

A 12-gene genomic instability signature predicts clinical outcomes in multiple cancer types

Rama K.R. Mettu^{1*}, Ying-Wooi Wan^{1*}, Jens K. Habermann^{2,3,4}, Thomas Ried², Nancy Lan Guo¹

¹Mary Babb Randolph Cancer Center/Community Medicine, West Virginia University, Morgantown, WV - USA

²Genetics Branch, National Cancer Institute, NIH, Bethesda, MD - USA

³Department of Surgery, University Hospital Schleswig-Holstein, Lübeck - Germany

⁴Karolinska Biomic Center (KBC), Karolinska Institute, Stockholm - Sweden

*These authors contributed equally to this study

ABSTRACT

Background and aims: Genomic instability, as reflected in specific chromosomal aneuploidies and variation in the nuclear DNA content, is a defining feature of human carcinomas. It is solidly established that the degree of genomic instability influences clinical outcome. We have recently identified a 12-gene expression signature that discerned genomically stable from unstable breast carcinomas. This gene expression signature was also useful to predict, with high accuracy, the clinical course in independent multiple published breast cancer cohorts. From a biological point of view, this result confirmed the central role of genomic instability for a tumor's ability to adapt to external challenges and selective pressure, and hence for continued survival fitness. This prompted us to investigate whether this genomic instability signature could also predict clinical outcome in other cancer types of epithelial origin, including colorectal tumors, non-small cell lung carcinomas, and ovarian cancer.

Results: The results show that the gene expression signature that defines genomic instability and poor outcome in breast cancer contributes significantly more accurate ($p < 0.05$ compared with random prediction) prognostic information in multiple cancer types independent of established clinical parameters. The 12-gene genomic instability signature stratified patients into high- and low-risk groups with distinct postoperative survival in three non-small cell lung cancer cohorts ($n=637$) in Kaplan-Meier analyses (\log -rank $p < 0.05$). It predicted recurrence in colon cancer patients ($n=92$) with an overall accuracy greater than 69% ($p=0.04$) in cross-cohort validation. It quantified relapse-free survival in ovarian cancer ($n=124$; \log -rank $p < 0.05$). Functional pathway analysis revealed interactions between the 12 signature genes and well-known cancer hallmarks.

Conclusion: The degree of genomic instability has diagnostic and prognostic implications. It is tempting to speculate that pursuing genomic instability therapeutically could provide entry points for a target that is unique to cancer cells.

Key words: Genomic instability, Gene expression signatures, Breast cancer, Ovarian cancer, Colon cancer, Non-small cell lung cancer

Received: February 26, 2010; Accepted: October 12, 2010

INTRODUCTION

Aneuploidy, chromosomal instability, and resulting genomic imbalances are among the hallmarks of human cancers of epithelial origin (1). While specific chromosomal imbalances are usually acquired before the transition to invasive disease, global destabilization of the genome occurs at later stages (2). The degree of destabilization, measured for instance by the number of chromosomal aberrations or the amount of variability in the nuclear DNA content from one cell to another, is an important predictor of clinical outcome, independent of conventional morphological or clinical parameters. For instance, women with genomically stable breast

carcinomas have considerably prolonged disease-free survival times compared to women whose tumors are genomically unstable (3). A similar picture emerges in prostate carcinomas (4). More recently, the search for predictors of clinical outcome was extended to embrace global gene expression profiling, and numerous signatures of poor prognosis or treatment failure were described (5-7). In an attempt to understand the biological basis of these signatures we analyzed, using gene expression profiling, breast carcinomas that had been characterized as genomically stable or unstable, hence of good and poor prognosis, respectively (8). These analyses led us to identify a set of 12 genes which were differentially expressed between stable and unstable tumors (8). The

biological and clinical relevance of these genes, i.e., the gene expression signature of genomic instability, was validated with large independent data sets and resulted in an excellent prediction of the clinical course. Inversely, when using established signatures of poor prognosis to predict the degree of genomic instability in our dataset, the results were equally convincing. That means that the basis for clinically used gene expression signatures of poor prognosis is biologically linked to chromosomal instability, which ultimately determines outcome. Since most solid tumors are defined by aneuploidy, we were interested to explore whether the gene expression signature of genomic instability that determines outcome in breast cancer is also useful for the prediction of the clinical course in other carcinomas, namely colorectal, lung, and ovarian cancer.

