

Breakout Session 2B: Disease Classifier Assays

Session Facilitators: Dr. Fred Waldman,
Dr. Mei Polley, Dr. Magdalena Thurin

1. INTRODUCTION

Morning session

The morning session will focus on the **definitions** and **evidence needed** to achieve clinical utility using illustrative two case examples.

A. Definition of clinical utility of prognostic assays

- Does clinical utility of a prognostic assay apply only in the setting of a difficult clinical decision regarding treatment/management?
- Must the assay drive a treatment/management decision?
- Is it necessary to understand the biology behind the clinical association?
- Must clinical utility always relate to improved clinical outcome?
- What are other factors in play and how should we balance among them? Efficacy, toxicity, cost of drug and cost of assay?

B. Evidence generation

- Does the assay work with the specimens of choice?
- Does the assay clearly define the population of interest? Does evidence always have to be generated in the same organ or context it will be used?
- What are clinically relevant endpoints? Does the choice depend on disease?
- Is it sufficient to identify different risk groups? How large does a difference in outcome need to be?
- Evidence-based approach: what are common statistical issues?

Afternoon session

The afternoon session will focus on **how to generate the data we need and how organizations such as the NCI can help facilitate and accelerate evaluation of clinical utility?**

C. Source of data

- Randomized controlled trials are the standard? Why or why not? Are there any cases in which nothing less can be considered?
- What are other options? When are they appropriate and when are they not?
- How do low-prevalence markers and small sample sizes affect the ability to evaluate clinical utility?

D. Role of NCI

- Does the NCI have a unique role?
 - Can the NCI help fund the development of promising assays?
 - Can the NCI help facilitate interactions with the FDA?
- Could the clinical trials network be utilized differently to generate the evidence? What are the hurdles?

2. DIAGNOSTIC, PROGNOSTIC OR PREDICTIVE

Diagnostic Assays (Session2A)

An assay that classifies disease into distinct subtypes. Assays of this type usually arise from *unsupervised* clustering analysis.

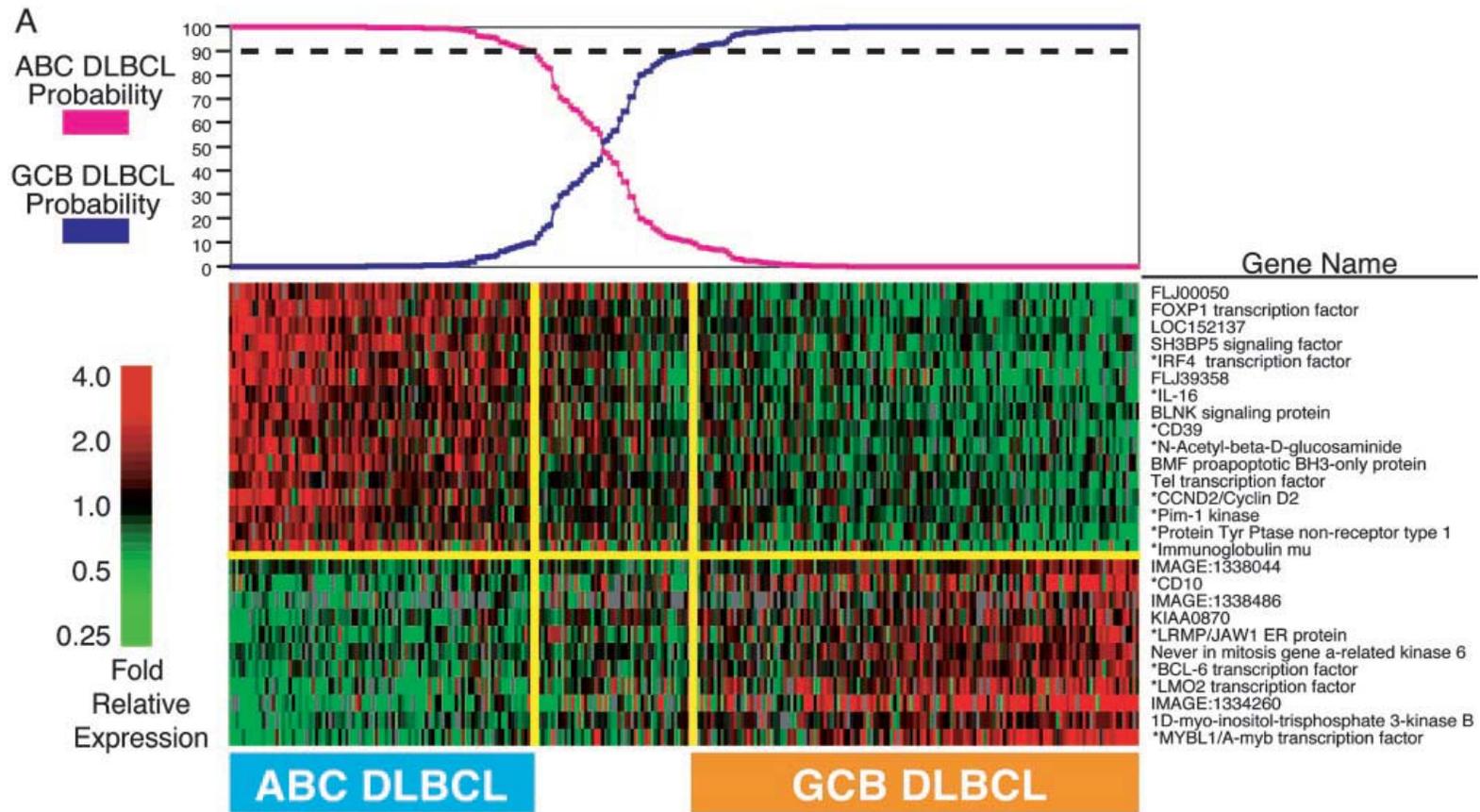
Example: A gene expression-based method to diagnose distinct subgroups of diffuse large B cell lymphoma (DLBCL)



A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma

George Wright*, Bruce Tan†, Andreas Rosenwald†, Elaine H. Hurt†, Adrian Wiestner†, and Louis M. Staudt†‡

*Biometric Research Branch, Division of Cancer Treatment and Diagnosis and †Metabolism Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892



B

DLBCL Subgroup by Hierarchical Clustering

		Model Prediction		
		ABC	GCB	Other
DLBCL Subgroup by Hierarchical Clustering	ABC	37	1	4
	GCB	1	58	8

