International BCB-Workshop on
Machine Learning in Bioinformatics
October 10, 2005 - Berlin

Predicting clinical outcome of breast cancer patients treated with Tamoxifen using gene expression data

James F. Reid

http://bio.ifom-firc.it/User/reid/

IFOM
ISTITUTO FIRC
DI ONCOLOGIA MOLECOLARE

Molecular Genetics of Cancer Group
http://www.ifom-ieo-campus.it/

ISTITUTO NAZIONALE
PER LO STUDIO
E LA CURA DEI TUMORI

Motivation

- Urgent need of new prognostic factors for treatment outcome of breast cancer patients
- Predicting the benefit of a particular treatment can prevent over-treating and tailor more beneficial treatment for individual patients

Microarray breast cancer studies

- Several studies have demonstrated that breast cancers with distinct pathologic features can be recognized by their gene expression profile.
- Microarrays have been used to identify expression patterns capable of predicting outcome or response after specific treatments

Tamoxifen treatment

- Breast cancer is estrogen-dependent, reducing estrogen secretion can cause the cancer to regress
- Tamoxifen is a standard adjuvant treatment for patients with primary, estrogen-receptor-positive breast cancer.
- It acts by inhibiting the biding of estrogen to estrogen receptors
- BUT approx. 40% of ER breast cancers do not respond to treatment and develop resistance leading to disease progression.
- ‘Standard’ clinicopathological features (e..g tumor stage/grade, PGR, ERBB2 etc.) do not work accurately.

Limiting factors of microarray based predictive studies

- Limited starting material
- New and expensive technology

→ poor experiment design

→ push for discovery and lack of validation

A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen

Xiao-Jun Ma,[1] Zuncai Wang,[2] Paula D. Ryan,[3] Steven J. Isakoff,[4,5] Anne Barmettler,[2] Andrew Fuller,[2] Beth Muir,[2] Gayatry Mohapatra,[2] Ranelle Salunga,[1] J. Todd Tuggle,[1] Yen Tran,[1] Diem Tran,[1] Ana Tassin,[1] Paul Amon,[1] Wilson Wang,[1] Wei Wang,[1] Edward Enright,[1] Kimberly Stecker,[1] Eden Estepa-Sabal,[1] Barbara Smith,[3] Jerry Younger,[3] Ulysses Balis,[2] James Michaelson,[2] Atul Bhan,[2] Karleen Habin,[3] Thomas M. Baer,[1] Joan Brugge,[4] Daniel A. Haber,[3] Mark G. Erlander,[1] and Dennis C. Sgroi[2]

[1]Arcturus Bioscience, Inc., 2715 Loker Avenue West, Carlsbad, California 92008
[2]Department of Pathology, Harvard Medical School, Molecular Pathology Research Unit, Massachusetts General Hospital, Boston, Massachusetts 02129
[3]Massachusetts General Hospital Cancer Center, Harvard Medical School, Boston, Massachusetts 02129
[4]Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115
[5]Department of Adult Oncology, Dana Farber Cancer Institute, Boston, Massachusetts 02115
*Correspondence: dsgroi@partners.org (D.C.S.), merlander@arctur.com (M.G.E.)

