2011 EORTC-NCI-ASCO Annual Meeting
Abstract #54

A genomic-based signature of response to chemotherapy in ovarian cancer fails to predict clinical outcome in two independent cohorts.

William T Barry, Wei Chen, Gregory Sfakianos, Pamela Isner, Michael Datto, Nicole Kuderer, Gary Lyman, Amy Abernethy, Geoff Ginsburg, Andrew Berchuck

Duke University School of Medicine
Duke Institute for Genome Science and Policy
Durham, NC USA

BACKGROUND

In Dressman (2007) a gene expression signature was developed for response to platinum–based therapy in advanced-stage ovarian cancer.

119 samples split into:

- ❑ 83 training samples
- ❑ 36 test samples

Clinical endpoint: clinical resp.

- ❑ WHO criteria in meas. dis.
- ❑ CA-125 in nonmeas. dis.
- ❑ 71.4% complete CR

<u>Goal</u> To evaluate the prognostic value of the multiplex marker, two ovarian cohorts were identified as external datasets for retrospective validation.

VOLUME 25 · NUMBER 5 · FEBRUARY 10 2007

JOURNAL OF CLINICAL ONCOLOGY | ORIGINAL REPORT

An Integrated Genomic-Based Approach to Individualized Treatment of Patients With Advanced-Stage Ovarian Cancer

Holly K. Dressman, Andrew Berchuck, Gina Chan, Jun Zhai, Andrea Bild, Robyn Sayer, Janiel Cragun, Jennifer Clarke, Regina S. Whitaker, LiHua Li, Jonathan Gray, Jeffrey Marks, Geoffrey S. Ginsburg, Anil Potti, Mike West, Joseph R. Nevins, and Johnathan M. Lancaster

Baggerly KA, Coombes KR, Neeley ES. J Clin Oncol. 2008; 26:1186-1187
Dressman HK, Potti A, Nevins JR, Lancaster JM. J Clin Oncol. 2008; 26:1187-1188

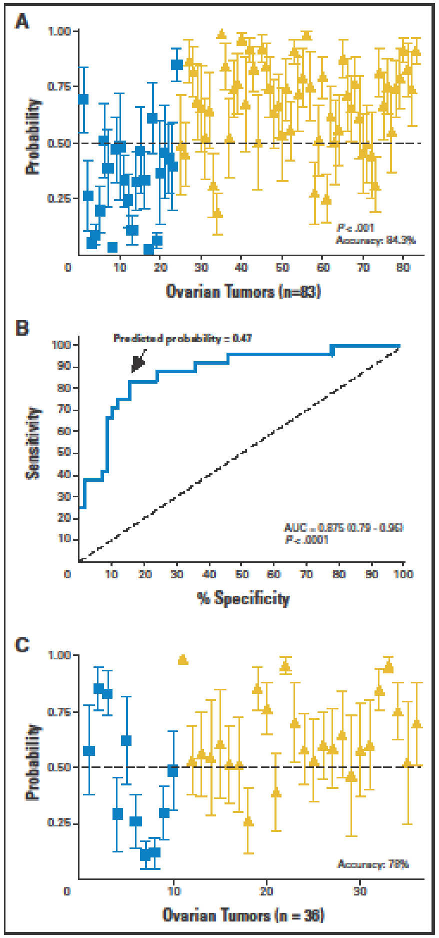


Fig 1. Gene expression pattern associated with platinum response. (A) Leave-one-out cross validation of training set (blue = incomplete responders, yellow = responders). (B) Receiver operating characteristic (ROC) curve of the training set. (C) Validation of the platinum response prediction based on a cutoff of 0.47 predicted probability of response as determined by ROC curve. AUC, area under the curve.

BACKGROUND

– In developing the Dressman (2007) signature:

<u>Model:</u> Shotgun Stochastic Search (SSS)
Hans, Dobra, and West, JASA 2007

"A Bayesian framework for regression with uncertainty about which predictors are in the model; uncertainty is represented in terms of a prior variable inclusion probability to penalize model dimension."

$$p(\gamma_1), \quad g_1(\boldsymbol{y}\,|\gamma_1) = \hat{\beta}_{1,0} + \hat{\beta}_{1,1} \cdot gene_{154} + \hat{\beta}_{1,2} \cdot gene_{827} \quad \hat{V}_1(\hat{\beta}_1)$$
$$p(\gamma_2), \quad g_2(\boldsymbol{y}\,|\gamma_2) = \hat{\beta}_{2,0} + \hat{\beta}_{2,1} \cdot gene_{398} \quad \hat{V}_2(\hat{\beta}_2)$$
$$p(\gamma_3), \quad g_3(\boldsymbol{y}\,|\gamma_3) = \hat{\beta}_{3,0} + \hat{\beta}_{3,1} \cdot gene_{154} + \hat{\beta}_{3,2} \cdot gene_{21} \quad \hat{V}_3(\hat{\beta}_3)$$
$$\ldots. \qquad \ldots.$$

– In validating the signature: Probabilities (scores) are obtained as posterior modes from the Bayesian model.