Training Set

		Model Prediction		
		ABC	GCB	Other
	ABC	38	1	2
	GCB	2	57	8
	Type 3	14	18	25

Validation Set

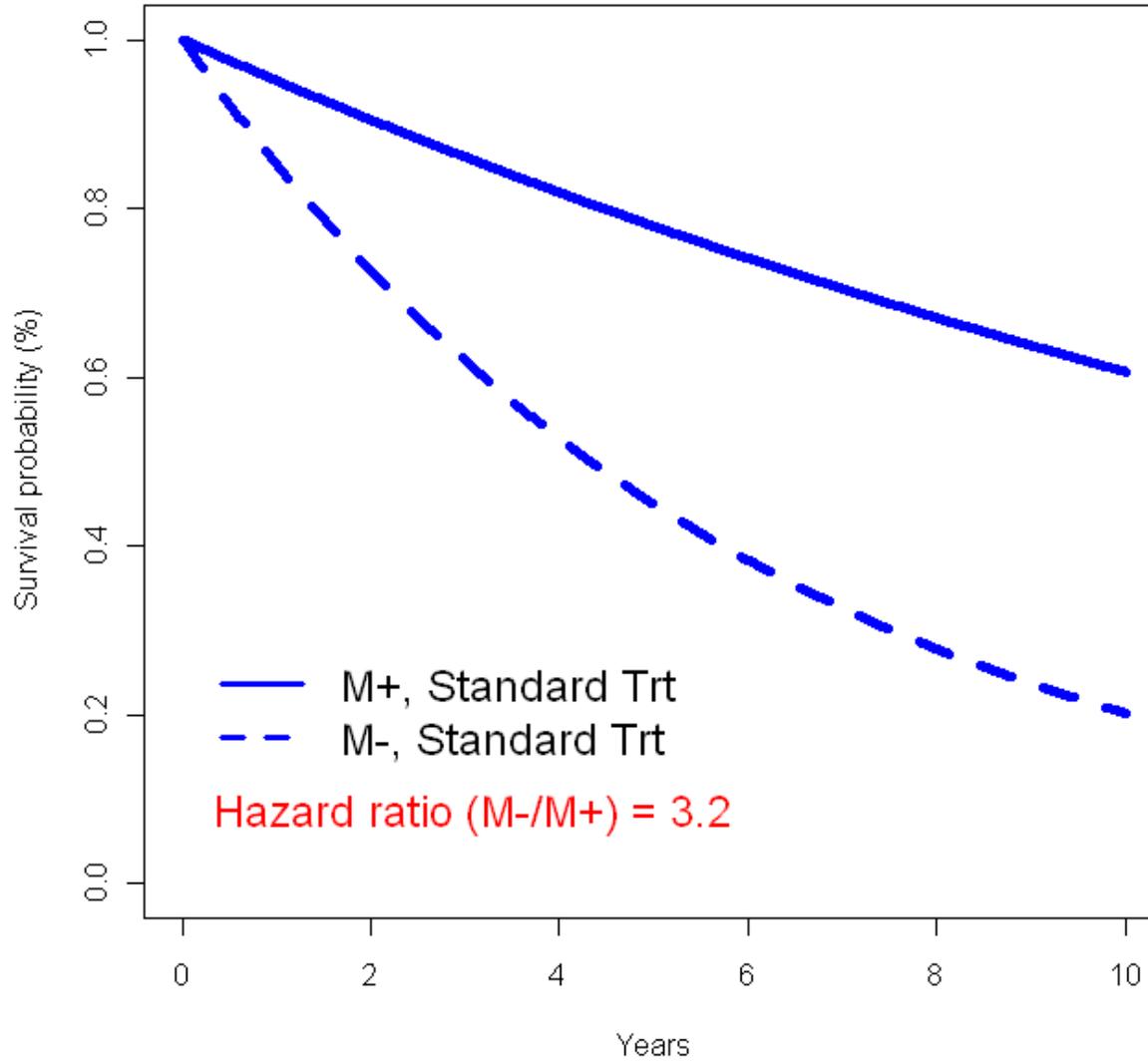
		Model Prediction		
		ABC	GCB	Other
	ABC	75	2	6
	GCB	3	115	16
	Type 3	14	18	25

All Samples

Prognostic Assay (Session 2B)

Biomarker measurement (test result) associated with clinical outcome in absence of therapy (natural course) *or with standard therapy all patients are likely to receive*

- **Clinical use:** Identify patients who have very favorable outcome in absence of *(additional)* therapy or who have poor outcome and might benefit from more aggressive *(additional)* therapy
- **Research use:** Disease biology, identify drug targets, stratification factor in clinical trials



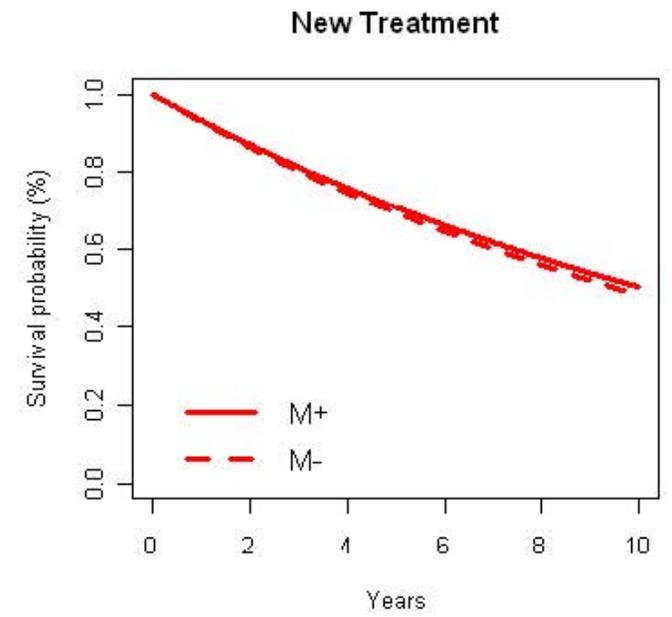
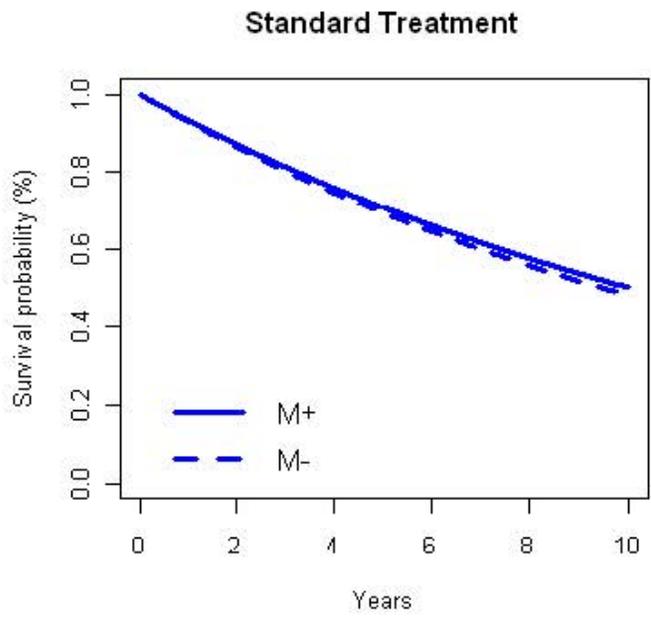
Predictive Assay (Session 1)

Biomarker measurement (test result) associated with benefit or lack of benefit (potentially even harm) from a particular therapy relative to other available therapy