Cancer Cell, June 2004

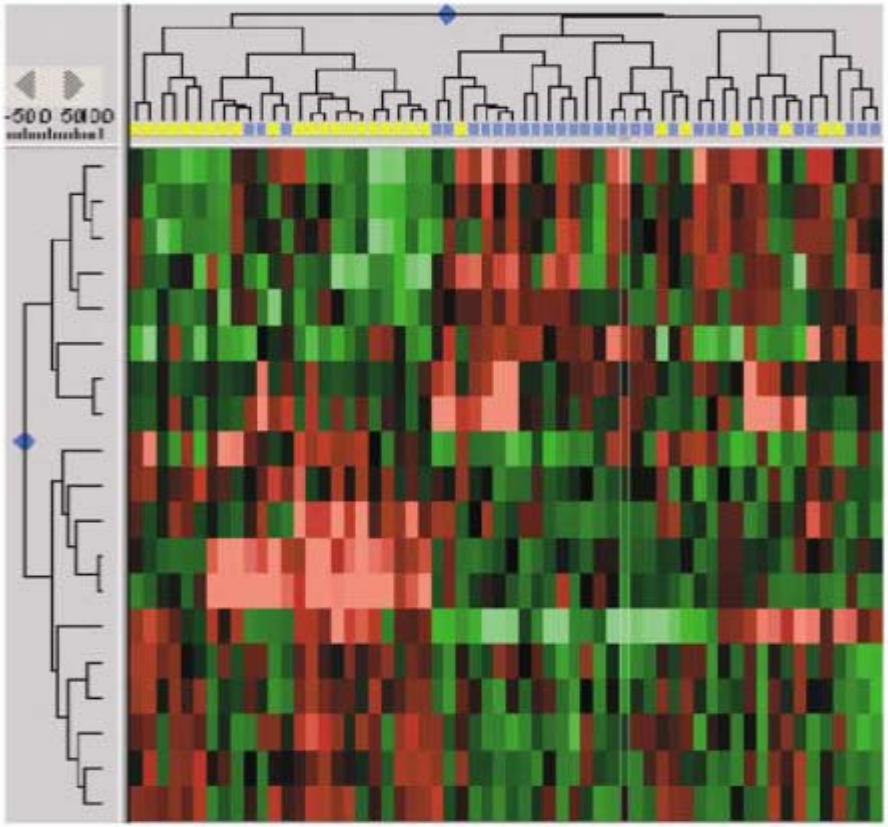
Ma et al. case material

Table 1. Patient and tumor characteristics in this study

		Cohort 1 (frozen)		Cohort 2 (FFPE)	
		Recurrence	Nonrecurrence	Recurrence	Nonrecurrence
Size (cm)	Total	28	32	10	10
	Mean	2.7	2.1	1.9	1.7
	Range	0.9–4.7	0.8–5.5	1.1–4.0	0.8–4.0
Grade	1	2	1	1	1
	2	15	24	6	8
	3	11	7	3	1
Nodes*	0	13	15	8	10
	1–3	6	11	1	0
	>3	6	2	0	0
Age	Mean	65.1	69.1	65.5	65.2
	Range	48–84	54–85	54–93	57–74
DFS (Months)	Mean	54.8	115.6	51.4	95.8
	Range	5–137	61–169	15–117	25–123
Receptor status	ER+	27/28	32/32	10/10	10/10
	PR+	23/28	27/32	8/9*	10/10

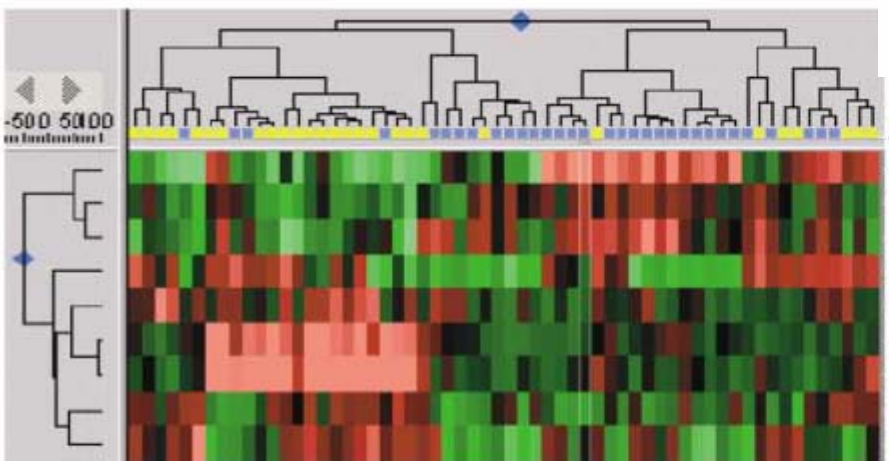
*Cases with missing data omitted. DFS, disease-free survival; FFPE, formalin-fixed and paraffin-embedded.

A



Accession	P	Description
AF208111	**9.9e-5**	**IL17BR | interleukin 17B receptor**
AI240933	**2.4e-5**	**EST**
AJ272267	0.0002	CHDH | choline dehydrogenase
AW613732	0.0006	Homo sapiens cDNA FLJ31137 fis, clone I.
AK027250	0.001	Homo sapiens cDNA: FLJ23597 fis, clone L.
AL050227	0.0006	Homo sapiens mRNA; cDNA DKFZp586M0.
N30081	0.0002	ESTs, Weakly similar to I38022 hypothetical.
X81896	0.0009	C11orf25 | chromosome 11 open reading fr.
M92432	0.0005	GUCY2D | guanylate cyclase 2D, membrane.
AB000520	9.6e-5	APS | adaptor protein with pleckstrin homol.
BC004960	0.0009	MGC10955 | hypothetical protein MGC10955
AI700363	**0.0004**	**HOXB13 | homeo box B13**
BC007092	**0.0004**	**HOXB13 | homeo box B13**
AL117406	0.0004	ABCC11 | ATP-binding cassette, sub-family.
BC007783	0.0009	SCYA3 | small inducible cytokine A3
AW006861	2.0e-5	SCYA4 | small inducible cytokine A4
X59770	8.1e-5	IL1R2 | interleukin 1 receptor, type II
AI820604	0.0002	ESTs
AI087057	0.0002	DOK2 | docking protein 2, 56kD

B



-8 -4 -2 1 2 4 8 Fold change

Accession	P	Description
AI688494	0.0005	FLJ13189 | hypothetical protein FLJ13189
AI240933	**0.0005**	**EST**
AF208111	**0.0004**	**IL17BR | interleukin 17B receptor**
AF033199	0.0004	ZNF204 | zinc finger protein 204
AL157459	0.0007	Homo sapiens mRNA; cDNA DKFZp434B0.
AI700363	**0.0003**	**HOXB13 | homeo box B13**
BC007092	**2.7e-5**	**HOXB13 | homeo box B13**
BC002480	0.0008	FLJ13352 | hypothetical protein FLJ13352
NM_014298	0.0004	QPRT | quinolinate phosphoribosyltransfera.

