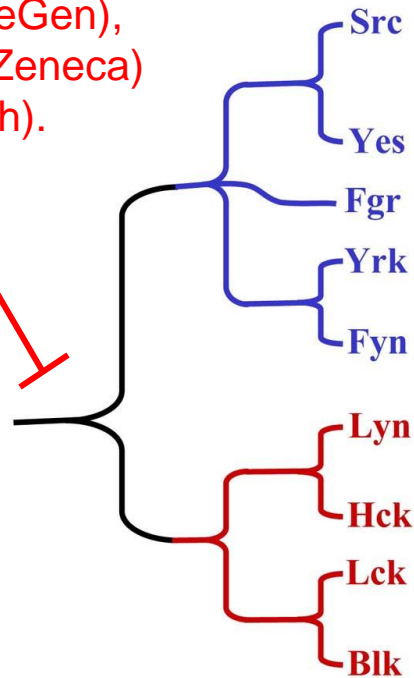
W+I	BA1		BA2		HER2I		HER2NI		LA		LB	
Gene	Centrality	Outlier score θ	Centrality	Outlier score θ	Centrality	Outlier score θ	Centrality	Outlier score θ	Centrality	Outlier score θ	Centrality	Outlier score θ
LYN	4.35	80%	1.9	39%	3.32	29%	0.21	5%	0	0%	0	0%
YES1	3.67	70%	1.82	52%	0	0%	1.24	25%	0	0%	0.24	6%
HCK	3.82	63%	0.38	10%	4.36	47%	0.21	6%	0.59	6%	0.31	8%
FYN	2.43	41%	0.92	31%	7.59	55%	0.38	7%	1.65	13%	0.43	8%
LCK	3.1	52%	0.5	15%	11.9	87%	0	0%	0.91	10%	0.39	7%
PIM2	4.11	65%	0.29	10%	5.87	79%	0	0%	0.6	9%	0.43	13%
HER2	0	0%	0	0%	6.5	100%	4.52	100%	0.01	0%	0.05	2%
TGFBR2	0.04	1%	0.72	9%	3.39	42%	0.83	13%	13.51	66%	0.45	9%
ERG	0	0%	0.73	12%	1.67	21%	2.09	32%	10.5	64%	1.19	26%
FOS	0	0%	0.11	2%	1.5	28%	0.96	21%	5.74	77%	0.76	34%
ETS2	0.46	11%	1.6	32%	2.36	27%	0.71	19%	5.83	34%	0.49	11%
ESR1	0	0%	0	0%	0.82	14%	1.56	26%	7.12	71%	3.48	83%
EGFR	0.75	11%	2.34	38%	1.25	18%	1.32	24%	1.56	19%	5.07	41%

$$g'_y = \frac{g_y - \text{median}(g_i)}{\text{MAD}(g_i)}$$

$$\theta = \frac{1}{N} \sum_j \Delta_j, \text{ where } N = 10, \Delta_j = 1 \text{ if } g'_j > 1, \text{ and } \Delta_j = 0$$

SRC kinase family and its inhibitors

Dasatinib® (Bristol-Myers Squibb),
AP 23846 (Ariad),
TG 100598 (TargeGen),
AZD 0539 (AstraZeneca)
or SKI-606 (Wyeth).



SRC pathway involvement

- Development
- Cell growth
- Immune response
- DNA damage
- ...

Validation: YES1 levels in public BC datasets GSE2034 and GSE4922 (upper)

IF Staining on 13 FFPE samples (lower) shows high (left), medium (middle) and low(right) levels of YES1