- **Clinical use:** Select one treatment versus another treatment

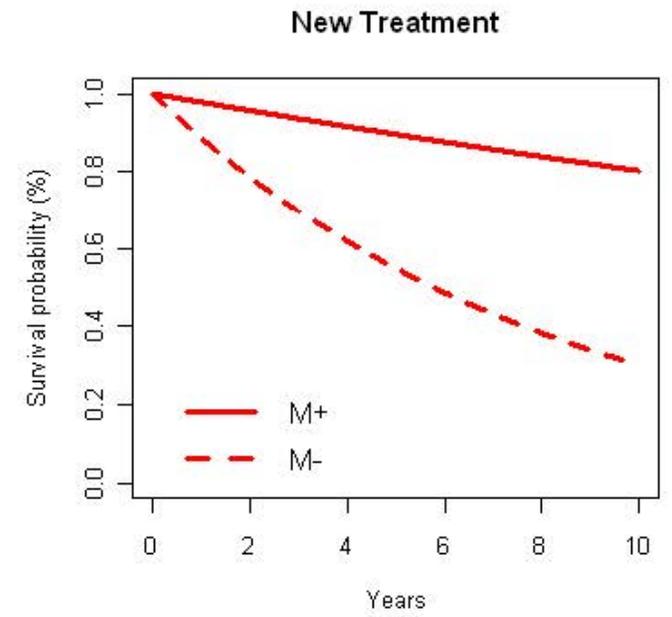
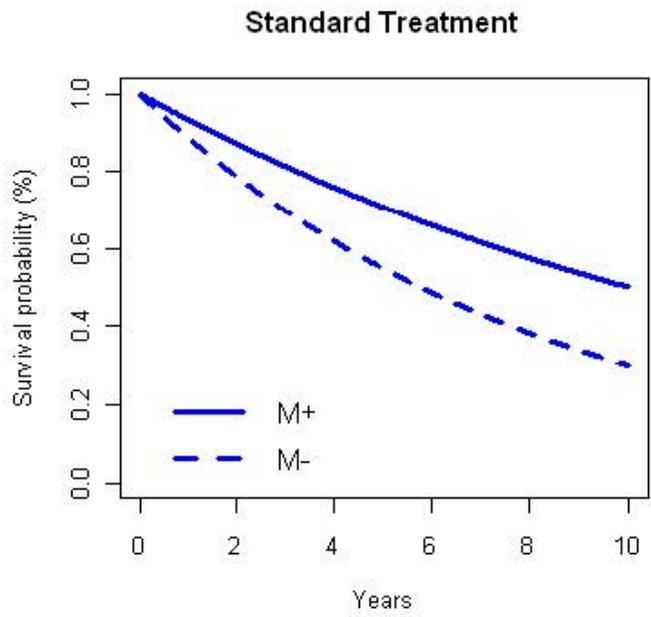
(I)

~~prognostic~~
~~predictive~~



(II)

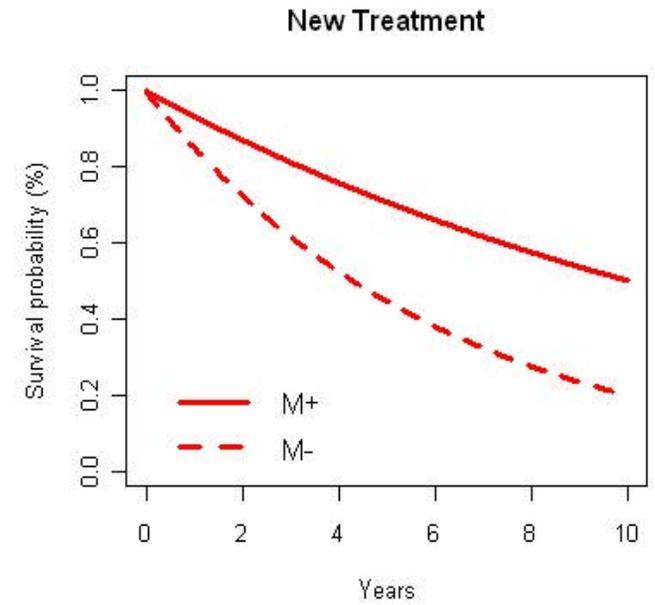
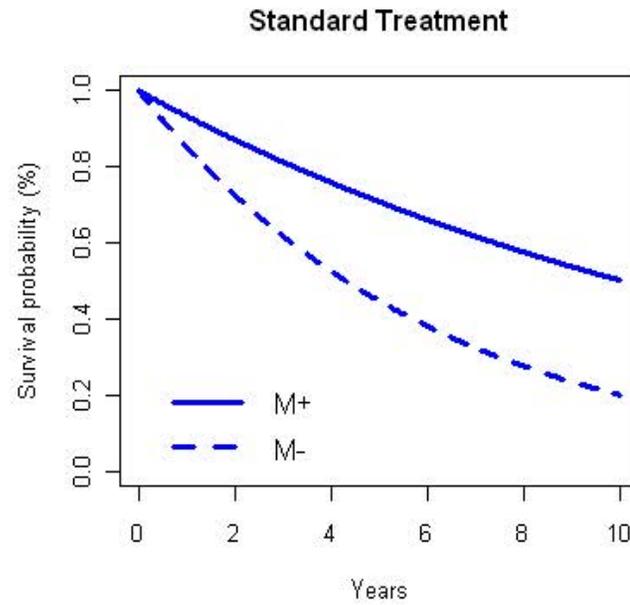
prognostic
+
predictive



(III)

prognostic

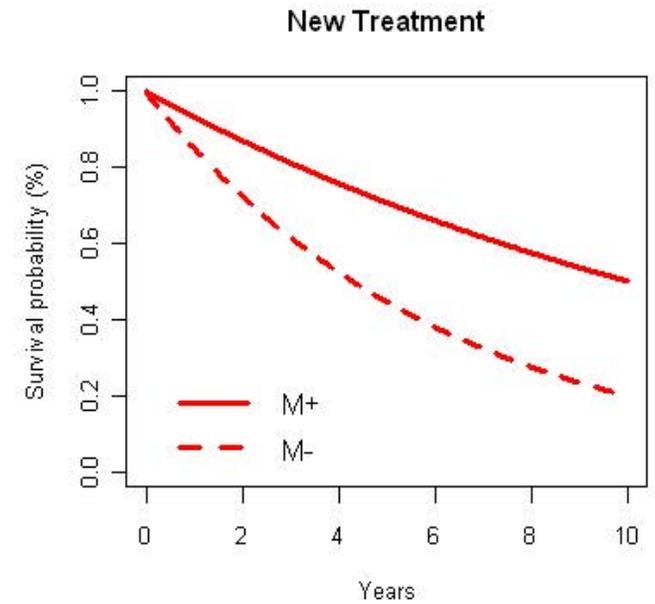
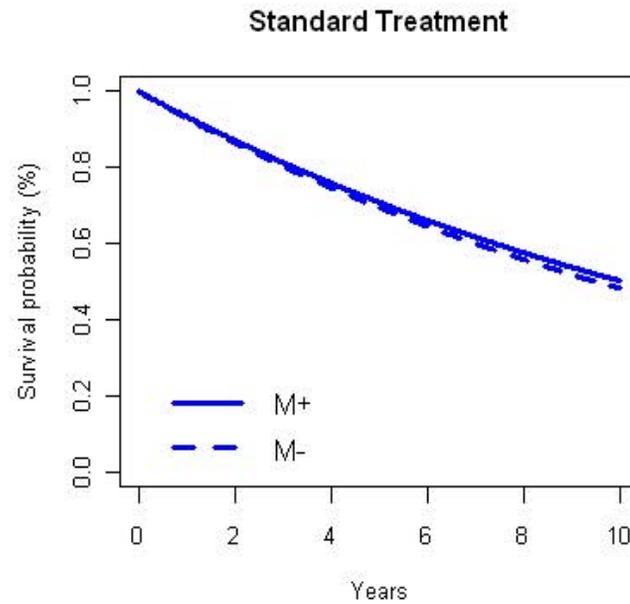
~~predictive~~



(IV)