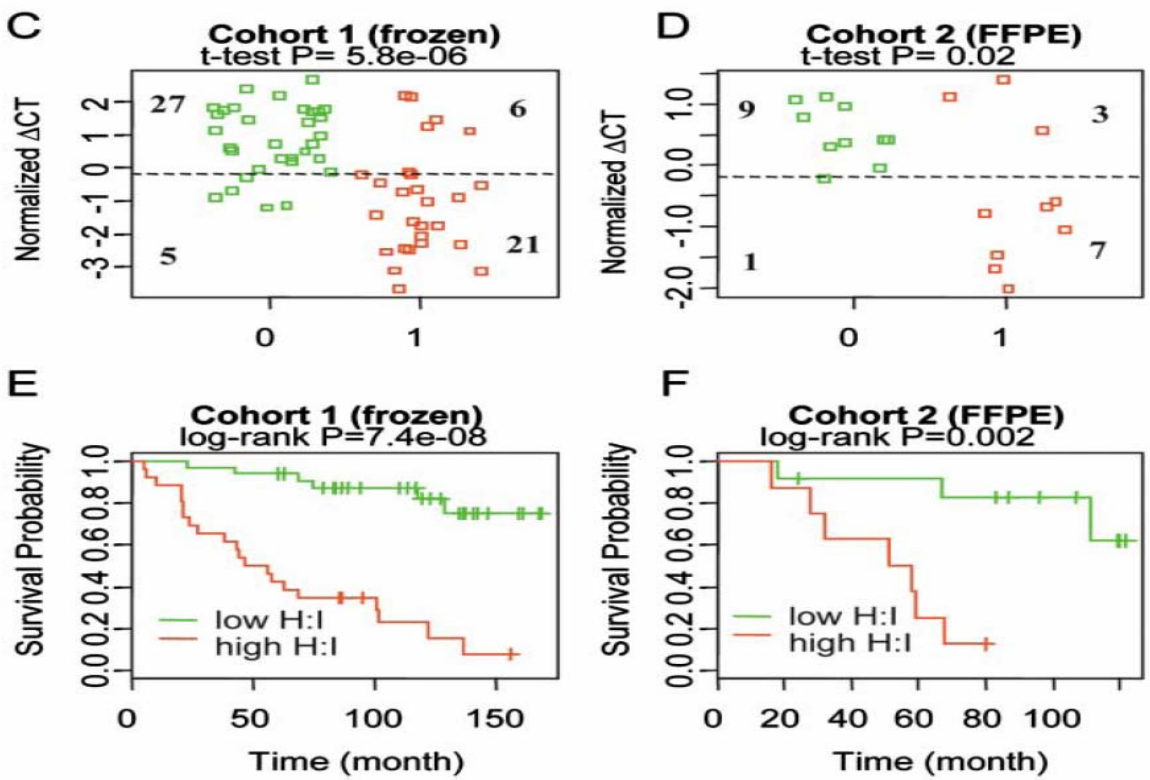
A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen.

Ma XJ, Wang Z, Ryan PD, Isakoff SJ, Barmettler A, Fuller A, Muir B, Mohapatra G, Salunga R, Tuggle JT, Tran Y, Tran D, Tassin A, Amon P, Wang W, Wang W, Enright E, Stecker K, Estepa-Sabal E, Smith B, Younger J, Balis U, Michaelson J, Bhan A, Habin K, Baer TM, Brugge J, Haber DA, Erlander MG, Sgroi DC.