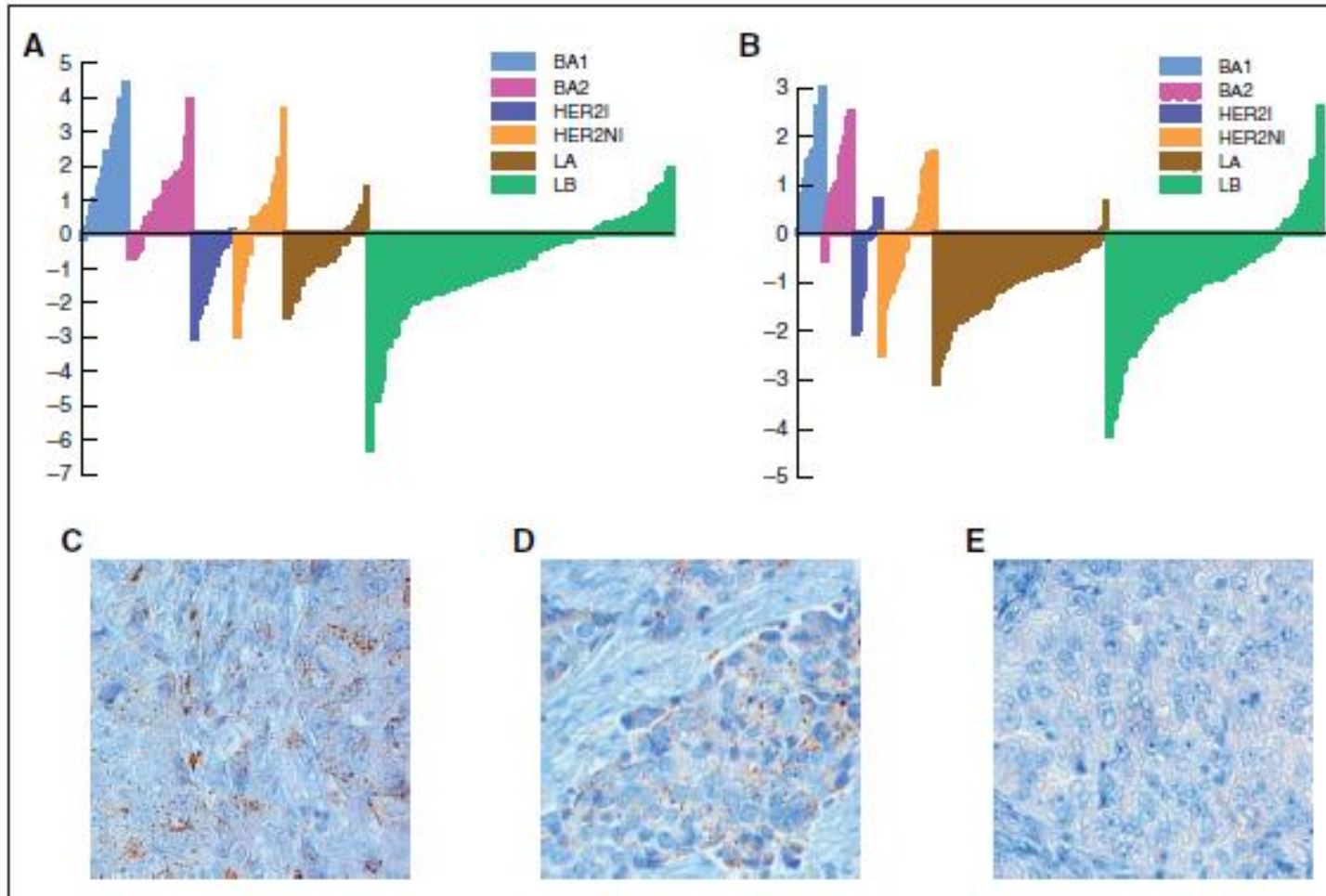
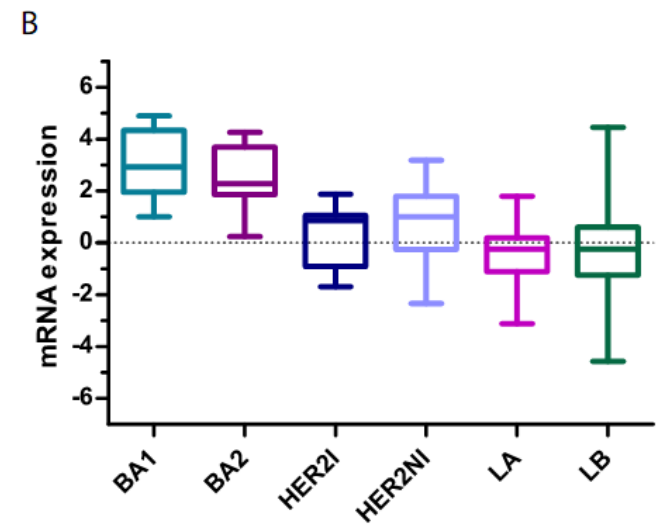
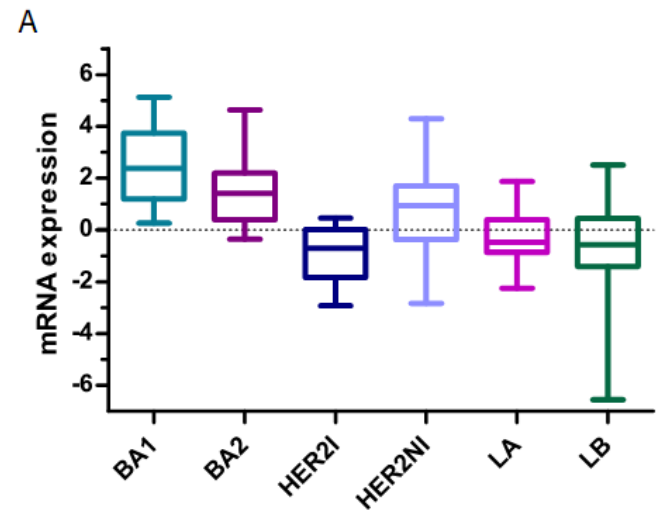
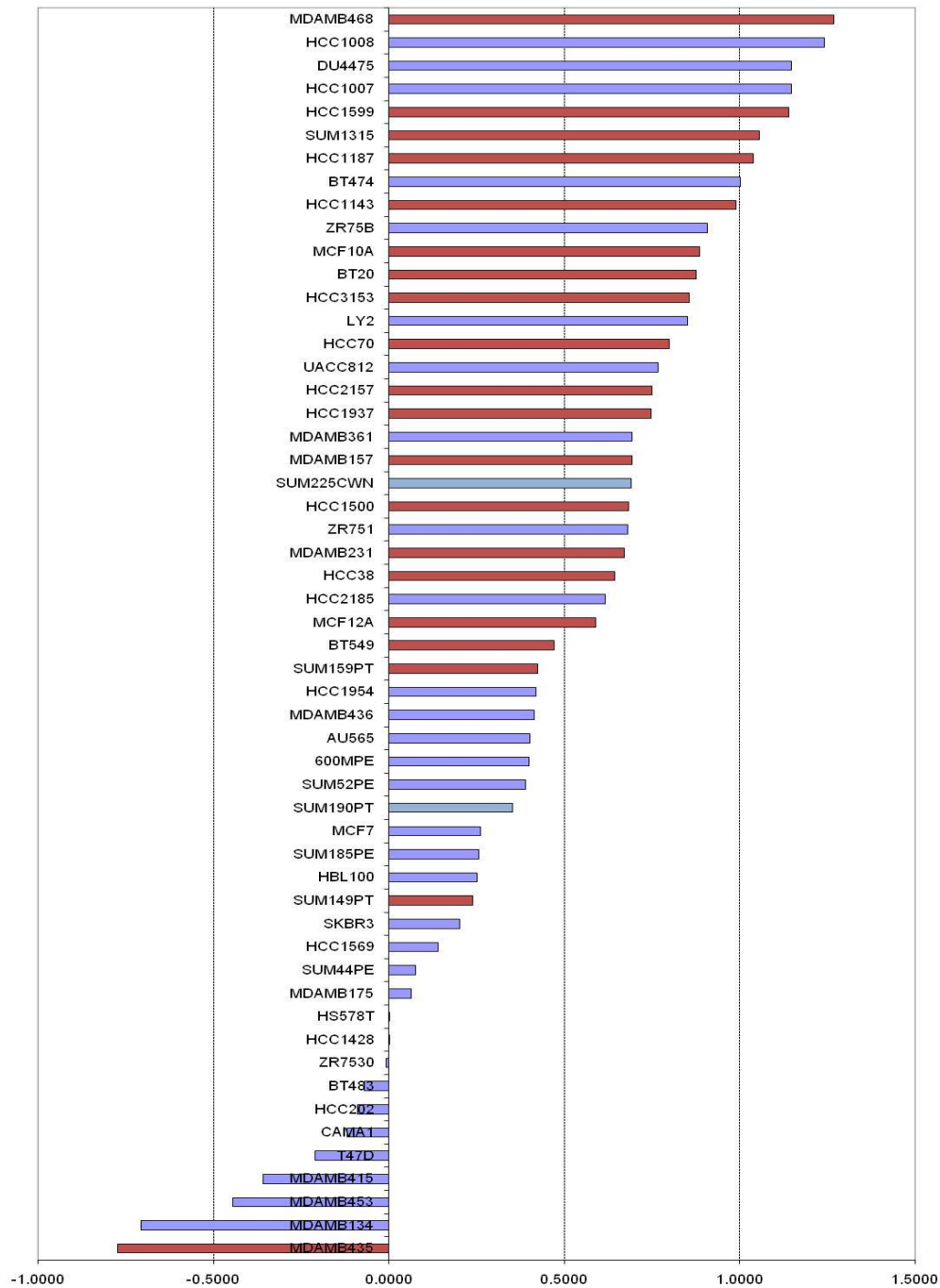
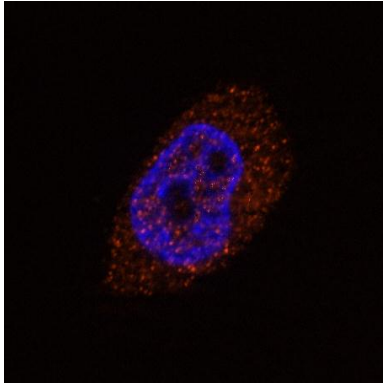