~~prognostic~~

predictive



3. ILLUSTRATIVE EXAMPLES

Suggested reading:

- *Zhu et al. (2010) Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. JCO 28(29): 4417-4424*
- *Oh et al. (2012) Prognostic gene expression signature associated with two molecularly distinct subtypes of colorectal cancer. Gut 61: 1291-1298*
- *Kang, et al. (2012) A DNA repair pathway-focused score for prediction of outcomes in ovarian cancer treated with platinum-based chemotherapy. J Natl Cancer Inst 104(9):1-12*
- *Subramanian and Simon (2010) Gene expression-based prognostic signatures in lung cancer: Ready for clinical use? J Natl Cancer Inst 102(7):464-474*

Case Example 1

Prognostic and Predictive Gene Signature for Adjuvant Chemotherapy in Resected Non–Small-Cell Lung cancer

Chang-Qi Zhu, Keyue Ding, Dan Strumpf, Barbara A. Weir, Matthew Meyerson, Nathan Pennell, Roman K. Thomas, Katsuhiko Naoki, Christine Ladd-Acosta, Ni Liu, Melania Pintilie, Sandy Der, Lesley Seymour, Igor Jurisica, Frances A. Shepherd, and Ming-Sound Tsao

See accompanying editorial doi: [10.1200/JCO.2010.31.0144](https://doi.org/10.1200/JCO.2010.31.0144)

Background and Aim

- Benefits of adjuvant cisplatin-based chemotherapy (ACT) in patients with resected stages IB to IIIA NSCLC vary widely.
- No trial showed significant survival benefit in stage IB; a potential detrimental effect was observed in stage IA.
- The current standard of care for stage I NSCLC remains surgery alone.
- However, 30% to 40% of stage I patients will relapse.
- Hypothesis: **Some stage I patients have poorer prognosis, and may potentially benefit from ACT**

Methods: source of data

- Data from JBR.10, a randomized trial of ACT vs. Observation alone (OBS)
- Trial N = 482
 - 445 consented to banking;
 - 169 had frozen tissues;
 - 166 contained more than 20% tumor cellularity
 - 133 completed gene expression profiling (GEP)
- Out of 133: 62 in OBS, 77 received ACT

Table 1. Baseline Demographics of JBR.10 Patients With and Without Microarray Profiles

Factor	All Patients (N = 482)	Microarray Profiled (n = 133)		No Microarray (n = 349)		P
		No.	%	No.	%	
Treatment received						.14
Adjuvant chemotherapy	231	71	53	160	46	
Observation alone	251	62	47	189	54	
Age, years						.6
< 65	324	87	65	237	68	
≥ 65	158	46	35	112	32	
Sex						.35
Male	314	91	68	223	64	
Female	168	42	32	126	36	
Performance status						.72
0	236	67	50	169	49	
1	245	66	50	179	51	
Stage of disease						.01
IB	219	73	55	146	42	
II	263	60	45	203	58	
Surgery						.66
Pneumonectomy	113	33	25	80	23	
Other resection	369	100	75	269	77	
Pathologic type						.56
Adenocarcinoma	256	71	53	185	53	
Squamous	179	52	39	127	36	
Other	47	10	8	37	11	
RAS mutation status						.12*
Present	117	28	21	89	26	
Absent	333	105	79	228	65	
Unknown	32	0	0	32	9	

*P value: without including those with missing or unknown values.

Methods: signature building

- Using 65 patient data in the OBS arm, statistical model building techniques were employed to arrive at a gene signature (Cox model)
- The risk score of each patient was then derived from the Cox model based on 15-gene expressions
- Median of the risk scores was used as the cutoff point to divide patients into high- and low-risk prognostic groups
- Multivariate Cox model used to assess prognostic effect while adjusting for pre-defined baseline factors.

Methods: validation

- The prognostic signature was tested in four independent microarray datasets (stage IB to II)
 - NCI Director's Challenge Consortium (n = 96)
 - Duke (n = 48)
 - U of Michigan (n = 79)
 - Netherlands Cancer Institute (n = 133)
- Results also verified with RT-qPCR
 - All 62 samples in OBS arm
 - 30 additional samples in JBR.10 that were not profiled by microarray

Results

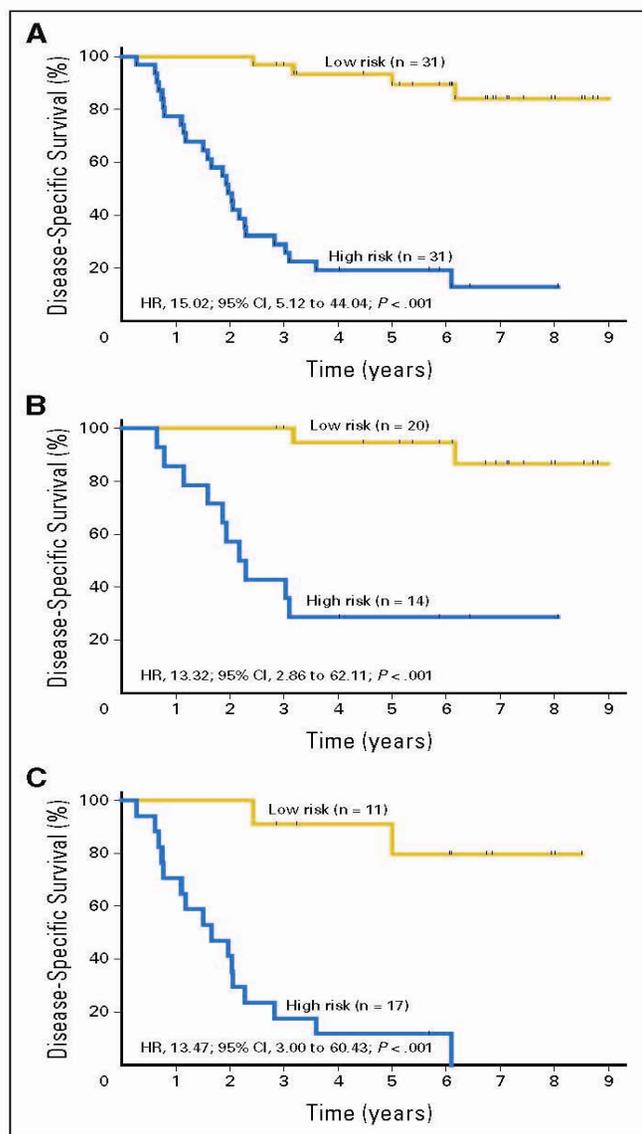


Fig 1. Disease-specific survival outcome based on the 15-gene signature in the JBR.10 training set. (A) Observation all; (B) observation stage IB; (C) observation stage II. HR, hazard ratio; ACT, adjuvant chemotherapy arm.