Cancer Cell. 2004 Jun;5(6):607-16

"The HOXB13:IL17BR expression ratio may be useful for identifying patients appropriate for alternative therapeutic regimens in early-stage breast cancer"

Authors' validation on an independent group of 20 patients but with 7 out of 10 correctly classified (P-value=0.34) [0: non-recurrent, 1: recurrent]



RE: A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen

We recently reported a novel biomarker of two-gene expression ratio that outperformed the current positive and negative predictors of outcome in patients with estrogen receptor (ER)-positive early-stage breast cancer treated with adjuvant tamoxifen (Ma et al., 2004). While the cases in our original 60-patient discovery cohort were closely matched for tumor size, grade, and lymph node status (28 cases node-negative, 25 node-positive and 7 cases not evaluated), almost all cases (19/20) in the initial validation cohort were lymph node-negative. The bias toward lymph node-negative patients in our initial validation cohort was not by design, but was consistent with the fact that lymph node-positive patients usually receive chemotherapy in addition to adjuvant tamoxifen (EBCTCG, 1998) and would have been excluded from our study. Subsequent to this publication, we have further assessed the predictive power of this two-gene ratio biomarker in an independent cohort of breast cancer patients from a randomized prospective clinical trial of adjuvant tamoxifen (D. Sgroi et al., 2004, ASCO Annual Meeting Proceedings, abstract). Results from this study confirm our initial observations and suggest that the two-gene signature is a more robust predictor in lymph node negative patients, as compared with lymph node positive patients. We are currently carrying out additional studies using archived tissue samples from large randomized prospective trials of adjuvant tamoxifen to further evaluate the clinical utility of our gene expression ratio biomarker.

Dennis C. Sgroi,[1,2,*] Daniel A. Haber,[1] Paula D. Ryan,[1] Xiao-Jun Ma,[3] and Mark G. Erlander[3]

[1]Department of Pathology, Harvard Medical School
Molecular Pathology Research Unit
Massachusetts General Hospital
Boston, Massachusetts 02129
[2]Massachusetts General Hospital Cancer Center
Harvard Medical School
Boston, Massachusetts 02129
[3]Arcturus Bioscience, Inc., 400 Logue Avenue
Mountain View, California 94043
*Correspondence: dsgroi@partners.org

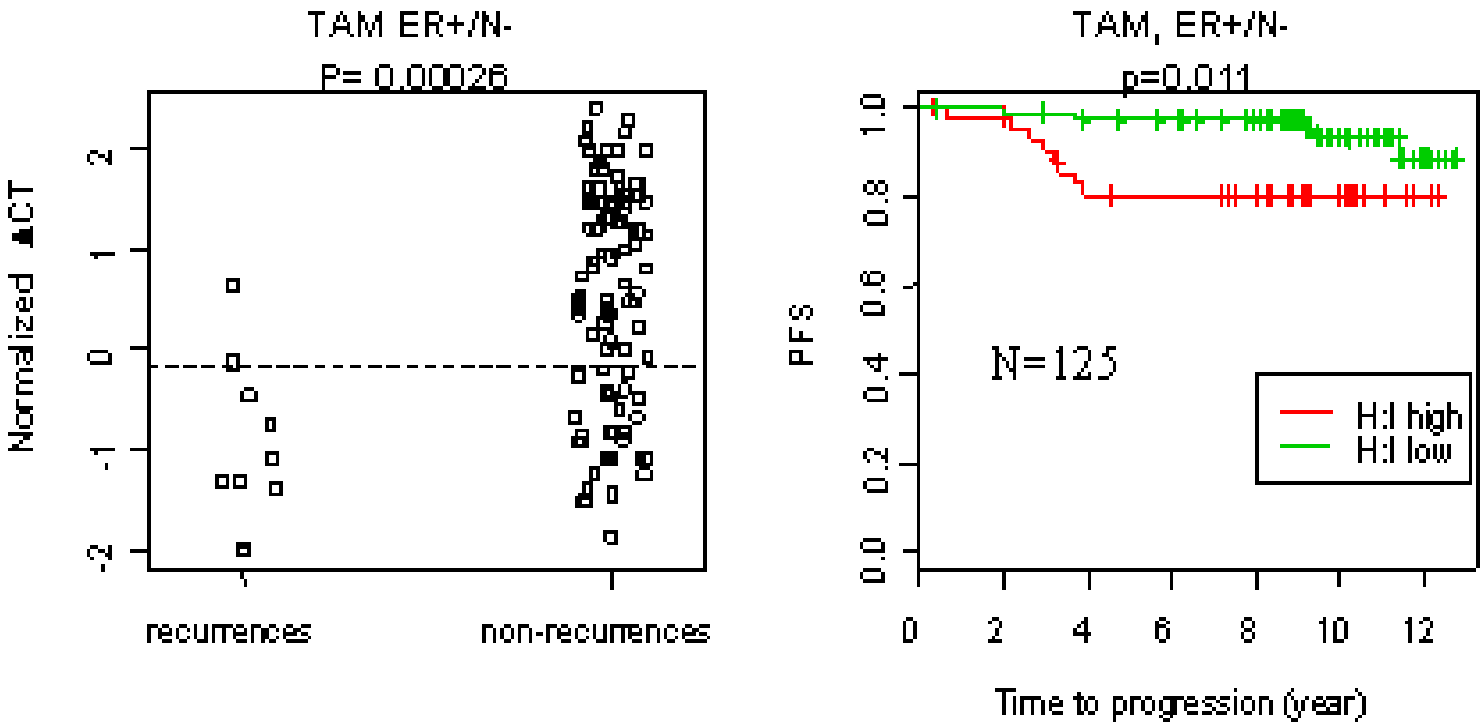
References

Ma, X.J., Wang, Z., Ryan, P.D., Isakoff, S.J., Barmettler, A., Fuller, A., Muir, B., Mohapatra, G., Salunga, R., Tuggle, J.T., et al. (2004). A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. Cancer Cell *5*, 607–616.