Figure I. YES1 is overexpressed in a subset of basal-like breast tumors. Bar plots showing relative overexpression of YES1 in a subset of basal-like breast tumors in the GSE2034 (A) and GSE4922 (B) gene expression datasets. To confirm this, 13 ER-/PR-/HER2- paraffin-embedded breast cancer tissue slides were probed for expression of YES1 by immunohistochemistry with an appropriate YES1-specific antibody. Of the 13 samples, 2 had high expression levels of YES1, 6 had medium expression, and 5 had low or no expression. Shown are examples of the staining protocol on slides showing high (C), medium (D), and low/zero (E) expression of YES1 in cancer cells on the slides.

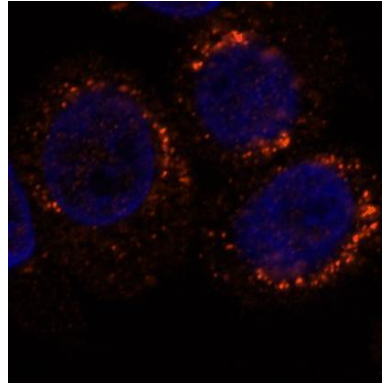


YES1 expression

MDAMB231

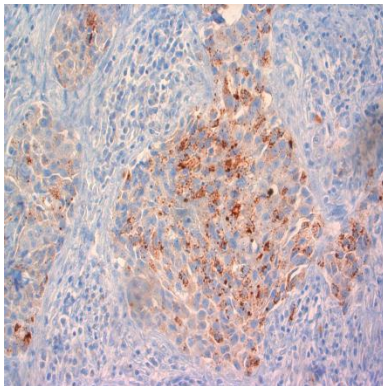


MDAMB468

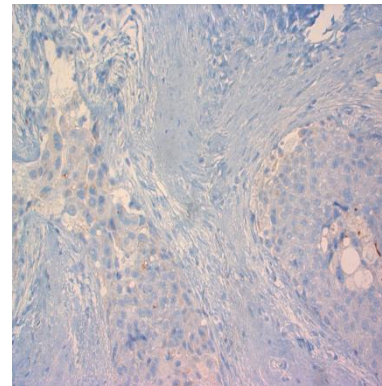


IF staining with anti-Yes1 antibody of two ER-/PR-/HER2- breast cancer cell lines.

ER-/HER2-/YES1+



ER-/HER2-/YES1-

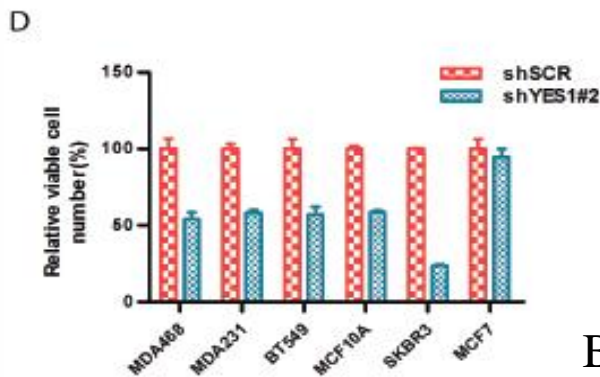
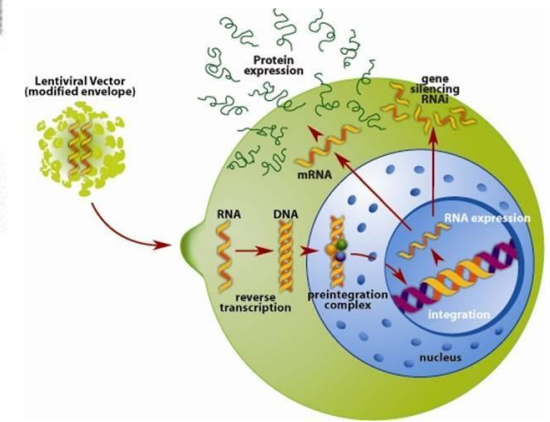
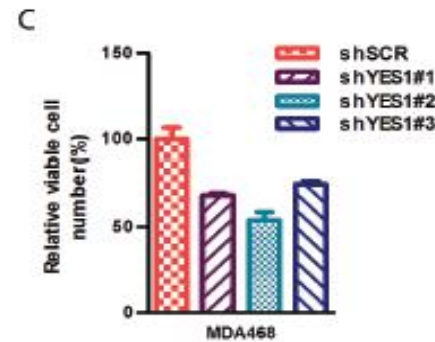
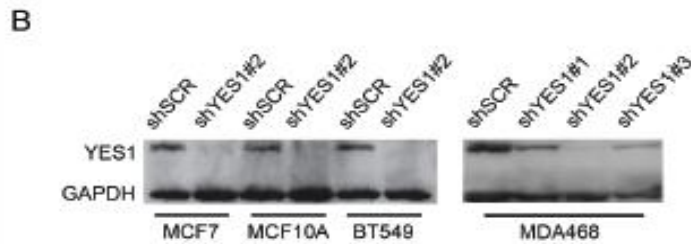
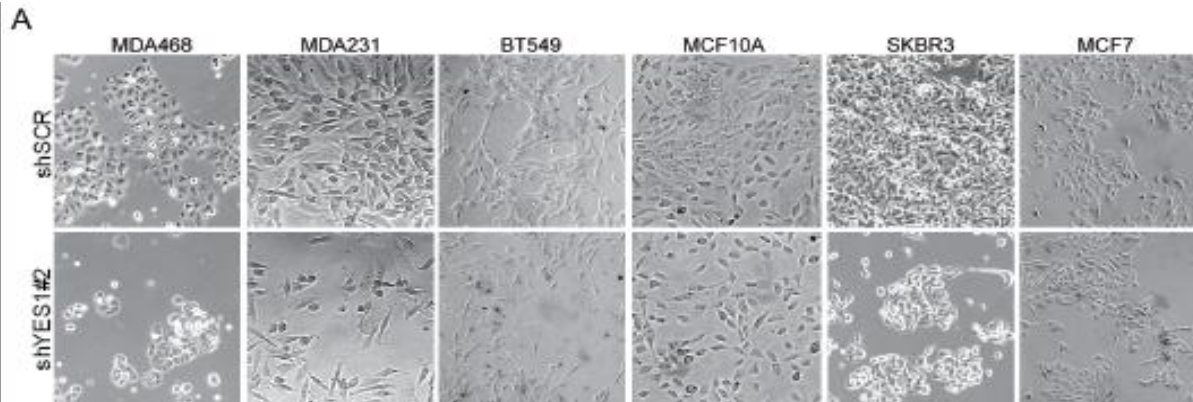