BACKGROUND

- Clinical data and batch-corrected gene expression provided as supplemental material (updated in 2009)

  http://data.genome.duke.edu/platinum.php

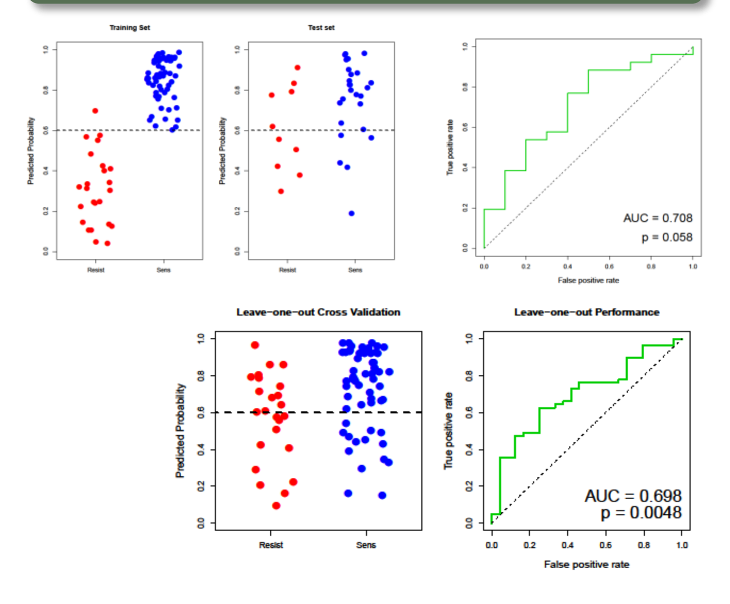
Table 1: Top 10 of 1704 Genes

Probe ID	Inclusion Prob.	Gene
201929_s_at	0.865	PKP4
201484_at	0.790	SUPT4H1
214352_s_at	0.176	KRAS
214843_s_at	0.157	USP33
221971_x_at	0.114	AGAP7
218053_at	0.114	PRPF40A
219206_x_at	0.106	TMBIM4
202172_at	0.071	VEZF1
212277_at	0.066	MTMR4
201928_at	0.058	PKP4

Split sample validation:
AUC = 0.708,
72.2% accuracy

Leave-one-out cross validation:
AUC = 0.698
67.5% accuracy

Plot: Evaluation of internal accuracy



<u>Posterior importance:</u> The relative influence of each gene.

Genelist closely matches results in supp. material (updated)

NEW VALIDATION COHORTS

Two ovarian datasets were identified for validation:

- Duke (2010): 47 banked specimen selected as complete vs. incomplete response to primary platinum-based therapy.
- TCGA (2011): Subset of serous tumors (n = 402) that received chemo-therapy and with clinical response known.

	Dressman (2007)		Duke (2010)		TCGA (2011)[1]	
	Incomplete Responder	Complete Responder	Incomplete Responder	Complete Responder	Incomplete Responder	Complete Responder
N	34	85	17	30	117	285
Age, mean	65	63	57	60	59	59
Stage						
I	0	0	0	1	1	7
II	0	0	0	1	0	21
III	27	72	16	23	92	216
IV	7	13	1	5	22	41
Grade						
1	1	2	0	0	1	3
2	15	42	11	13	17	33
3	18	41	6	17	98	249
Surgical Debulk.						
Optimal	12	51	5	16	-NA-	-NA-
Subopt.	22	34	11	14	-NA-	-NA-
Sensitivity						
Resistant	-NA-	-NA-	17	4	42	37
Sensitive	-NA-	-NA-	0	26	21	163
Median OS (yrs)	2.6	3.7	3.6	6.1	2.7	4.8

1. Raw Affymetrix data and clinical information downloaded from the TCGA portal (http://tcga-data.nci.nih.gov/tcga/) on 4/20/2011 Data were cross-checked with the supplemental material subsequently provided to the Nature publication.

METHODS – PREPROCESSING

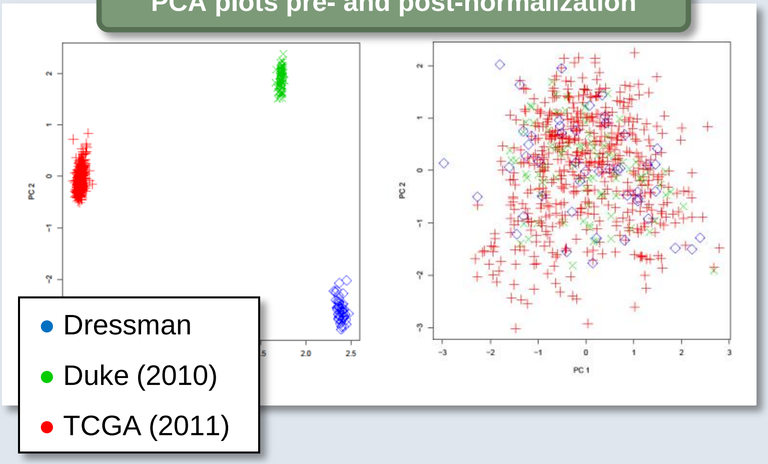
Due to differences in methods, labs, and protocols (shown in Table 2), any integrated analysis of the gene expression will require batch normalization.