Early Breast Cancer Trialists' Collaborative Group (EBCTCG). (1998). Tamoxifen for early breast cancer: An overview of the randomized trials. Lancet *351*, 1451–1467.

Cancer Cell 20 Nov. 2004

Additional data presented by Ma et al.



Authors' validation on an independent group of 125 patients, Randomized Trial NCCCTG#89-30-52 (from ASCO'04)

INT patient and tumor characteristics

	Total cases (58)	Disease-free (40)	Relapsed (18)
Age (yrs):			
≤50	4	3	1
>50	54	37	17
Nodal Status*:			
N-	13	11	2
N+	45	29	16
Tumor size(cm):			
≤2	22	16	6
>2	36	24	12
PgR (LBA) §:			
Negative	12	7	5
Positive	46	33	13

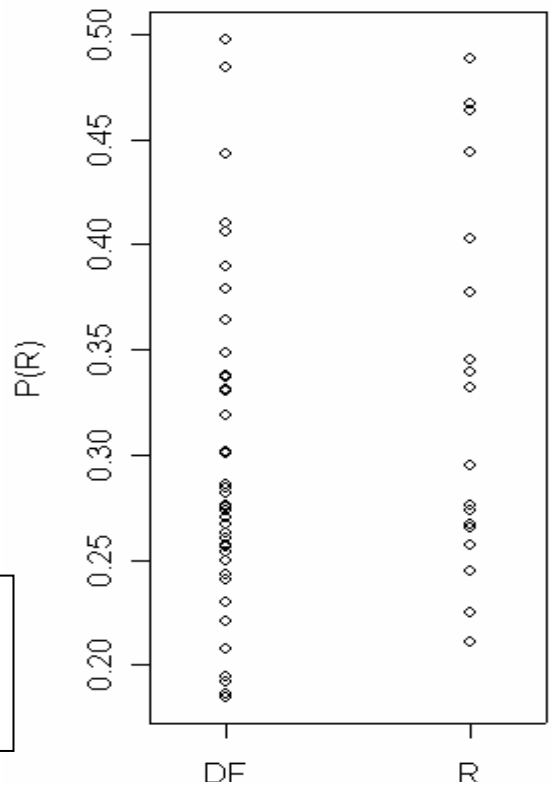
* N-: no lymph nodes, N+: 1 to 10 lymph nodes § Negative ≤ 25 pmol/mg protein

Main difference with case material from Ma et al. is the lymph node status (LN): our cases are mostly positive

Limits of Predictive Models Using Microarray Data for Breast Cancer Clinical Treatment Outcome

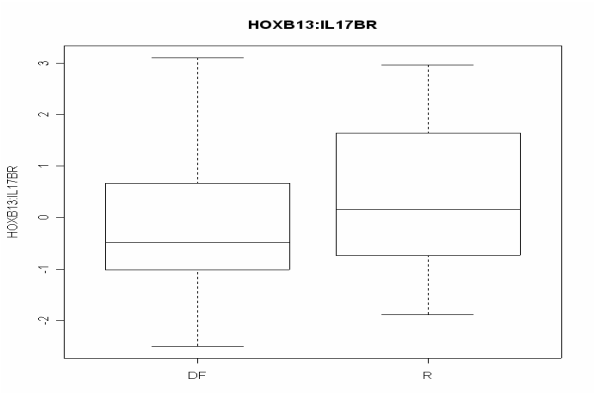
James F. Reid, Lara Lusa, Loris De Cecco, Danila Coradini, Silvia Veneroni, Maria Grazia Daidone, Manuela Gariboldi, Marco A. Pierotti

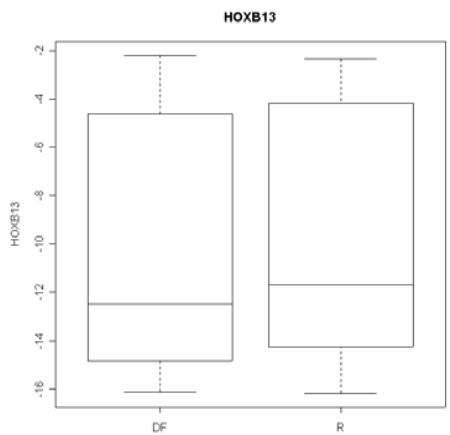
Journal of the National Cancer Institute, Vol. 97, No. 12, June 15, 2005