Table 3. Validation of the Independent Prognostic Value of the 15-Gene Signature in Four Other Separate Stage IB-II Patient Cohorts Who Received No Adjuvant Treatment

Cohort	Tumor Type	Platform	No.	Hazard Ratio*	95% CI	Adjusted P
Training set						
JBR.10	NSCLC	U133A	62	18.00	5.78 to 56.05	< .001
JBR.10	NSCLC	RT-qPCR	62	2.29	1.06 to 4.94	.034
Validation sets						
DCC	ADC	U133A	96	2.26	1.02 to 4.97	.044
NLCI	NSCLC	44K	133	2.27	1.18 to 4.35	.014
Duke	NSCLC	U133 + 2	48	1.96	0.87 to 4.42	.11
UM-SQ	SQC	U133A	79	3.57	1.48 to 8.58	.005
JBR.10†	NSCLC	RT-qPCR	19	7.65	0.85 to 69.04	.037

Abbreviations: NSCLC, non-small-cell lung cancer; U133A, Affymetrix U133A chip; RT-qPCR, quantitative reverse-transcriptase polymerase chain reaction; DCC, Director's Challenge Consortium adenocarcinoma data set; ADC, adenocarcinoma; NLCI, Netherlands Cancer Institute; 44K, Agilent 44K gene expression array; Duke, Duke University; U133 + 2, Affymetrix U133 plus2 chip; UM-SQ, University of Michigan, squamous cell carcinoma data set; SQC, squamous cell carcinoma.

*HR compares the overall survival of the high-risk (poor prognosis) patient group to that of the low-risk (good prognosis) group, after adjustment for tumor histologic subtype, stage, age, and sex.

†Values were not adjusted for clinical factors due to small sample size. The model was not adjusted for histology for UM-SQ. Since the NLCI data set did not contain information on sex this covariate was not included in the model.

Results

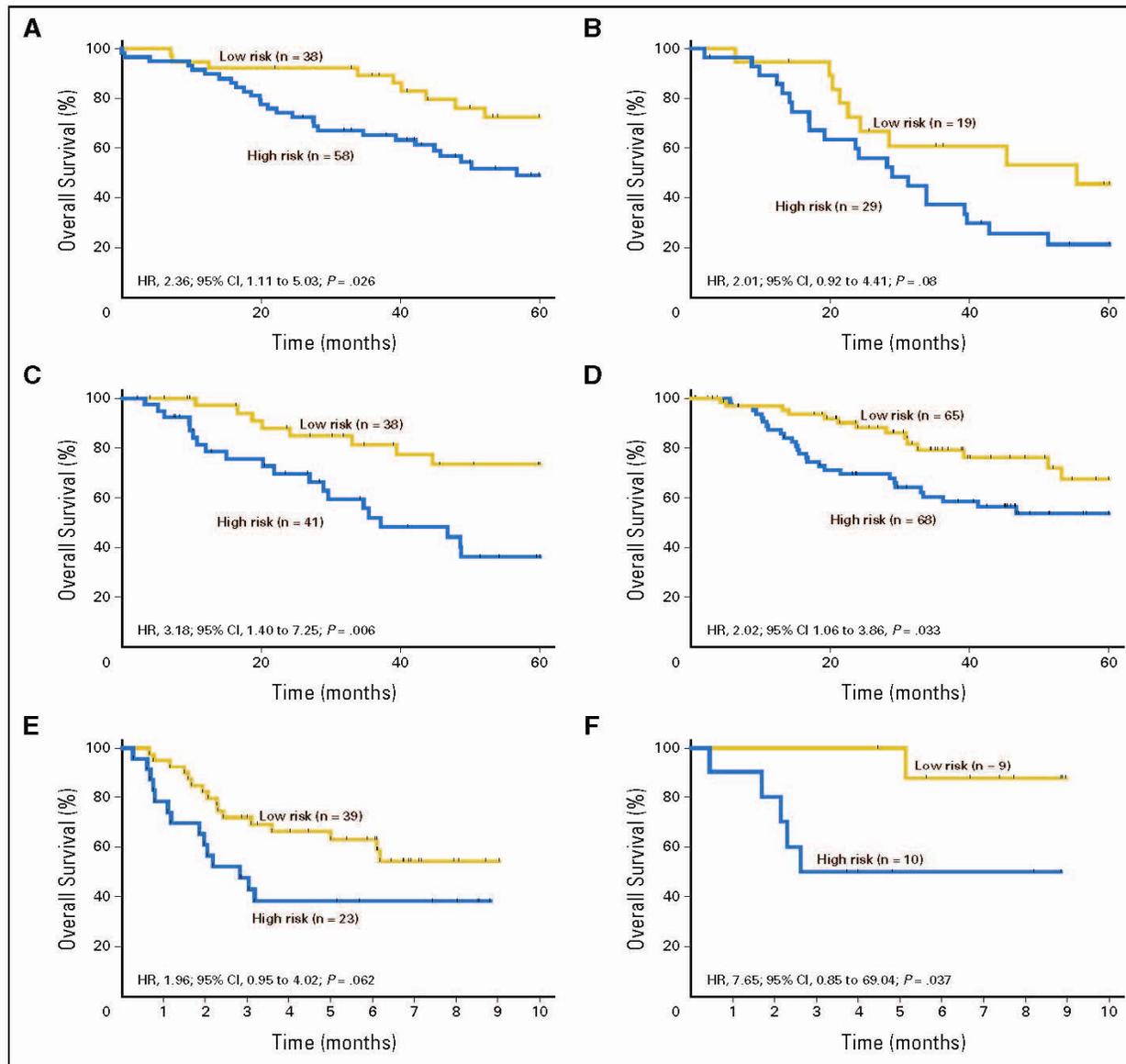


Fig 2. In silico and quantitative reverse-transcriptase polymerase chain reaction (RT-qPCR) validation of the signature in stage IB to II patients who received no adjuvant therapy. (A) Director's Challenge Consortium adenocarcinoma data set; (B) Duke University data set; (C) University of Michigan squamous cancer data set; (D) Netherlands Cancer Institute data set; (E) observation with RT-qPCR; (F) observation with RT-qPCR with additional samples. HR, unadjusted hazard ratio.

Strengths

- well-defined clinical question: Can we identify a group of high-risk stage IB/II NSCLC patients for whom adjuvant chemotherapy may be beneficial?
- appropriate selection of patient population for model building (all patients had complete surgical resection and none received adjuvant chemotherapy)
- multiple independent external validation cohorts were included

Limitations

- The “resubstitution statistics” were reported which were known to yield severely biased results (Figure 1).
- Did NOT successfully demonstrate the prognostic value of the gene signature above and beyond standard clinical variables (multivariable Cox regression models are suitable for assessing associations but inadequate for assessing predictive accuracy)

Limitations (continued)

- Endpoint: The genetic signature was built on the training set using DFS as the primary clinical outcome. However, the clinical endpoint became OS when validating the signature in external datasets.

Limitations (continued)

- Clinical Utility? : Figure 2A shows that the 5-year OS was roughly 75% and 55% for low risk and high risk patients, respectively, in the NCI validation dataset. The difference was even smaller for the Netherlands Cancer Institute dataset: 65% vs. 55% for low and high subgroups, respectively (Figure 2D). Was the prognostic effect of the signature sufficiently profound in these data to support a clinical decision to treat early stage NSCLC patients with ACT?