Table 2: Technical factors

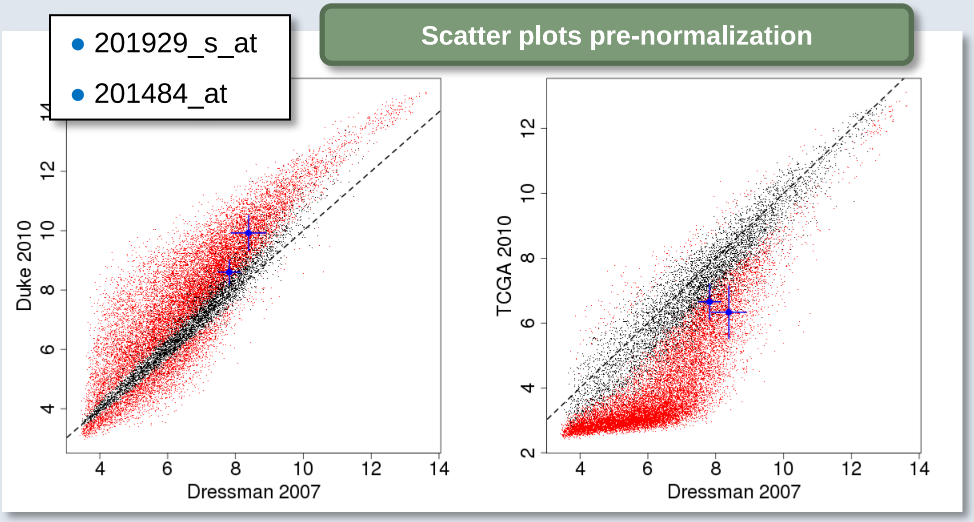
	Dressman (2007)	Duke (2010)	TCGA (2011)
Laboratory	Duke MaF	Duke MaF	TCGA GCC
Labeling Kit	Affy 1 cycle	Ambion Message Amp	HT IVT
Platform	U133A	U133Plus2	HT-U133A

Similar estimates and results were obtained using other methods: (1) Combat, (2) DWD