13 ER-/PR-/HER2- breast cancer tissue samples:

- 6 samples with LOW/NO Yes1 expression
- 5 samples with MEDIUM Yes1 expression
- 2 samples with HIGH Yes1 expression

YES1 is a drug target in Triple Negative Breast Cancer

Cell fitness upon silencing YES1 with RNAi using a lentivector



Amplicons and Recurrence in ER+ Breast Cancer

Erhan Bilal



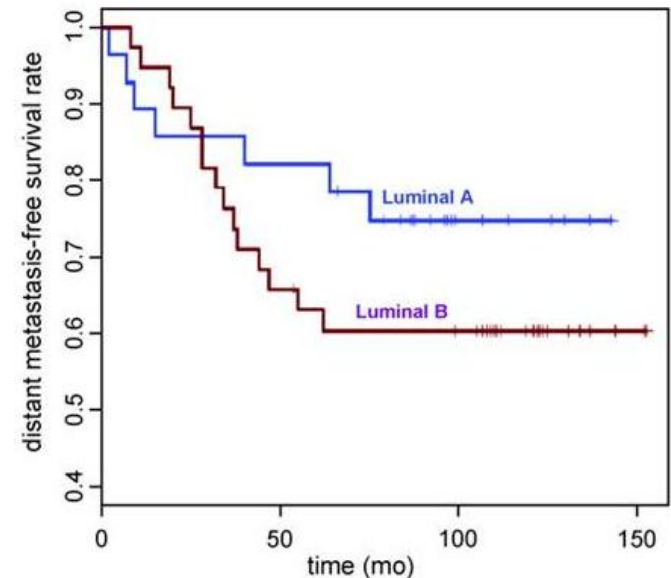
Hege Russnes (OUH)



Vanessa Almendro
(Dana Farber)

Tamoxifen resistance in ER+ breast cancers

- Pathways associated with Tamoxifen resistance in vitro:
 - Estrogen associated transcription factors and activators (E α / β , NF- κ B, NCOA1, NCOA3, PELP1, CBP and p300)
 - Growth factor receptors (ERBB2, EGFR, EGF1R, FGFR)
 - MAPK signaling (MEK, ERK, CDK10)
 - PI3K signaling (AKT, PTEN)
 - SRC interacting proteins (BCAR1, BCAR3)
 - Cyclins, MYC, CDK inhibitors
 - BCL-2 family members (BCL-2, BIK, BAD)
 - Survival signaling (XBP1)
- Signatures predicting response to endocrine therapies:
 - Breast cancer subtypes (LB)
 - Oncotype Dx (21 genes)
 - Genomic grade signature
 - HOXB13/IL17RB expression ratio
 - TuM1 (33 genes), etc.



~ 30% of ER+ patients on Tamoxifen suffer early relapse.

Outlier analysis

$$g'_{ij} = \frac{g_{ij} - \text{median}(g_i)}{\text{MAD}(g_i)}$$

High/low outlier \Rightarrow $> 90\%$ or $< 10\%$ quantile
for each sample.

$\mathbf{B} =$	1	0	0	1	0	0	0	0	0	0
	0	1	0	0	1	0	0	0	0	0
	0	0	0	0	1	0	0	0	0	0
	1	0	0	1	1	0	0	0	0	0
	0	0	1	1	0	1	1	0	0	0
	0	0	0	0	1	1	0	1	0	0
	0	0	0	0	1	0	0	0	0	1
	0	0	0	0	1	1	0	1	1	1
	0	0	0	0	0	0	0	1	0	1
	0	0	0	0	0	0	1	1	1	1

— bad prognosis

— good prognosis

Only genes associated with differential expression were kept
(logrank test, $p < 0.05$)

Three gene expression datasets from Desmedt et al (*Lancet* 2005) GSE 6532

81, 109 and 87 ER+ Affymetrix samples treated with Tamoxifen

Long term follow-up (9 years median)