Case Example 2

**Prognostic gene expression signature
associated with two molecularly distinct
subtypes of colorectal cancer.**

Sang Cheul Oh, Yun-Yong Park, Eun Sung Park, Jae Yun
Lim, Soo Mi Kim, Sang-Bae Kim, Jongseung Kim, Sang
Cheol Kim, In-Sun Chu, J Joshua Smith, R Daniel
Beauchamp, Timothy J Yeatman, Scott Kopetz, Ju-Seog
Lee. *Gut* 2012;61:1291-1298

Challenges

- Poorer survival of CRC patients with stage II (T stage 4, lymph node-negative) than with stage IIIA (T stage 1-2, lymph node-positive).
- There are no clinically useful biomarkers that can reliably predict the prognosis and response to adjuvant chemotherapy in stage II and stage III CRC.
- KRAS mutations represent first biomarker integrated into clinical practice for CRC with negative predictive value for anti EGFR antibody treatment.
- MSI is considered to be a robust prognostic marker in stage II but not stage III CRC and marker of response in the adjuvant setting in MSI-L/MSS but not MSI-H stage II.
- 18q LOH in combination with MSI-H was prognostic in stage II but not stage III.
- BRAF is associated with poor prognosis and accounts for resistance to Cetuximab.

Prognostic tests based on GEP in CRC

- Kerr (JCO 2009, 27, 4000 and 2011,35, 4498) 12 gene prognostic signature that is stage independent in stage II CRC (n=1,851, NSABP C-01/C-04/C-06 and Cleveland Clinic.) RT-PCR based test, available for FFPE tumor tissue, [Oncotype Dx](#) colon cancer, Genomic Health, Inc.).
- Salazar (JCO 2011, 29, 17) 18-gene profile ([Coloprint](#), Agendia) from whole genome oligonucleotide array (Agilent 44K) for low and high risk of recurrence groups in retrospective cohort from fresh-frozen stage II and III CRC (n=110).
- Agesen et al. (Gut, 2012, 61, 1560) [ColoGuideEx](#) classifier based on 13 genes differentially expressed between stage I and IV CRC patients fresh frozen samples (n=207) (Human exon 1.0 ST Affymetrix) for prognosis in stage II CRC but not stage III CRC patients.

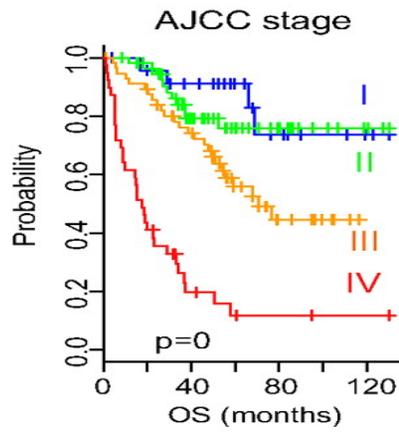
Aims & Methods

- To establish a prediction model to establish risk of recurrence and help guide treatment strategies.
- Unsupervised hierarchical clustering analysis was applied to already available gene expression data from specimens from stage I-IV CRC (Moffitt CC n= 177) (Affymetrix U133 version 2 and Affymetrix U133 platforms).
- The association between the signature and prognosis of patients was assessed by Kaplan-Meier plots, log-rank tests and the Cox model using two data cohorts (Vanderbilt Univ. /Max Planck Inst., n= 117 and Royal Melbourne Hospital, n=96).

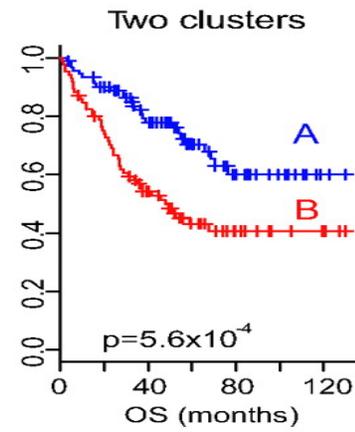
Results

- Two subgroups of patients (A and B) were identified using a gene signature (114 genes) that was associated with overall survival and disease-free survival.
- The signature was independent of the current staging system suggesting prognostic value superior to conventional risk factors.
- The gene signature was an independent predictor of response to chemotherapy and clinical outcome in subtype B but was not significant for patients in subtype A in stage III patients.

A

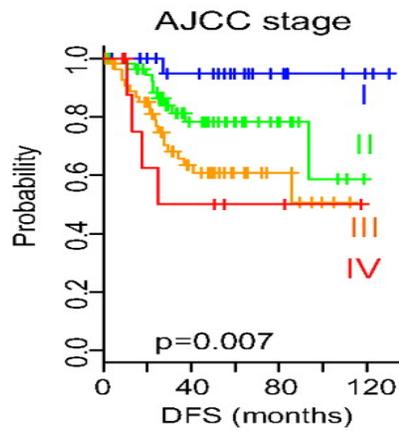


No. at risk			
Stage I	24	18	7
Stage II	57	31	13
Stage III	57	38	8
Stage IV	39	6	2

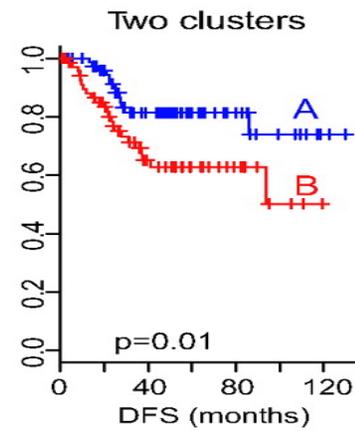


No. at risk				
A	91	54	20	3
B	86	39	10	3

B



No. at risk			
Stage I	24	17	7
Stage II	55	24	8
Stage III	56	24	7
Stage IV	10	4	2



No. at risk				
A	75	41	15	
B	70	28	9	

Strengths

- Clinically significant question (to identify subtypes of colorectal cancer that have distinct biological characteristics associated with prognosis and treatment response).
- Use of two independent validation cohorts.
- Identified genes associated with more aggressive subtype with poorer outcome (group B), *e.g.*, TGFb pathway associated with metastasis.
- Overlap with genes in the Oncotype DX test (FAP, INHBA and BGN)

Weaknesses:

- Overall, poorly defined study population based on specimens from heterogeneous patients cohorts (stage I-IV) and retrospective analysis (convenience samples?)
- Data generated using different Affymetrix platforms.
- No information was provided how the raw data were normalized and whether the expression levels were comparable across different platforms.
- Not clear how the algorithm for classification was developed.
- Strong confounding factors such as T stage, MSI and number of examined lymph nodes were not included in the analysis.
- What would be the clinical application of this test for early stage CRC? Stage and conventional risk factors outweigh the prognostic significance of this algorithm.
- What is the clinical use of the test for stage IV patients?

Case Example 3

4. DISCUSSION

Going forward

- What criteria need to be satisfied to establish clinical utility?
- How do we obtain reliable evidence?
- What can NCI do to help?

A. Definition of clinical utility of prognostic assays

- Does clinical utility of a prognostic assay apply only in the setting of a difficult clinical decision regarding treatment/management?