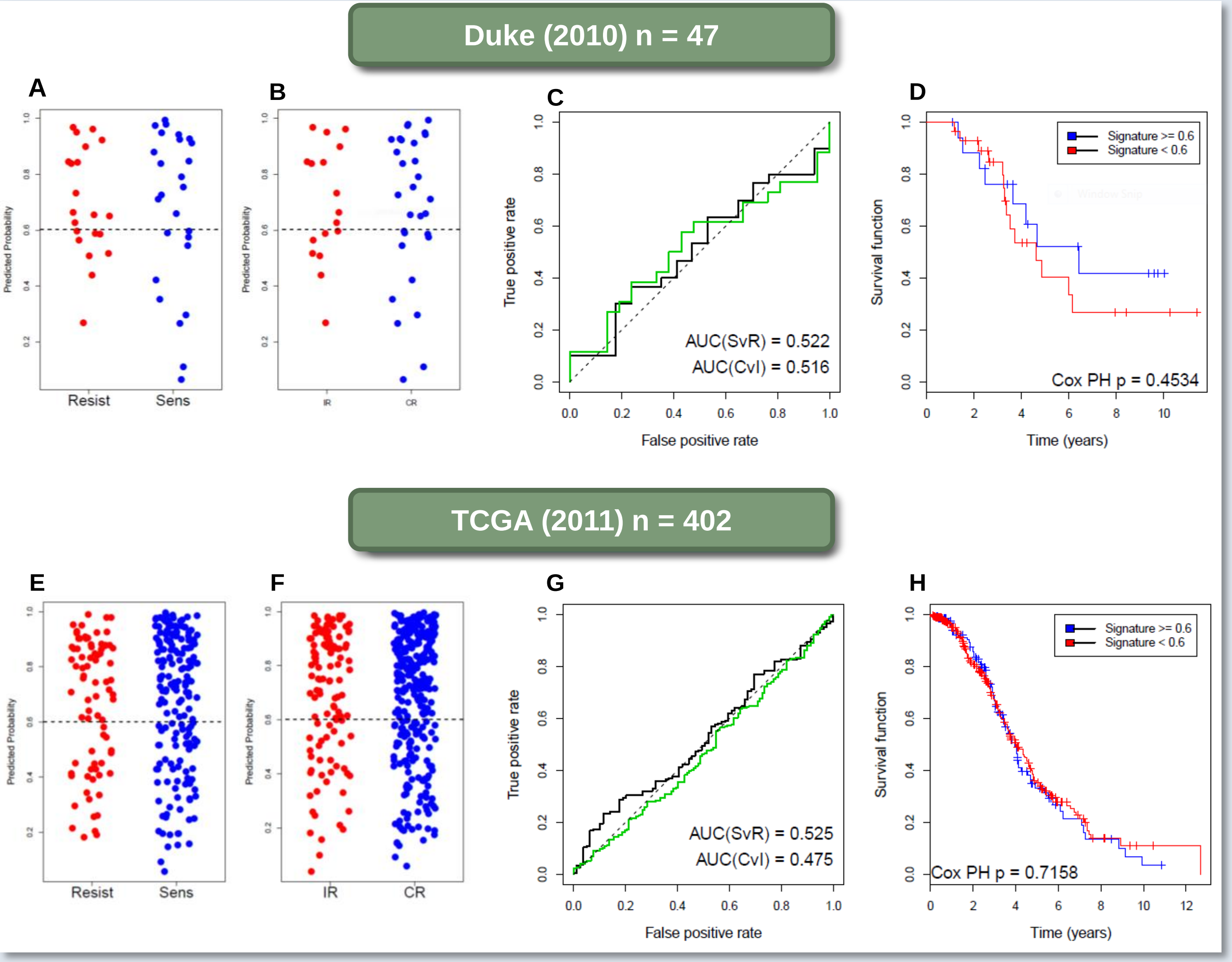
PCA plots pre- and post-normalization



Scatter plots pre-normalization



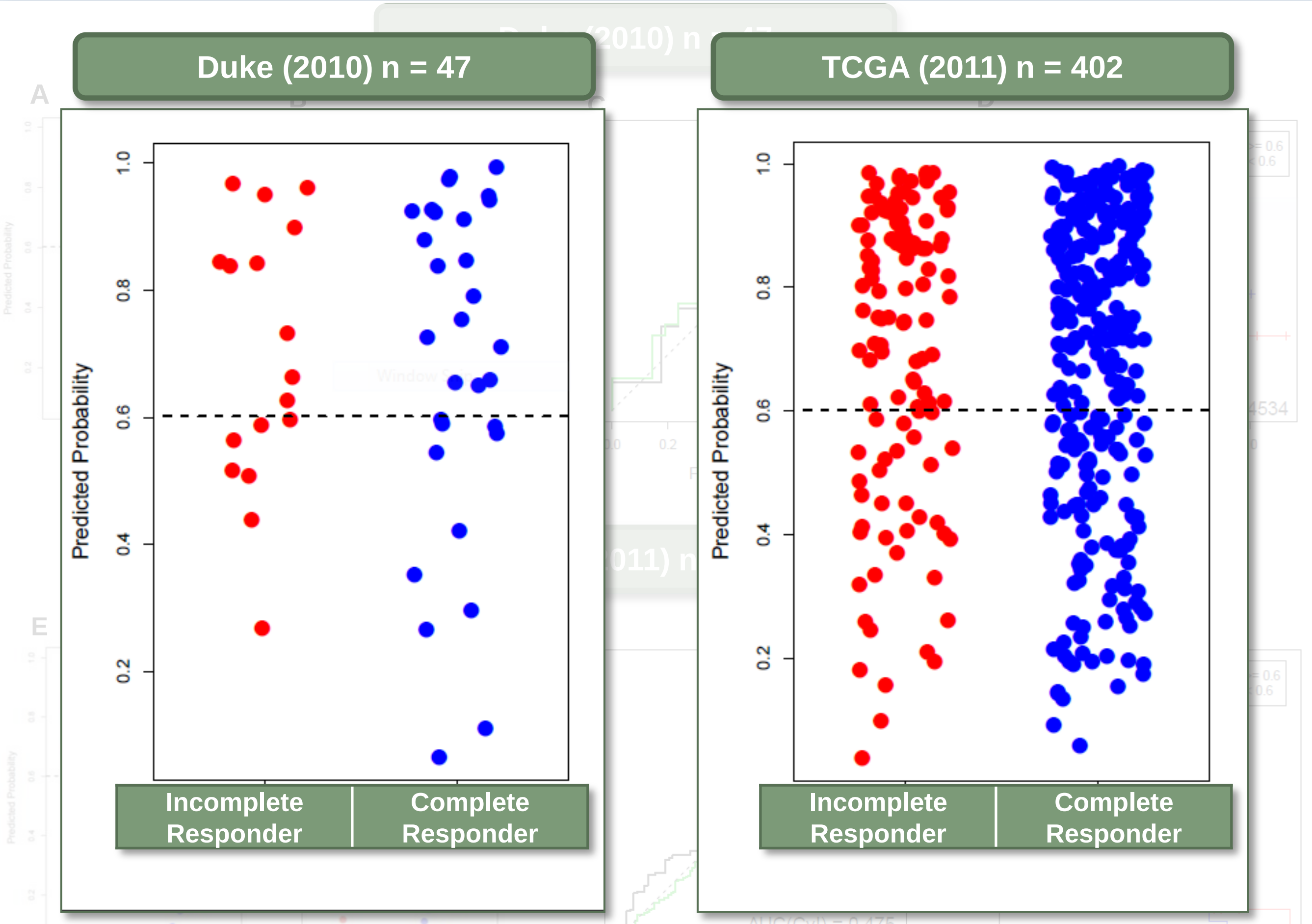
PERFORMANCE



Probabilities of a complete response from the Bayesian genomic model are plotted against:

- (Panels B and F) Complete clinical response,
- (Panels A and E) Sensitive versus resistant tumors.
- (Panels D and H) Overall survival (binary cutoff of P ≥ 0.6)