Let $\mathbf{C}_{1,2}$ be covariance matrices between rows of $\mathbf{B}_{1,2}$

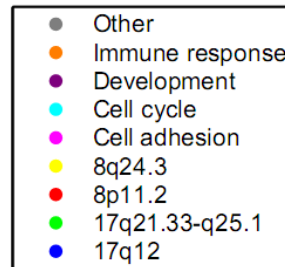
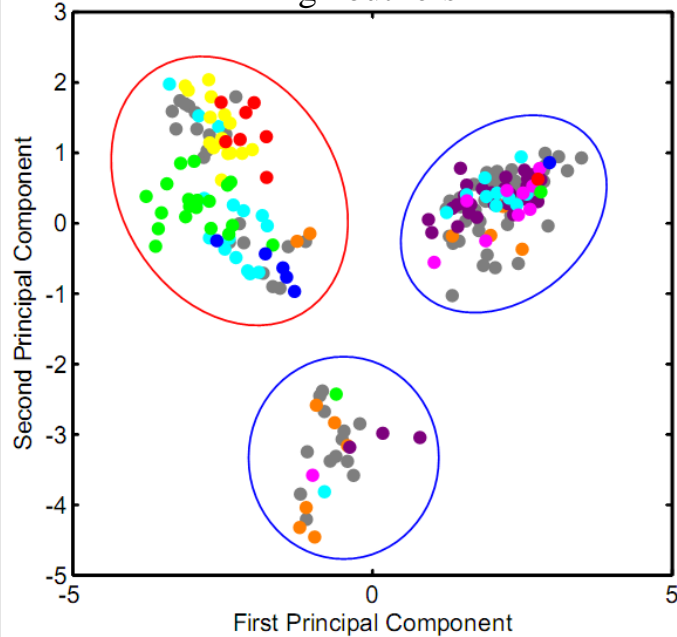
$$\mathbf{R}_{1,2}(i, j) = \mathbf{C}_{1,2}(i, j) / \sqrt{\mathbf{C}_{1,2}(i, i) \mathbf{C}_{1,2}(j, j)}$$

Clusters of tightly correlated genes identified by iteratively removing row/column i if

$$\sum_j \Delta_{ij} \leq 1, \text{ where } \Delta_{ij} = 1 \text{ if } \mathbf{R}_{1,2}(i, j) > 0.5 \text{ and } \Delta_{ij} = 0 \text{ otherwise}$$

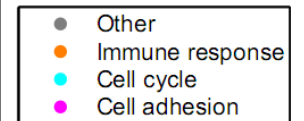
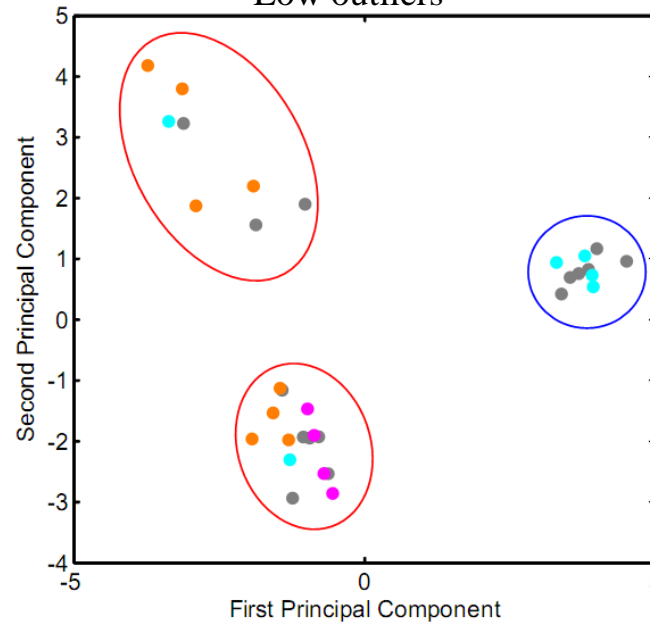
Gene patterns associated with Tamoxifen resistance in dataset GSE6532

High outliers

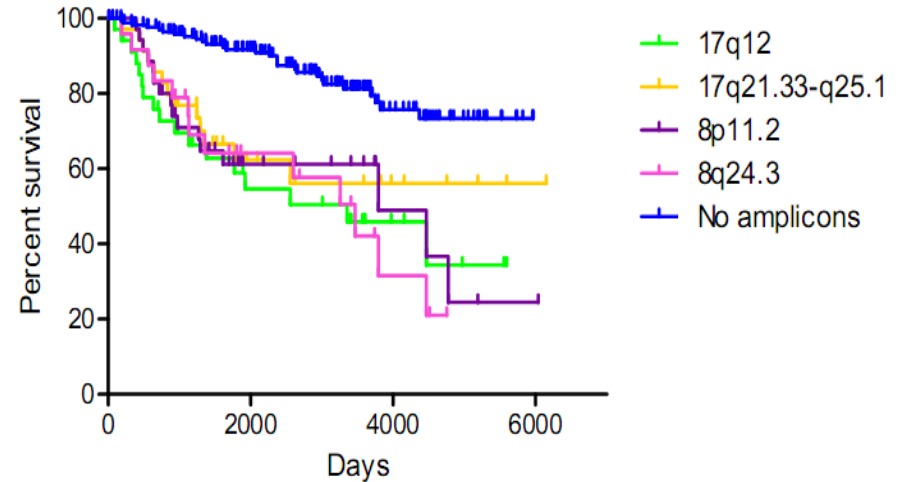
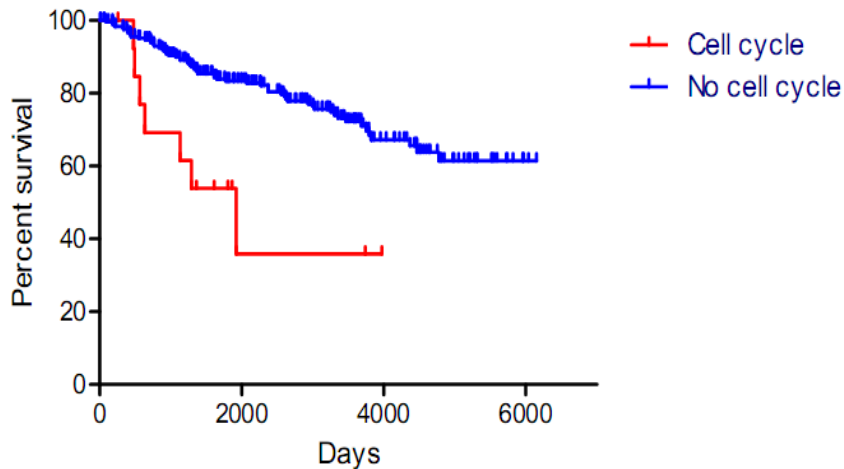


	Over-expression	p-values	Under-expression	p-values
Good Prognosis with Tamoxifen treatment	Immune response	1.61E-05	Cell cycle	1.10E-03
	Development	7.56E-08		
	Cell adhesion	1.68E-04		
Poor outcome with Tamoxifen treatment	Cell cycle	9.12E-07	Immune response	1.36E-05
	17q21.33-q25.1	3.87E-05	Cell adhesion	2.01E-08
	17q12	1.39E-08		
	8p11.2	1.11E-16		
	8q24.3	2.22E-16		