Probability of experiencing recurrence stratified by recurrence status (DF=Disease Free, R=Recurrent), estimated by logistic model including log(HOXB13:IL17BR) as a covariate

The two genes DO NOT predict clinical outcome in our samples (independent validation)





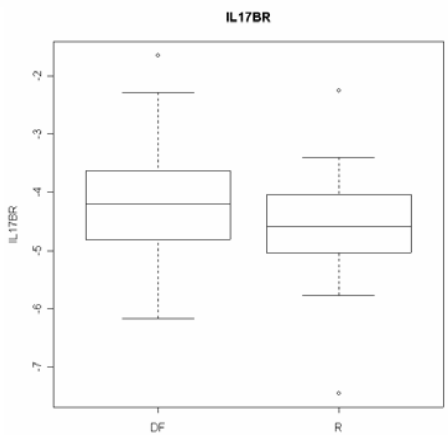


Table 1. Association and discrimination of reverse transcription–quantitative polymerase chain reaction expression data from 58 primary estrogen receptor-positive, lymph node-positive breast cancers from patients treated to adjuvant monotherapy with tamoxifen

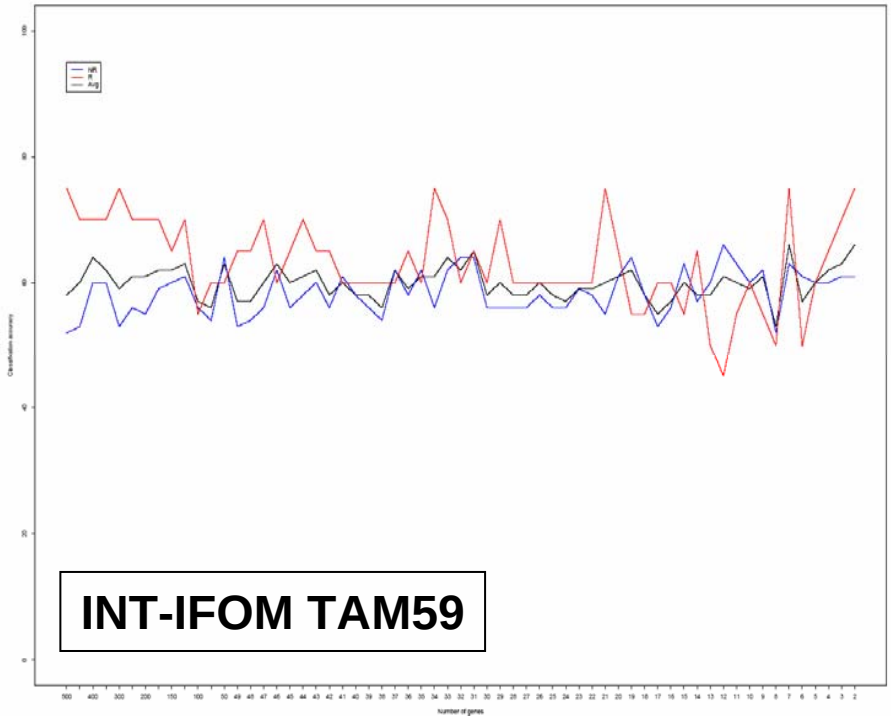
Analysis	HOXB13	IL17BR	HOXB13/IL17BR
Mean comparison*			
mean(DF)-mean(R)	-0.85	0.42	-0.55
95% CI	(-3.74 to 2.05)	(-0.22 to 1.06)	(-1.42 to 0.31)
t test *P*	.56	.19	.20
Mann–Whitney *P*	.49	.21	.23
AUC†			
Coefficient	0.55	0.59	0.58
95% CI	(0.40 to 0.71)	(0.43 to 0.75)	(0.41 to 0.74)
P	.51	.27	.20
Logistic regression‡			
Odds ratio	1.04	0.69	1.30
95% CI	(0.92 to 1.16)	(0.40 to 1.20)	(0.88 to 1.93)
P	.54	.18	.18