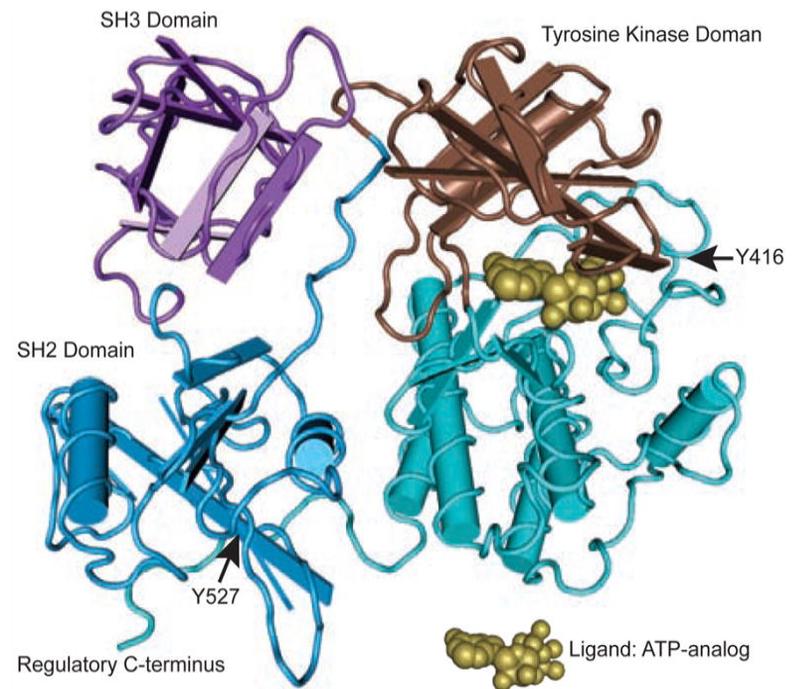
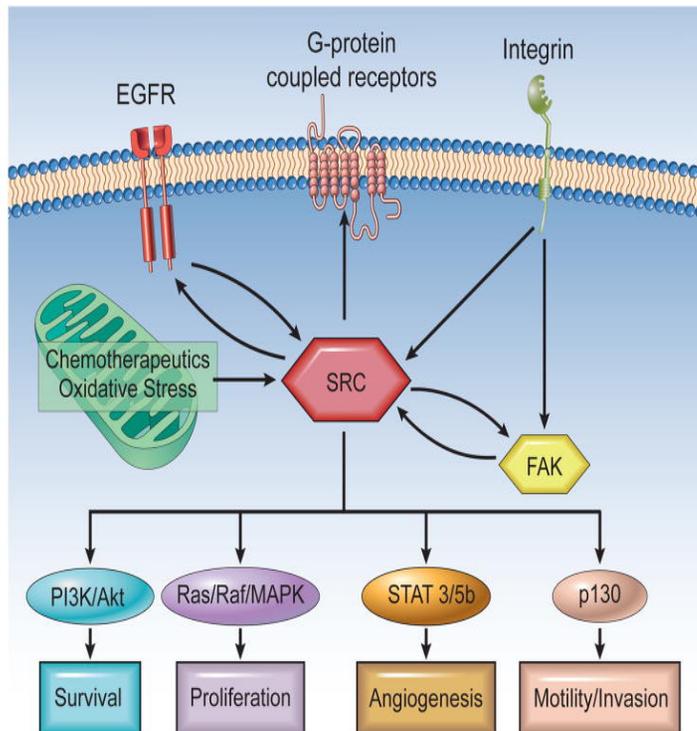
A. Definition of clinical utility of prognostic assays

- Must the assay drive a treatment/management decision?

A. Definition of clinical utility of prognostic assays

- Is it necessary to understand the biology behind the clinical association?

Targeted therapies can be designed for genes upregulated in CRC, e.g., SRC kinase blocking agents are in clinical development including including dasatinib (BMS-354825), saracatinib (AZD0530), bosutinib (SKI-606), KX2-391, and XL228



A. Definition of clinical utility of prognostic assays

- Must clinical utility always relate to improved clinical outcome?

A. Definition of clinical utility of prognostic assays

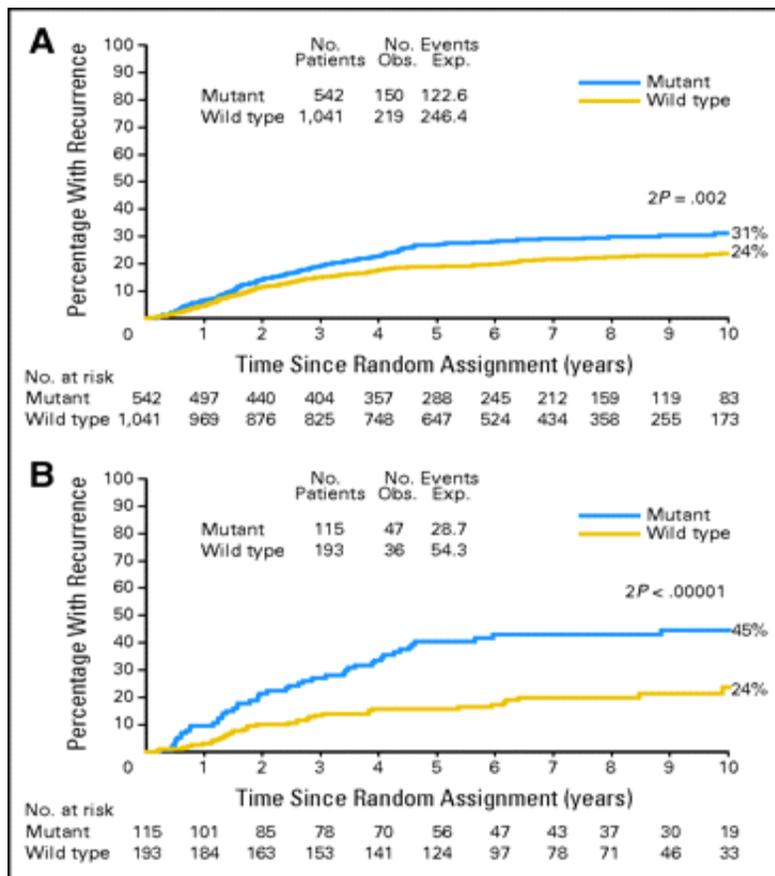
- What are other factors in play and how should we balance among them?
Efficacy, alternative treatment options, toxicity, cost of drug and cost of assay?

B. Evidence generation

- Does the assay work with the specimens of choice?

B. Evidence generation

- Does the assay clearly define the population of interest?
Does evidence always have to be generated in the same organ or context it will be used?



Recurrence by *KRAS* status in QUASAR trial: Increased risk of recurrence is observed in *KRAS* mutant rectal but not colon tumors (A) all patients, (B) rectum stage II only. Obs., observed number of recurrences

B. Evidence generation

- What are clinically relevant endpoints? Does the choice depend on disease?

B. Evidence generation

- Is it sufficient to identify different risk groups? How large does a difference in outcome need to be?

B. Evidence generation

- Evidence-based approach:
what are common statistical
issues?

Gene Expression–Based Prognostic Signatures in Lung Cancer: Ready for Clinical Use?

Jyothi Subramanian, Richard Simon

Manuscript received July 9, 2009; revised December 29, 2009; accepted January 15, 2010.

Correspondence to: Richard Simon, DSc, Biometric Research Branch, Department of Cancer Treatment and Diagnosis, National Cancer Institute, 9000 Rockville Pike, Bethesda, MD 20892-7434 (e-mail: rsimon@mail.nih.gov).