Low outliers

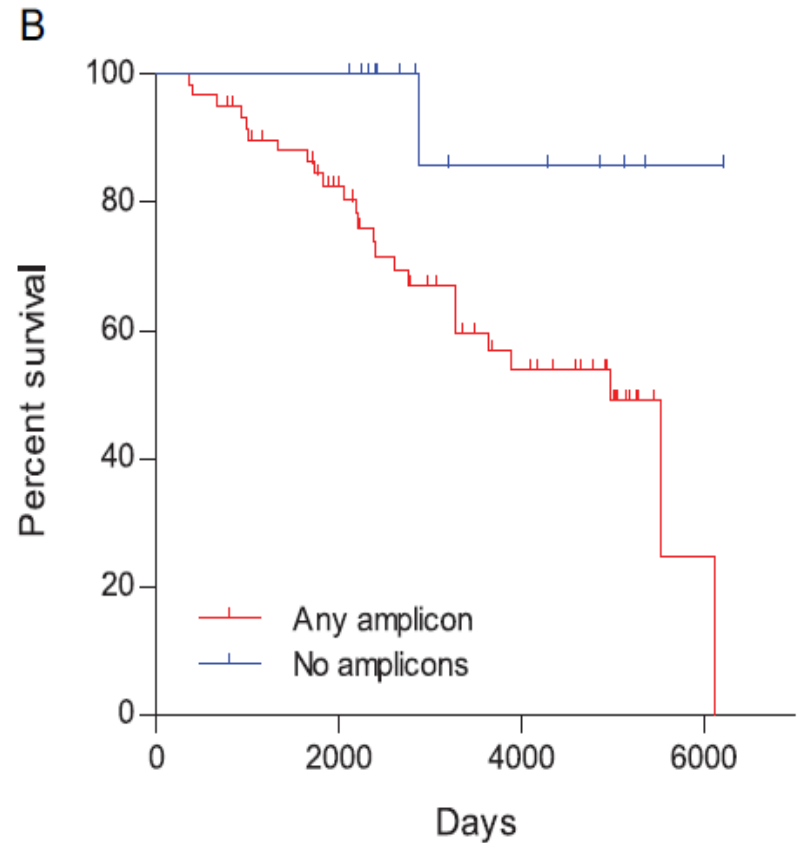
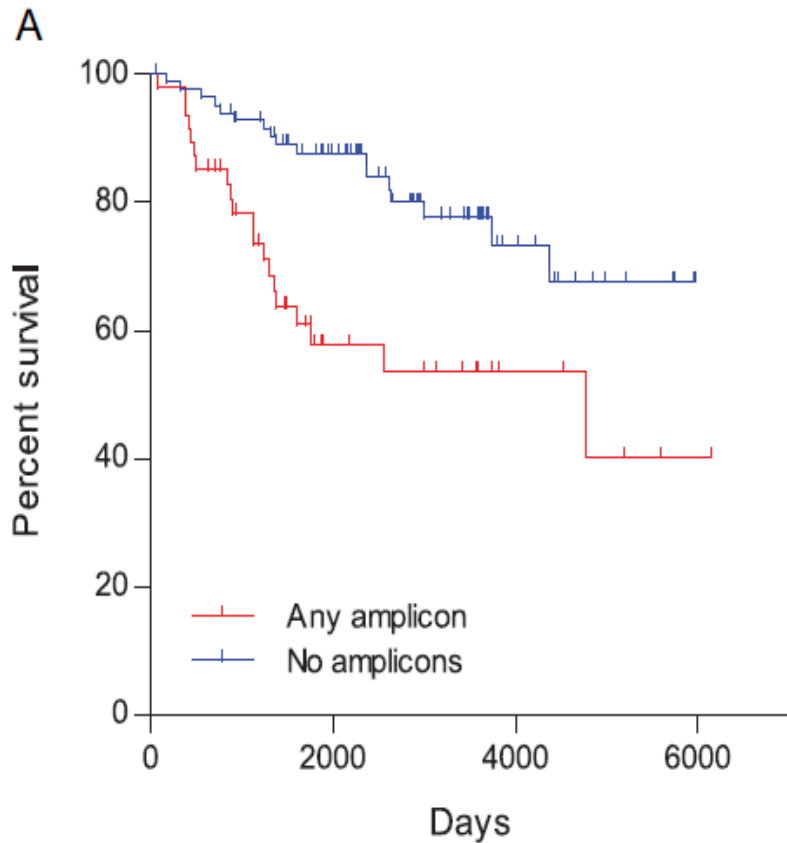