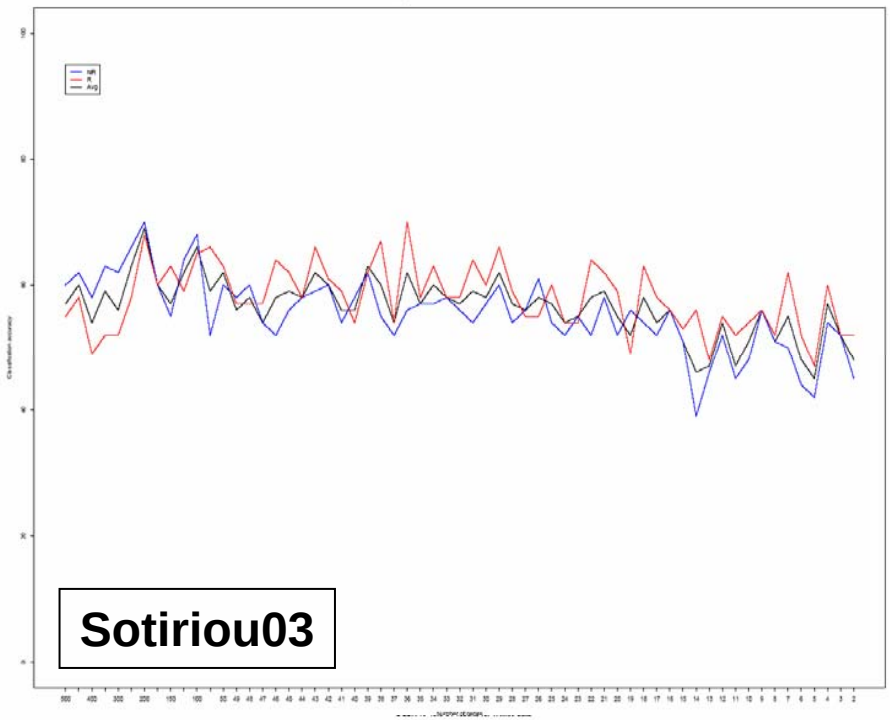
Predicting response to Tamoxifen in three breast cancer series

Datasets

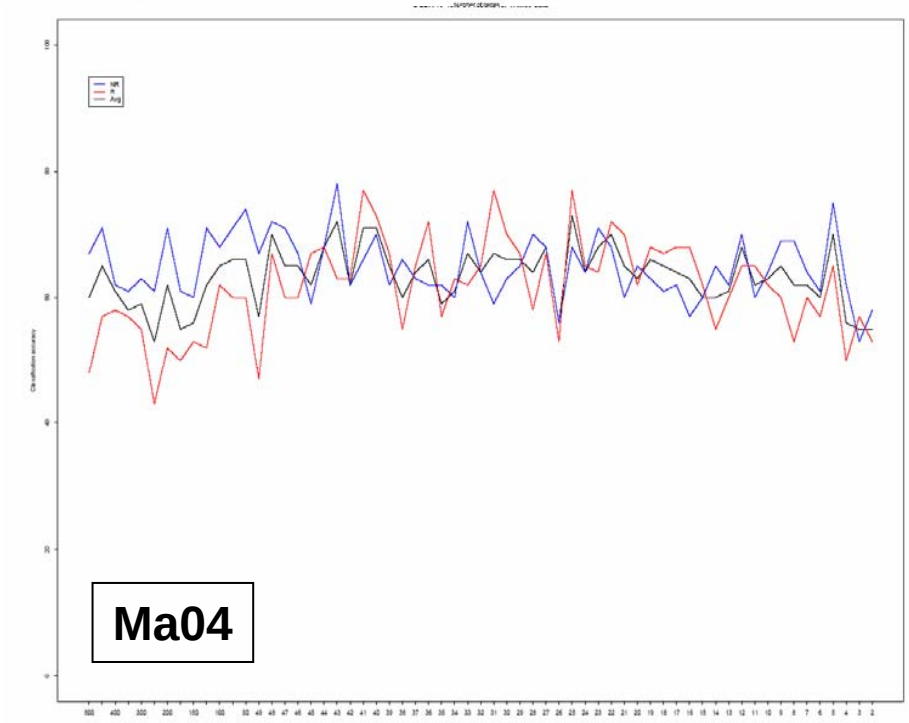
- IFOM TAM59 (18R,41N)
  cDNA microarray (18K IMAGE clones)
- Sotiriou03 (36R, 42N)
  cDNA microarray (8K IMAGE clones)
- Ma04 (28R, 32N)
  Oligo ink-jet microarray (22K Agilent 60-mer)

Experiment design of all three studies used a common reference design (Stratagene)