MATERIAL AND METHODS

Patients and samples

Colon cancer

The first cohort contained 50 patients with stage II colon adenocarcinoma (9). None of the patients had emergency surgery or received any adjuvant chemotherapy. Twenty-five patients developed a distant metastasis (liver in 20 patients, lung in 5 patients) within 52 months. The other 25 patients remained disease free for at least 60 months, with mean follow-up of 79 months. The second cohort contained 24 patients with stage II colon adenocarcinoma (10). None of these patients received adjuvant chemotherapy. Ten patients developed a liver metastasis within 55 months. The other 14 patients remained disease free for at least 60 months, with mean follow-up of 72.2 months. The third cohort contained 18 patients with colon adenocarcinoma (11). Ten of these patients had no lymph node metastasis (stage II) and did not receive any chemotherapy. The

other 8 patients had lymph node metastasis (stage III) and received 6-month adjuvant chemotherapy with 5-FU and levamisole. Patients were evaluated at 3-month intervals for the first postoperative year and at 6-month intervals thereafter. Nine of the 18 patients (5 stage II patients and 4 stage III patients) developed a distant metastasis within 53 months. The other 9 patients remained disease free for at least 60 months, with mean follow-up of 75 months. A summary of patient clinical characteristics is provided in Table I.

Non-small cell lung cancer

The cohort from Bild et al (12) contained 111 patients of which 67 were in stage I, 18 in stage II, 24 in stage III, and 2 in stage IV. There were 2 cell types in this cohort: lung adenocarcinoma (n=58) and squamous cell lung cancer (n=53). The cohort from Bhattacharjee et al (13) contained 84 patients with lung adenocarcinoma. Sixty-two patients were in stage I, 14 in stage II, and 8 in stage III. Twenty-six tumors were well differentiated, 43 moderately differentiated, and 15 poorly differentiated. The cohort from Shedden et al (14) contained 442 lung adenocarcinomas collected from multiple cancer centers and institutes. Two hundred and seventy-six patients were in stage I, 94 in stage II, and 68 in stage III; 4 patients had undefined disease stage. A summary of clinical characteristics including age, gender, and median follow-up time for each cohort is given in Table II.

Ovarian cancer

The ovarian cancer cohort (n=124) was retrieved from Bild et al (12); 94.4% (117/124) of these ovarian cancer patients had advanced stages (III and IV). The cohort was randomly partitioned into training and test sets of equal size. Table III presents the median follow-up time and the distribution of tumor stage for the training and test cohorts.

TABLE I - CLINICAL CHARACTERISTICS OF COLON CANCER PATIENT COHORTS

	Barrier et al (9) (n=50)	Barrier et al (10) (n=24)	Barrier et al (11) (n=18)
Mean follow-up (months)			
Patients with recurrence	52	55	53
Patients without recurrence	79	72.2	75
Tumor stage			
Stage II	100%	100%	55.6%
Stage III	0%	0%	44.4%
Recurrence within 5 years of surgery			
Yes	50%	58.3%	50%
No	50%	41.7%	50%

TABLE II - CLINICAL CHARACTERISTICS OF LUNG CANCER PATIENT COHORTS

	Bild et al (12) (n=111)	Bhattacharjee et al (13) (n=84)	Shedden et al (14) (n=442)
Histology			
Adenocarcinoma	52%	100%	100%
Squamous cell	48%		
Median follow-up (months)	31	38	47
Age (mean, SD)	65 (10)	63 (11)	64 (10)
Sex			
Male	57%	43%	50%
Female	43%	57%	50%
Tumor stage			
Stage I	60%	74%	62%
Stage II	16%	17%	22%
Stage III	22%	9 %	15%
Stage IV	2%	-	-
Unknown	-	-	1%

SD, standard deviation

TABLE III - CLINICAL CHARACTERISTICS OF OVARIAN CANCER PATIENT COHORTS

	Bild et al (12) (Training, n=62)	Bild et al (12) (Testing, n=62)
Median follow-up (months)	44	35
Tumor stage		
Stage I	5%	-
Stage II	5%	2%
Stage III	77%	80%
Stage IV	11%	18%
Unknown	2%	-

DNA microarray analysis

The RNA extraction and cDNA preparation in these studies was described in their original publications (9-14). Three colon cancer datasets were all generated with Affymetrix U133A arrays. The lung adenocarcinoma dataset from Bhattacharjee et al (13) was measured on Affymetrix U95A arrays. The non-small cell lung cancer dataset from Bild et al (12) was quantified with Affymetrix U133 Plus 2.0 arrays. The lung adenocarcinoma dataset from Shedden et al (14) was generated with Affymetrix U133A. The ovarian cancer dataset from Bild et al (12) was assayed with Affymetrix U133A (retrieved with record GSE3149 from Gene Expression Omnibus).