Survival Differences between classes

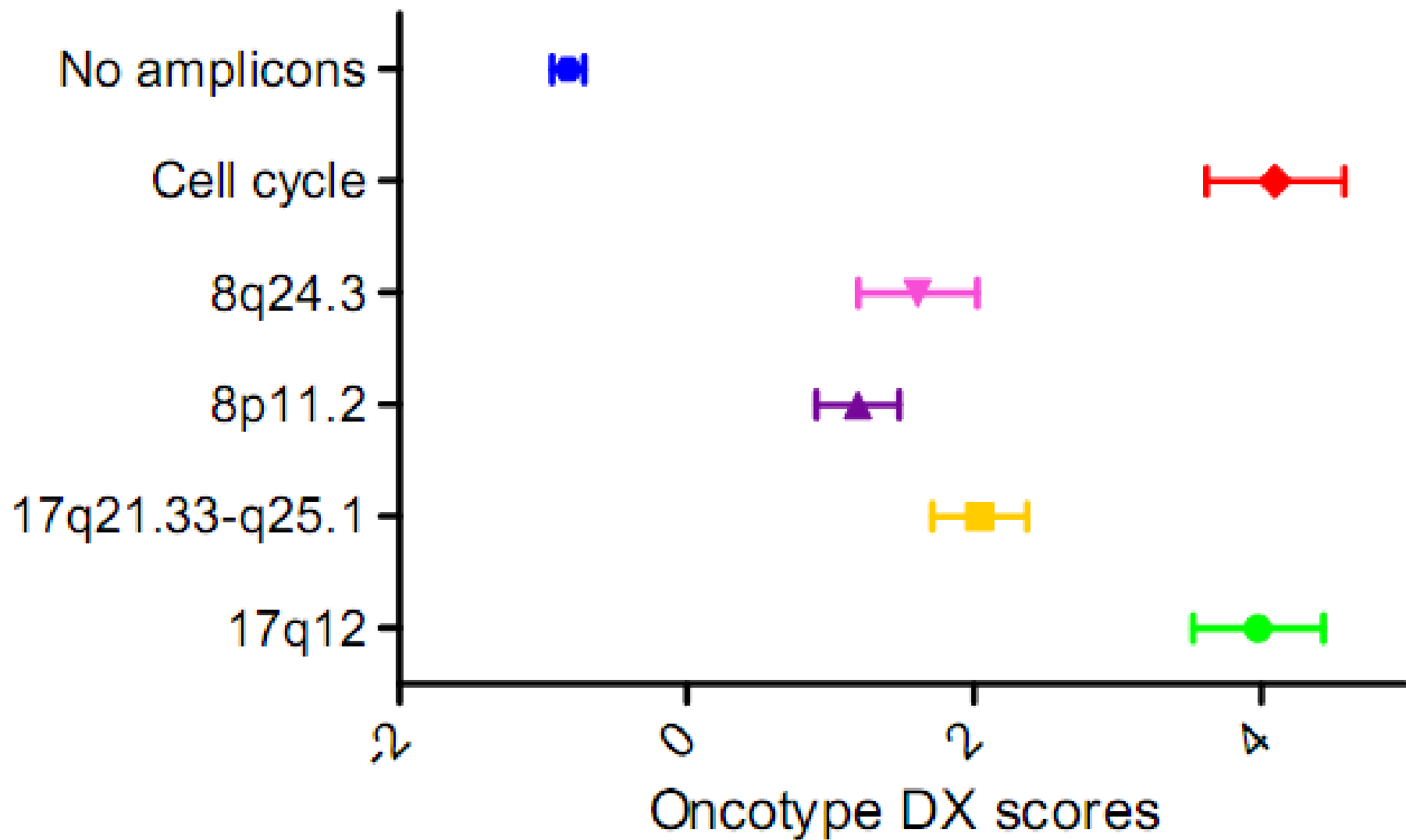


Amplicon	Med survival (days)	Hazard ratio	95% CI	p-value	H-Grade tumor enrich. p-value	Node stat association p-value	Oncogene
17q12	3355	4.1	3.8 – 21.0	<0.0001	0.0002	0.85	ERBB2
17q21.33-q25.1	—	3.1	2.2 - 13.6	0.0003	0.2	0.86	?
8p11.2	3795	3.8	3.2 - 18.3	<0.0001	0.04	0.83	? LSM1
8q24.3	3468	4.3	4.3- 34.1	<0.0001	0.002	0.86	? HSF1

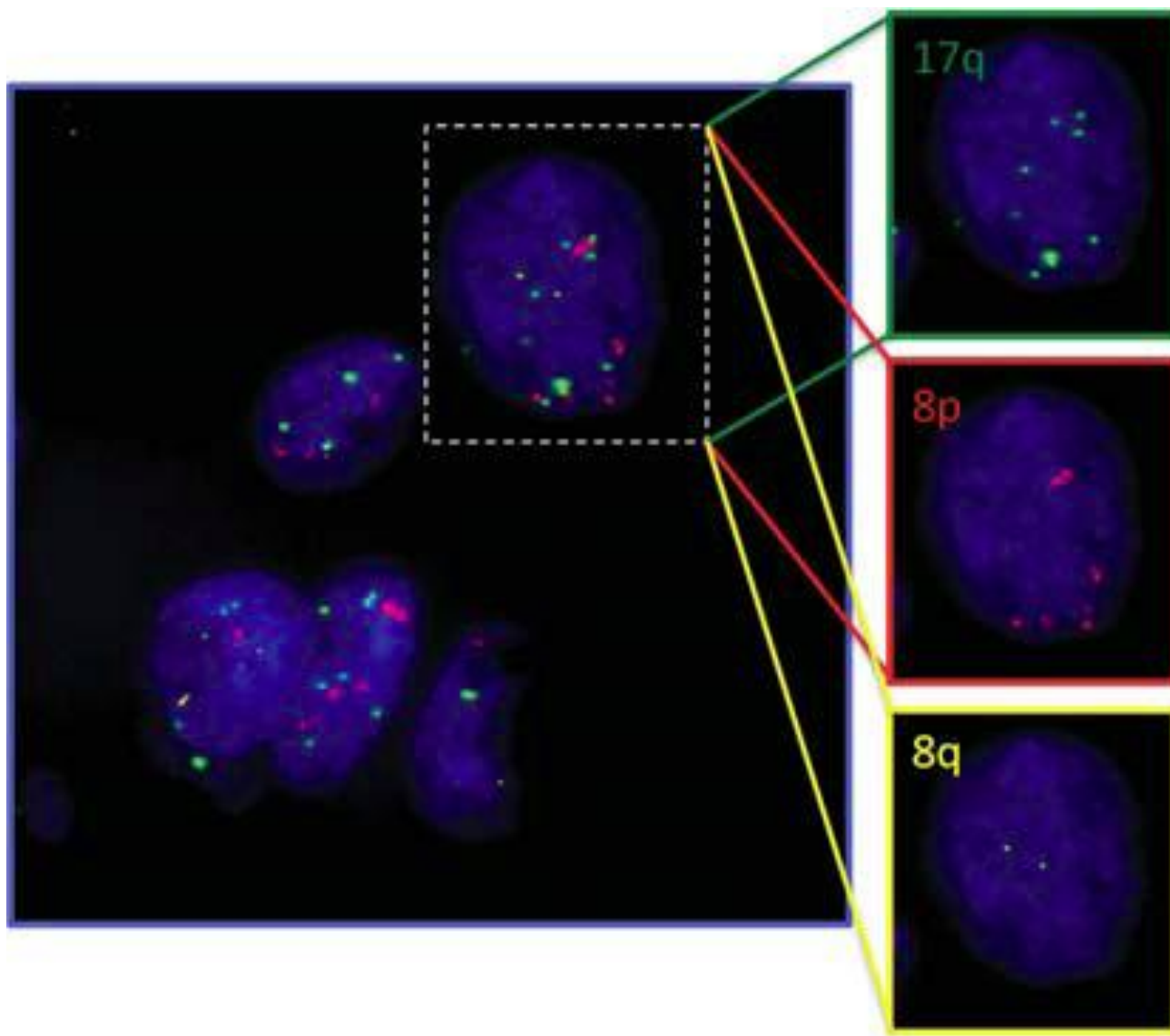
Survival for intermediate grade tumors in GSE6532 (gene expression) and GSE22133 (CGH array) datasets