PERFORMANCE



Wilcoxon p = 0.43

Wilcoxon p > 0.5

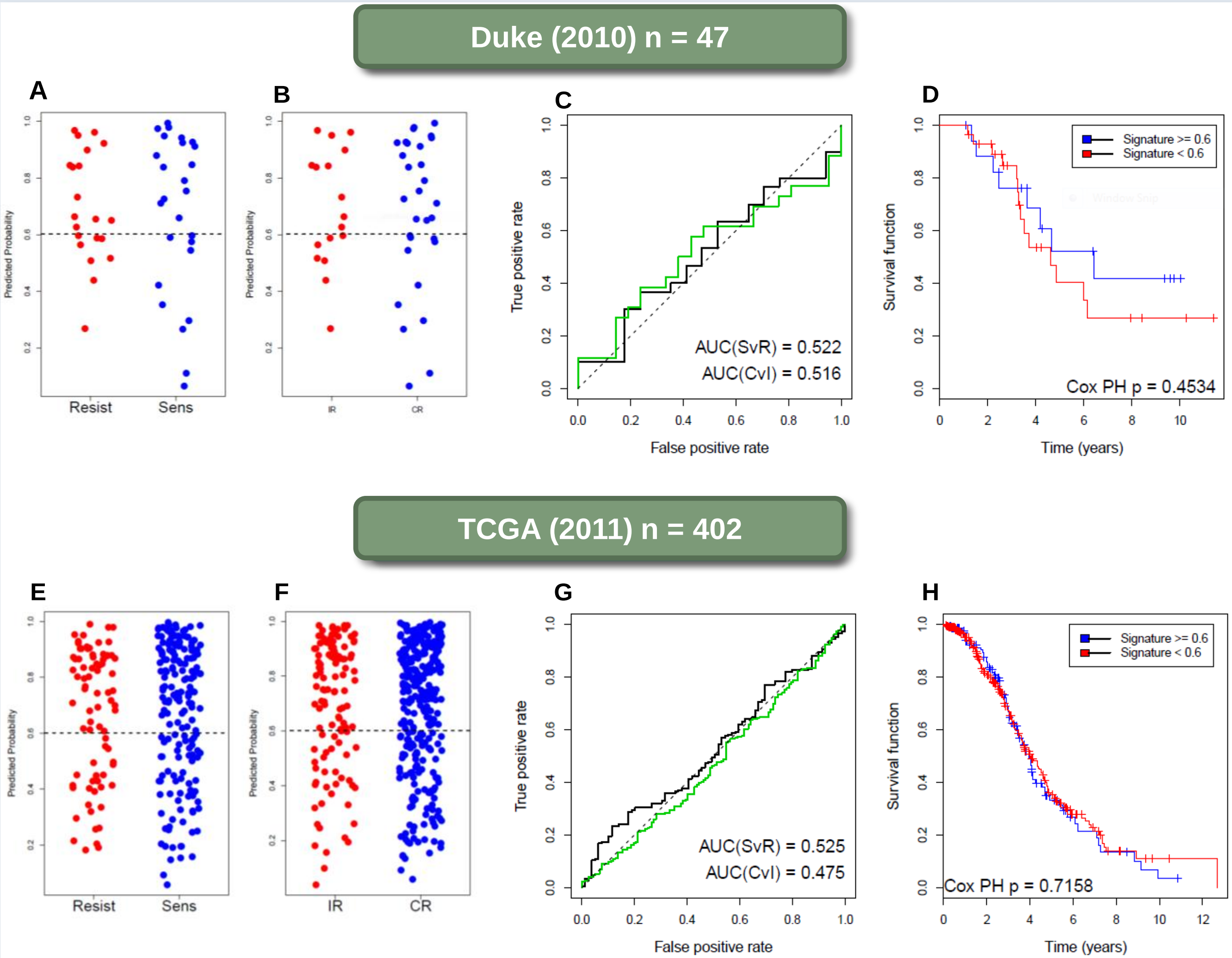
Probabilities of a complete response from the Bayesian genomic model are plotted against:

- **(Panels B and F) Complete clinical response**
- (Panels A and E) Sensitive versus resistant tumors.
- (Panels D and H) Overall survival (binary cutoff of P ≥ 0.6)

DukeMedicine

DUKE CANCER INSTITUTE
A National Cancer Institute-designated Comprehensive Cancer Center

PERFORMANCE



Probabilities of a complete response from the Bayesian genomic model are plotted against:

- (Panels B and F) Complete clinical response,
- (Panels A and E) Sensitive versus resistant tumors.
- (Panels D and H) Overall survival (binary cutoff of P ≥ 0.6)

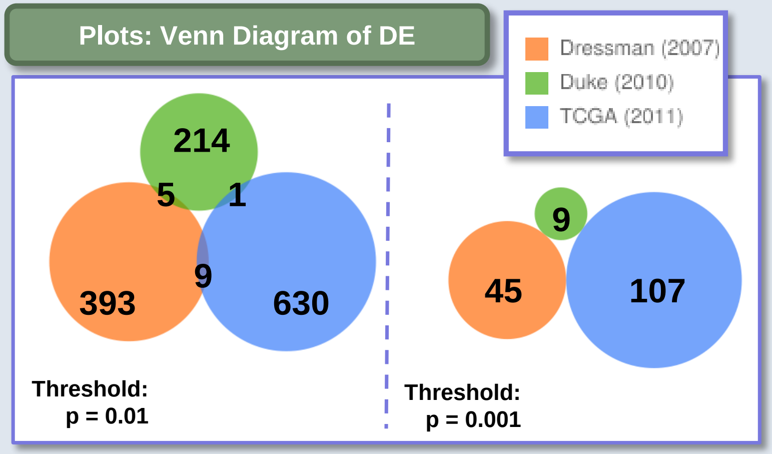
PERFORMANCE

To understand the lack of prognostic value in the two cohorts:

(1) Ranked gene-lists of differential expression between complete and incomplete responders.

Using nominal thresholds of p = 0.01 and p = 0.001, almost no overlap is seen.

Plots: Venn Diagram of DE



ID	T-stat	P.Value	adj.P.Va	Gene	Desc
DRESSMAN (2007)					
201929_s_at	4.80	4.58E-06	0.0744	PKP4	plakophilin 4
217608_at	-4.71	6.73E-06	0.0744	SFRS12IP1	SFRS12-interacting protein 1
TCGA (2011)					
216026_s_at	5.00	8.70E-07	0.0194	POLE	polymerase (DNA directed), epsilon
206267_s_at	-4.70	3.54E-06	0.0395	MATK	megakaryocyte-associated tyrosine kinase
205422_s_at	-4.42	1.27E-05	0.0946	ITGBL1	integrin, beta-like 1 (with EGF-like repeat domains)
DUKE (2010)					
244506_at	-4.24	1.02E-04	0.9995	TMTC1	transmembrane and tetratricopeptide repeat
203379_at	-4.09	1.65E-04	0.9995	RPS6KA1	ribosomal protein S6 kinase, 90kDa, polypeptide 1

<u>After correction for multiple testing</u>

- 2 genes in the Dressman 2007 set (PKP4 and SFRS12IP1),
- 3 genes in the TCGA data (POLE, MATK, and ITGBL1) had a FDR<10%.
- No genes in Duke (2010) were significant.

METHODS – SAFE ANALYSIS FOR SIGNATURES

GOAL: To evaluate whether signatures show concordant patterns of differential expression in external datasets

SAFE framework: Methods of pathway-analysis have been proposed that are unbiased by accounting for gene-gene correlation: GSEA (Mootha, 2003), SAFE (Barry, 2005), and GSA (Efron, 2006).