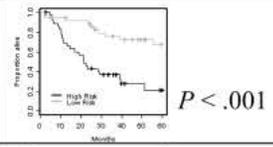
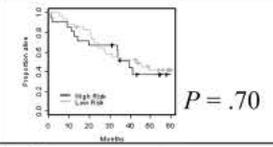
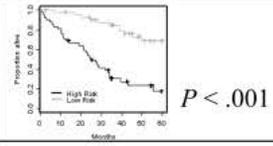
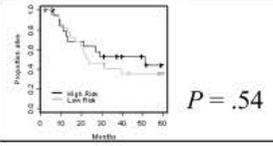
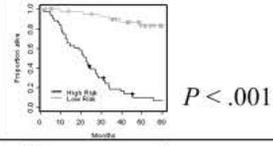
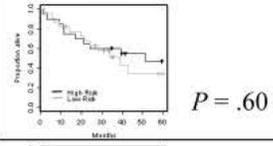
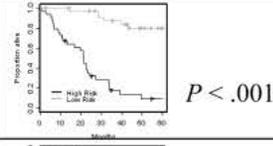
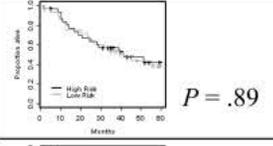
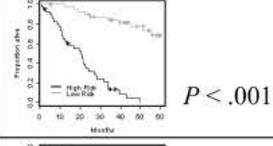
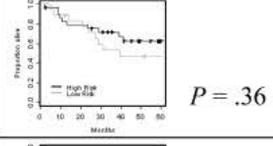
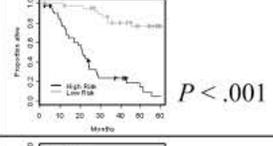
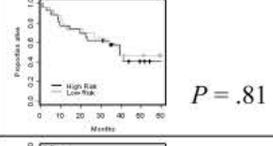
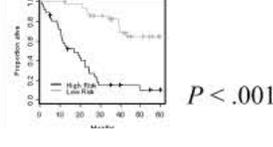
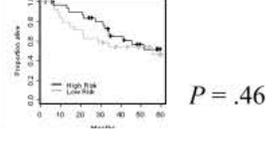
A substantial number of studies have reported the development of gene expression–based prognostic signatures for lung cancer. The ultimate aim of such studies should be the development of well-validated clinically useful prognostic signatures that improve therapeutic decision making beyond current practice standards. We critically reviewed published studies reporting the development of gene expression–based prognostic signatures for non–small cell lung cancer to assess the progress made toward this objective. Studies published between January 1, 2002, and February 28, 2009, were identified through a PubMed search. Following hand-screening of abstracts of the identified articles, 16 were selected as relevant. Those publications were evaluated in detail for appropriateness of the study design, statistical validation of the prognostic signature on independent datasets, presentation of results in an unbiased manner, and demonstration of medical utility for the new signature beyond that obtained using existing treatment guidelines. Based on this review, we found little evidence that any of the reported gene expression signatures are ready for clinical application. We also found serious problems in the design and analysis of many of the studies. We suggest a set of guidelines to aid the design, analysis, and evaluation of prognostic signature studies. These guidelines emphasize the importance of focused study planning to address specific medically important questions and the use of unbiased analysis methods to evaluate whether the resulting signatures provide evidence of medical utility beyond standard of care–based prognostic factors.

J Natl Cancer Inst 2010;102:464–474

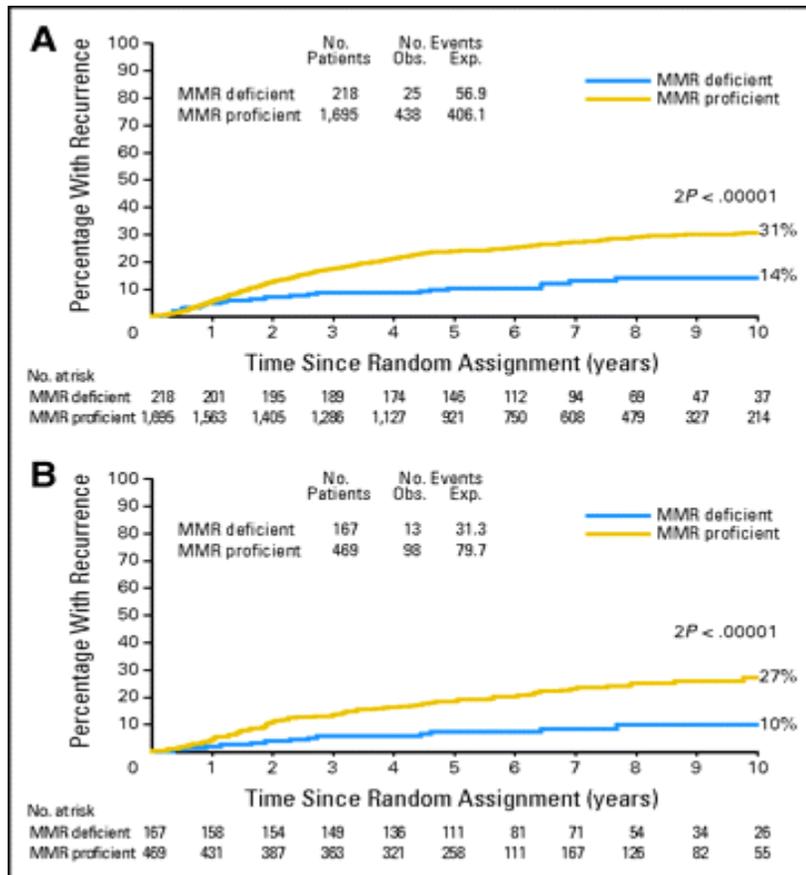
- Patient selection should be based on intended use of the prognostic signature.
- The signature must be validated on at least one independent dataset.
- Resubstitution statistics for the training should NOT be reported.
- The new signature should show better performance than other standard variables (regression model not sufficient).

Resubstitution Statistics

Figure 2. Kaplan–Meier survival estimates for the simulation study. *Prediction accuracy for the training and validation datasets with random gene expression profiles. For this simulation, survival data on 129 patients were obtained from Bild et al. (32). For each patient, 5000 random numbers obtained from the standard normal distribution formed the gene expression profile. This master dataset was divided randomly into training and validation sets. A model predicting survival based on gene expression was developed from the training data. This model was used to classify survival risk group for patients in the training set and the validation set. The Kaplan–Meier curves show the proportion alive (vertical axis) vs time in months (horizontal axis) for predicted high-risk (black line) and low-risk (gray line) groups. Tick marks indicate censored observations. The P values are two-sided and are from the log-rank test. The Kaplan–Meier survival curves for the training set are “resubstitution” estimates because the same data are used to develop the model and to test it. Additional details of the simulation methodology are provided in the Supplementary Methods (available

Simulation	Training set	Validation set
1	 $P < .001$	 $P = .70$
2	 $P < .001$	 $P = .54$
3	 $P < .001$	 $P = .60$
4	 $P < .001$	 $P = .89$
5	 $P < .001$	 $P = .36$
6	 $P < .001$	 $P = .81$
7	 $P < .001$	 $P = .46$

Quick and Simple and Reliable (QUASAR) trial, n=1,913).
 Recurrence by mismatch repair (*MMR*) status: (A) all patients, (B)
 colon stage II only.



IHC test is suggested for routine clinical practice. MSI-H is positive prognostic but negative predictive factor and MSI-H patients should not offered chemotherapy .

C. Source of data

- Randomized controlled trials are the standard? Why or why not? Are there any cases in which nothing less can be considered?

C. Source of data

- What are other options?
When are they appropriate
and when are they not?

C. Source of data

- How do low-prevalence markers and small sample sizes affect the ability to evaluate clinical utility?

D. Role of NCI

- Does the NCI have a unique role? Can the NCI help fund the development of promising assays? Can the NCI help facilitate interactions with the FDA?

D. Role of NCI

- Could the clinical trials network be utilized differently to generate the evidence? What are the hurdles?