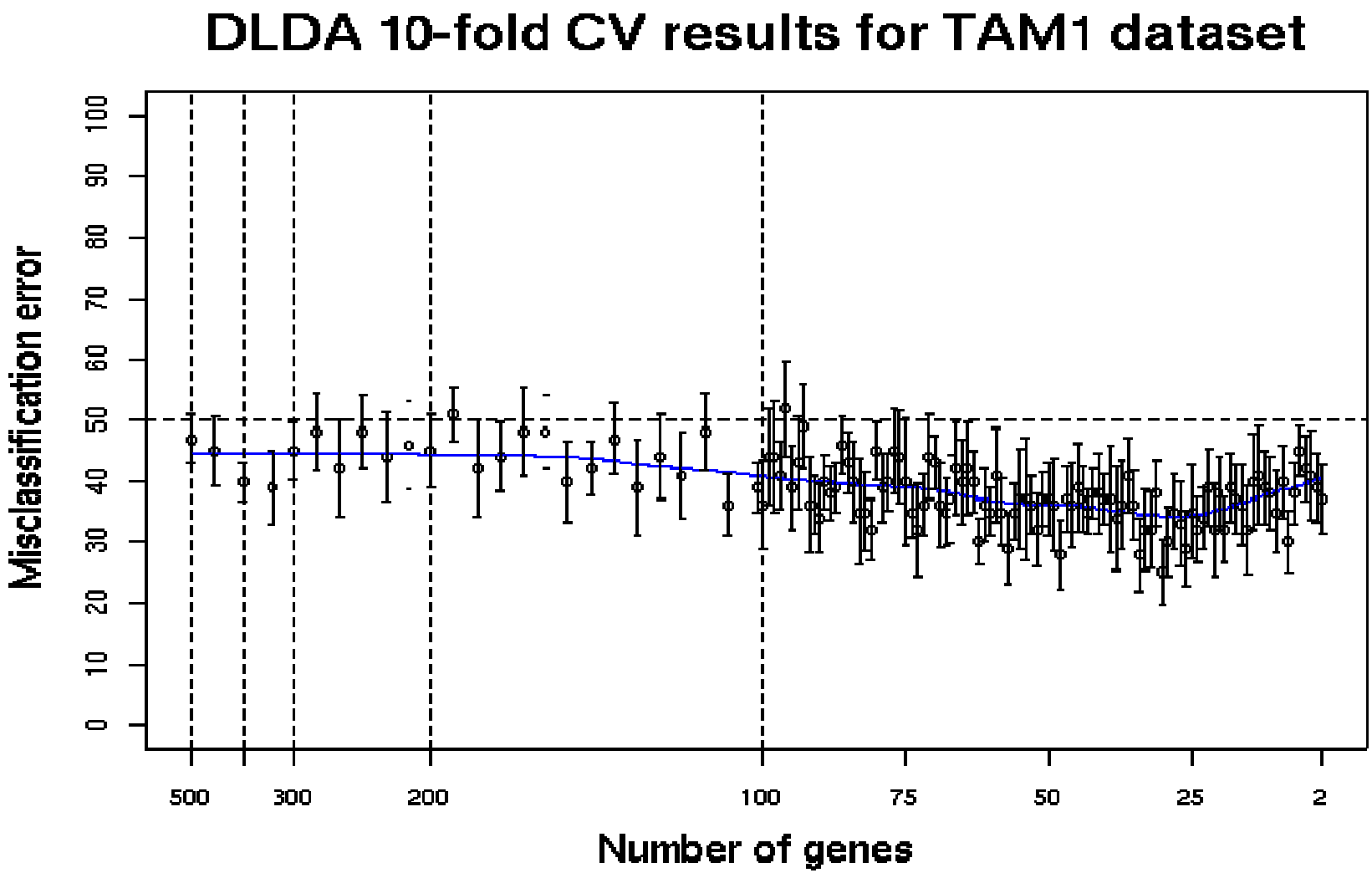




DLDA analysis for predicting response to Tamoxifen



Cross-validation is probably useful



Perspectives

- **What are the limiting factors in building predictive models from gene expression profiling experiments?**
  - **Size and homogeneity**
    - **sample selection, sample numbers**
    - **number of probed transcripts**
  - **Meta-analysis**
    - **cross-platform comparisons**
    - **cross-laboratory comparisons**

A Roadmap

- Develop classifier for addressing a specific important therapeutic decision
- Patients are sufficiently homogeneous and receiving uniform treatment so that results are therapeutically relevant
- Treatment options and costs of mis-classification are such that a classifier is likely to be used
- Perform internal validation of classifier to assess whether it appears sufficiently accurate relative to standard prognostic factors that it is worth further development
- Translate classifier to platform that would be used for broad clinical application
- Demonstrate that the classifier is reproducible
- Independent validation of the completely specified classifier on a prospectively planned study

From Table 1 in “Roadmap for Developing and Validating Therapeutically Relevant Genomic Classifiers”, R. Simon to appear in JCO

Aknowledgements

Molecular Genetics of Cancer group:

Loris De Cecco
Lara Lusa
Manuela Gariboldi

Marco A. Pierotti

Unit of Tumour Biology and Experimental Therapy:

Silvia Veneroni
Danila Coradini

Maria Grazia Daidone

IFOM
ISTITUTO FIRC
DI ONCOLOGIA MOLECOLARE

ISTITUTO NAZIONALE
PER LO STUDIO
E LA CURA DEI TUMORI