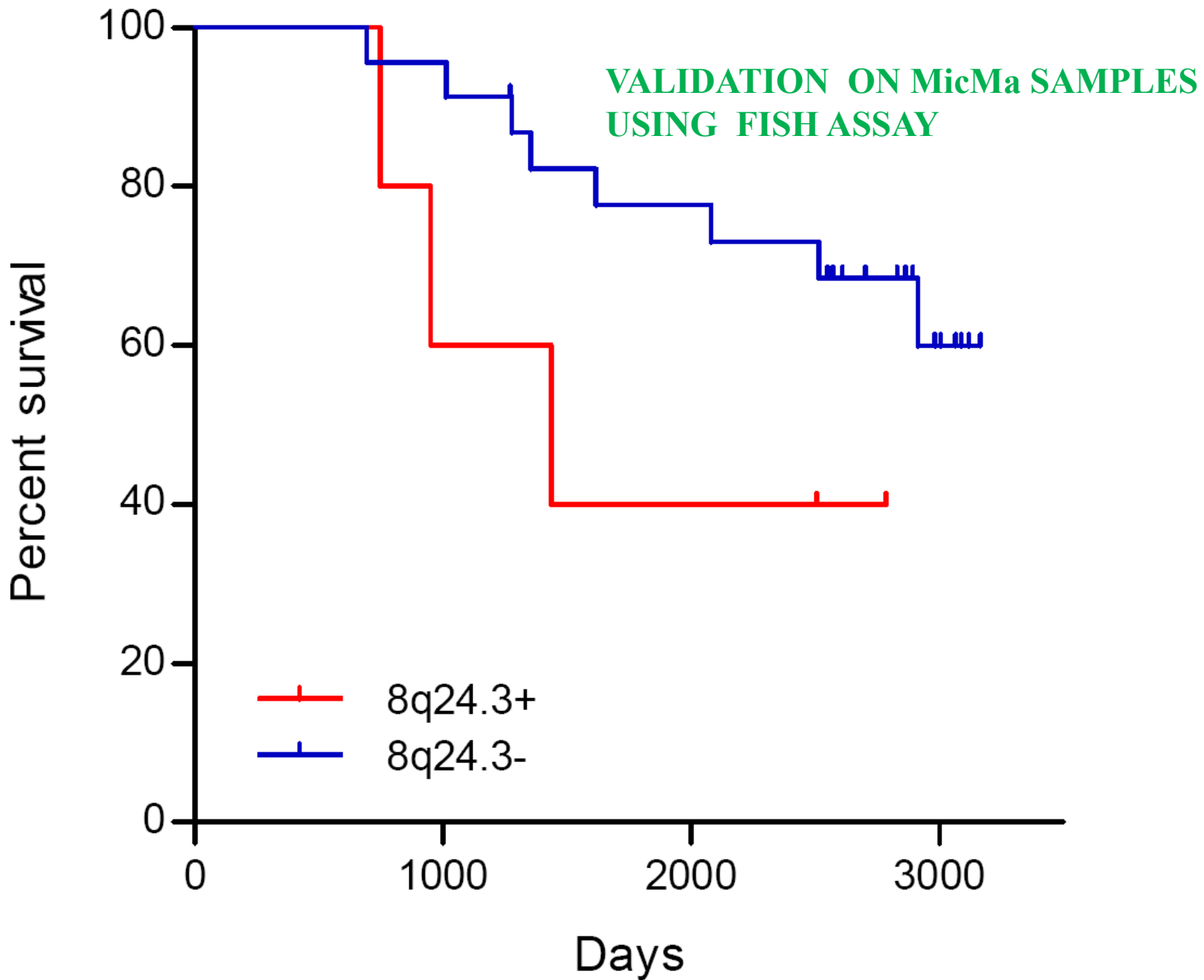


Amplicons and Inferred Oncotype Dx Score



Multiplex FISH Assay to detect Amplicons





Acknowledgments



Erhan Bilal
Rutgers/IBM



Gabriela Alexe
MIT/Harvard



Michael Seiler
Rutgers



Vanessa
Almendro
DFCI



Hege Russnes
DFCI/OUH



Arnie Levine
IAS/CINJ



Shridar Ganesan
CINJ



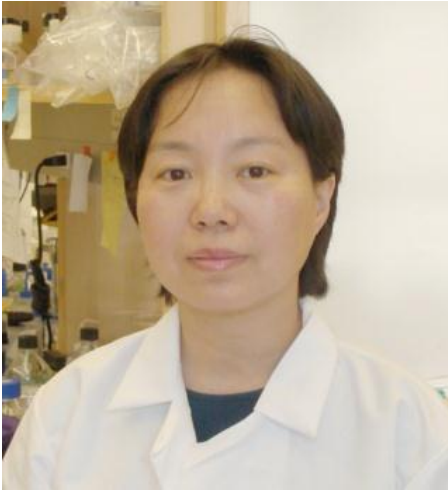
Anne-Lise
Børresen-Dale
OUH



Vessela
Kristensen
OUH

The Ganesan Lab

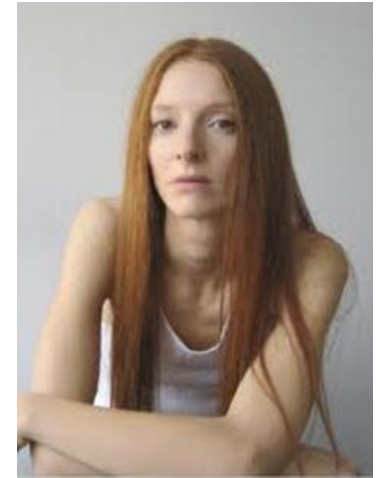
(Where The Real Biology Happens)



Ming Yao



Shridar Ganesan



Sunniva Bjorklund



Vasudeva Ginjala



Atul Kulkarni



Jay Oza