We propose two extensions :

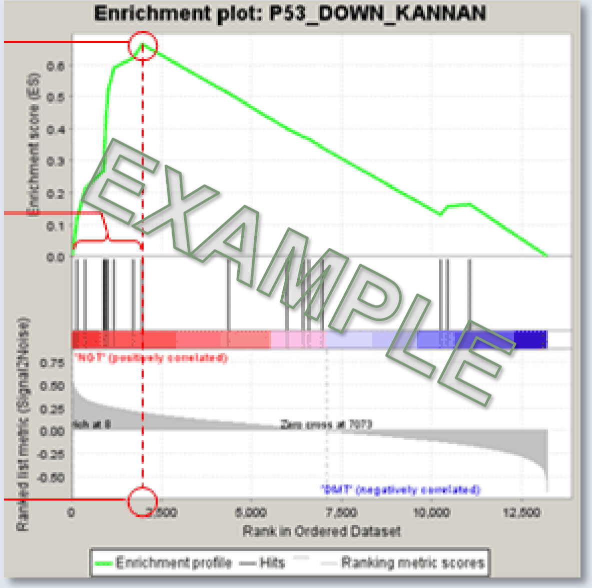
1. Define membership on the [0,1] scale. Apply a Wilcoxon linear score statistic (Hajek and Sigdak, 1967).
2. Gene-specific tests (T) are one-sided in the direction of differential expression seen in the training set.

$$U = \sum_{i=1}^{m} p_i \cdot Rank\big(T_i \cdot Sign(t_{obs,i})\big)$$

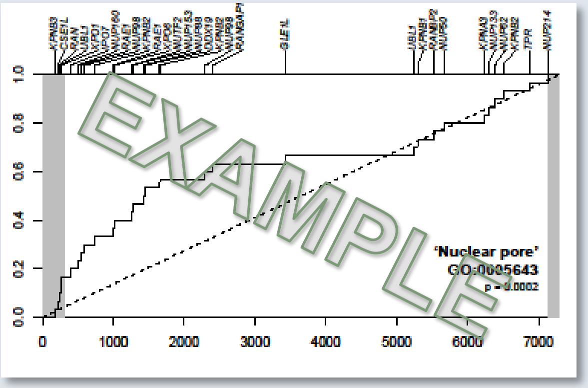
p = probability of membership, t_{obs} = the observed value from the discovery process for m genes.

$$E_{H_0}[U] = \frac{m_c \cdot (m+1)}{2}, \qquad m_c = \sum_{i=1}^{m} p_i$$

Resampling is used to (1) induce a global null under permutation, or (2) estimate parameters (95th percentiles) for bootstrap-based tests.

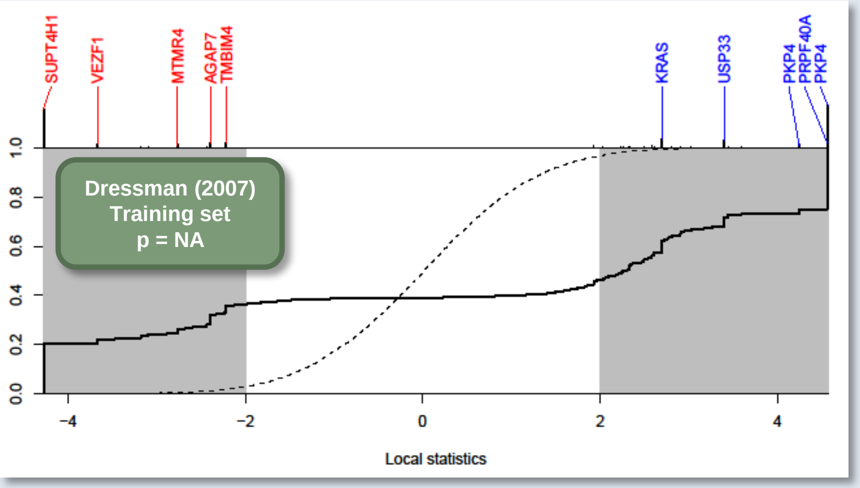


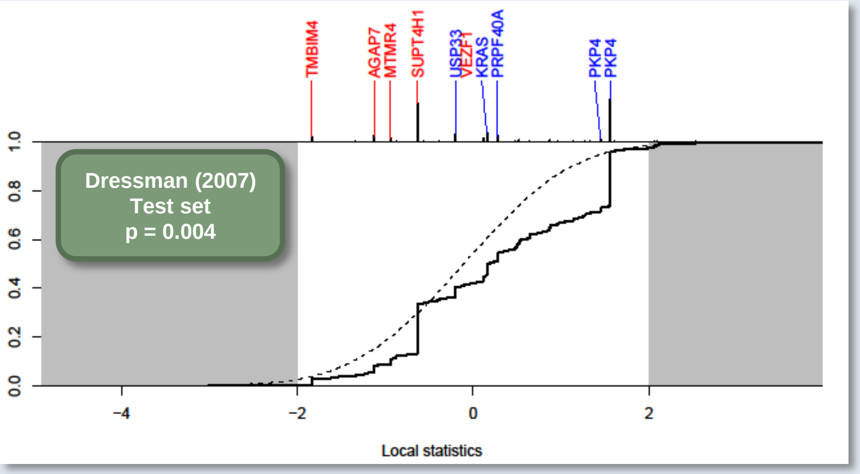
Plots of pathway DE: GSEA (↑) and SAFE (↓)



SAFE ANALYSIS OF THE OVARIAN SIGNATURE

- Weights are defined by the inclusion probabilities in the Bayesian SSS model, $\{p_i\}$
    - 1704 probesets are non-zero
    - 10 probesets have $p < 0.05$ (labeled in plots)
- Genes are colored as positively associated (blue) or negatively associated with complete clinical response (red).
    - $|t| > 2$, for the top 10 members
    - eCDF of all 1704 deviates from the unit line
- In the test set, differential expression within the signature was attenuated, ($|t| < 2$ for the top 10 members)
- The rank order of signature members was largely conserved, and signature-wide association to CR was seen ($p = 0.004$).





DukeMedicine

DUKE CANCER INSTITUTE

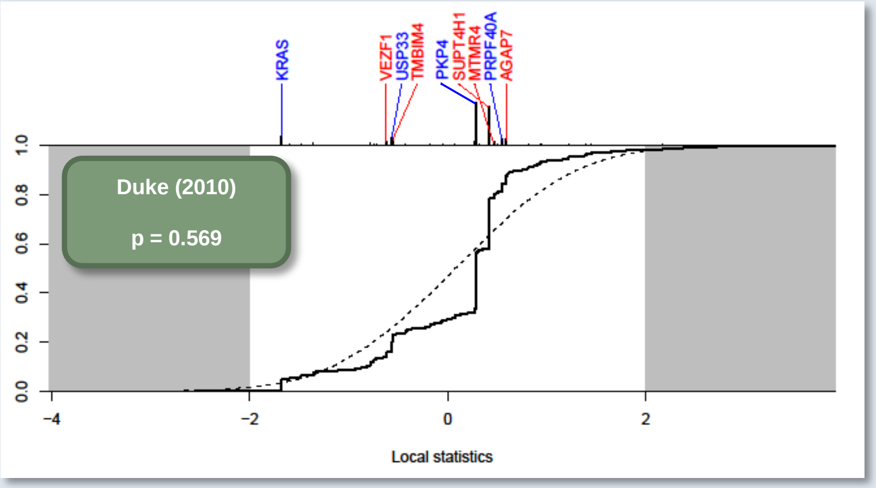
A National Cancer Institute-designated Comprehensive Cancer Center

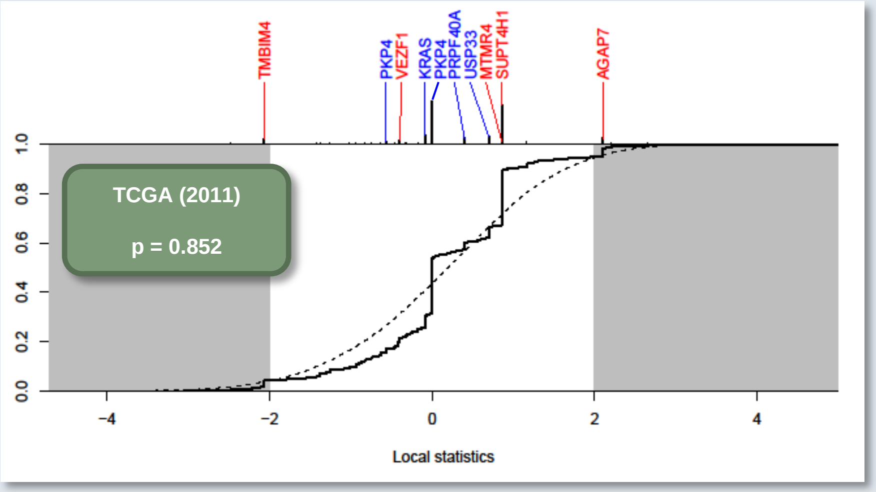
SAFE ANALYSIS OF THE OVARIAN SIGNATURE

- In the Duke (2010) cohort, no consistent patterns of differential expression was seen for genes in the signature (p > 0.5)

- In the TCGA (2011) cohort, an separate and inconsistent pattern of differential expression with was seen (p > 0.5)

- Based on the SAFE pathway analysis, methods to correction for technical batch effects between datasets should not improve performance.





TCGA – PROGNOSTIC SIGNATURE FOR SURVIVAL

- A genomic signature was developed for overall survival (split-sample), using Affymetrix-HT, -Exon and Agilent data.

- Genes with Cox PH $p < 0.01$ were included in the model
  - 108 genes positively correlated with outcome (labeled “good”),
  - and 85 genes negatively correlated with outcome (“poor”)

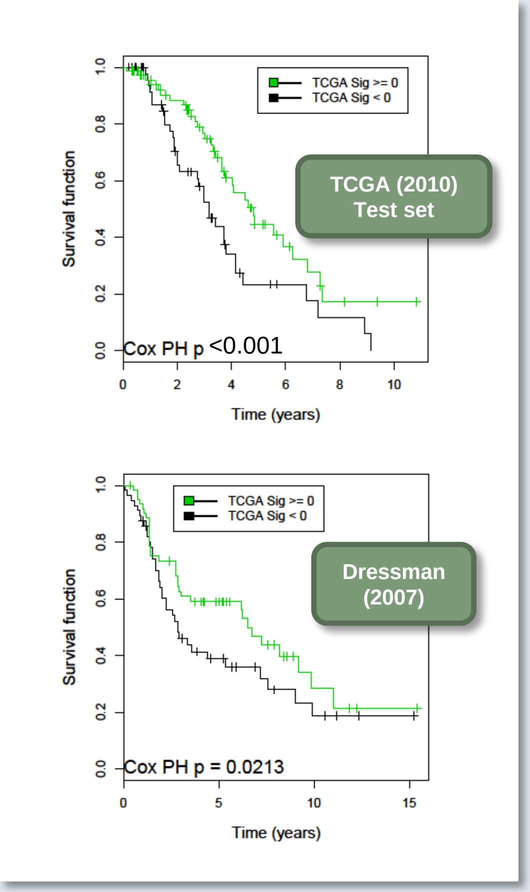
- Investigational dataset:
  - Gene expression is standardized,
  - “t-score” is a two-sample comparison of “good” and “poor” genes.

$$Score = \frac{\frac{1}{n_g}\sum_{i=1}^{n_g} Z_{g,i} - \frac{1}{n_p}\sum_{i=1}^{n_p} Z_{p,i}}{\widehat{sd}}$$

- The Dressman (2007) dataset was successfully used as an external validation of the signature.

- Using Affymetrix only, we applied the method to the internal and external validation sets and get concordant results.

Plots: Verification of TCGA sig.



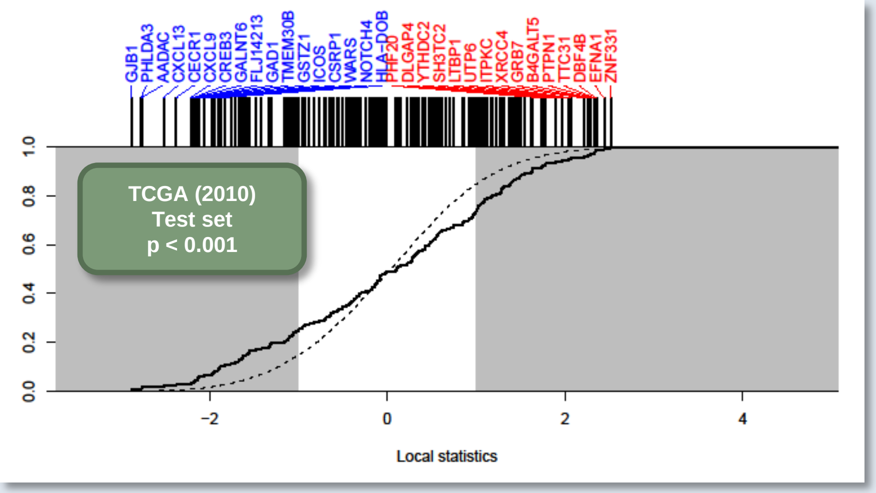
SAFE ANALYSIS OF THE TCGA SIGNATURE

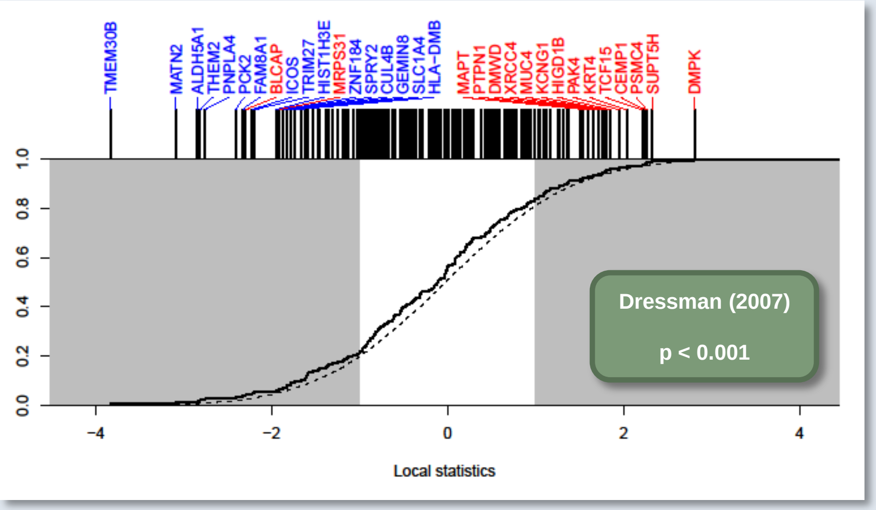
- As a proof-of-principle to the SAFE procedure, we applied the bootstrap-based test to the TCGA signature weighted by:

$$\{p_i, i \in g\} = \frac{1}{n_g} \qquad \{p_i, i \in p\} = \frac{1}{n_p}$$

- Statistically significant associations with outcome are observed, even when the eCDF does not strongly deviate from the global array.

Note: TCGA sig. did not validate in Duke (2010)





CONCLUSIONS

- The signature of response to platinum-based therapy (Dressman et al, 2007) was verified; but fails to predict clinical outcome in two independent cohorts.

- No overlap in the patterns of differential expression between incomplete and complete response are seen in the three cohorts.

  This could be attributed to a lack of standardized outcomes, or the failure to include known clinical factors, but highlights the limitations of attempts to do whole genome data-driven approaches to discovery in ovarian cancer.

- We propose a novel extension to pathway-analyses (Barry et al 2005, 2008, in preparation) as a first step to evaluate signatures before translating predictive models across platforms.

- The methods used in the TCGA signature are more readily translated across platforms. However, the algorithm can not be applied in a prospective manner.

ACKNOWLEDGMENTS

Members of Duke IGSP:
Holly Dressman
Andrew Berchuck
John Lancaster (Moffit)

Dept’s of Biostatistics and Bioinformatics:
Liz Delong
Steven George
Kouros Owzar
Herbert Pang
Fred Wright (UNC)
Andrew Nobel (UNC)

This work was supported by a grant from the National Cancer Institute (RC2CA148041-01).

Computational resources were supported by grants from the National Institute of Health (1S10RR025590-01) and the North Carolina Biotechnology Center (2009-IDG-1002).

DUKE CANCER INSTITUTE