Computational methods for prognostic classifications

RIPPER

RIPPER is a propositional rule learning algorithm proposed by Cohen (15) with improvements over original

incremental reduced error pruning (IREP). In the RIPPER algorithm, after an initial rule set is learned from IREP, the rule set is further pruned repeatedly based on a different metric and stopping condition on randomized data. The repeated pruning stops when the rule set learned from IREP is refined into a rule set with optimized size and performance. JRip learner with WEKA 3.4 (16) was employed in the analysis.

Cox proportional hazard model

In survival analysis, the hazard function assesses the instantaneous risk of death at time t , condition on survival to that time point:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

where T is the variable that represents the survival time with cumulative distribution function $P(t) = \Pr(T \leq t)$.

A Cox proportional hazard model defines the relationship between the survival of patients and a set of variables, such as gene expressions. The Cox model gives the hazard at time t for an individual with a given set of predictors denoted by X :

$$h(t, X) = h_0(t) e^{\sum_{i=1}^n \beta_i x_i}, X = (X_1, X_2, \dots, X_n)$$

A hazard ratio is defined as the hazard for one individual divided by the hazard for a difference individual:

$$\text{hazard ratio} = \frac{\hat{h}(t, X^*)}{\hat{h}(t, X)}$$

where X^* denotes the set of predictors for one

individual and X denotes the set of predictors for the other individual (17).

In our analysis, the hazard ratio represents the ratio of hazard (i.e., death from cancer) between the average risk scores of 2 prognostic groups.

Prognostic prediction of recurrence in colon cancer

The matching genes in the 12-gene genomic instability signature were identified with Affymetrix IDs. Nine common genes were found in each of the 3 colon cancer cohorts with 6 genes having matches to multiple probes. The mean expression of the duplicate probes for each gene was used in this study. The patient cohort from Barrier et al (9) was used as a training set ($n=50$), while the cohorts from another 2 papers by Barrier et al (10, 11) were combined into an independent validation set ($n=42$). A training model was built with the 9 signature genes to classify recurrence in colon cancer patients using a rule-based learner RIPPER. A 10-fold cross-validation was used to evaluate the performance of the training model. This training model was used to predict recurrence in each patient in the validation set.

Prognostic categorization of non-small cell lung cancer

The DNA microarray data were generated on 3 lung cancer cohorts using different Affymetrix platforms. Gene symbols were used to find the matching genes in the signature. In each patient cohort, a Cox proportional hazard model was constructed by using the matching genes as covariates to predict lung cancer survival after the initial treatment. The non-small cell lung cancer cohort ($n=111$) from Bild et al (12) was used as a training set. A risk score was generated for each patient in this cohort. A high risk score represents a high probability of postoperative treatment failure, and similarly for a low risk score. Based on the distribution of the risk scores in this cohort, a cutoff point was identified to stratify patients into high-risk or low-risk groups. This cutoff risk score was applied in prognostic categorization in the 2 lung adenocarcinoma patient cohorts from Bhattacharjee et al (13) ($n=84$) and Shedden et al ($n=442$) (14). In each cohort, the survival probability of each prognostic group was assessed with Kaplan-Meier analysis. The difference between the survival probabilities in the 2 groups was estimated with log-rank tests. These analyses were performed with software packages in *R* (17).

Patient stratification in ovarian cancer

The ovarian cancer patient cohort ($n=124$) from Bild et al (12) was randomly partitioned into a training set ($n=62$) and a testing set ($n=62$). Using the training set, a Cox model was built based on the 12-gene genomic instability

signature as covariates. A risk score was generated for each patient. Based on the distribution of the risk scores in the training model, a cutoff value was identified for patient stratification. This training model and cutoff value were applied to the testing set.

Biological pathway analysis

Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems, Redwood City, CA, USA) is a proprietary web-based curated database which provides contents of gene and protein interactions reported in the literature. In this study, we used IPA to delineate molecular networks of genes interacting with the 12-gene signature. Core analysis was used to identify the most significant biological processes and functions from the merged network related to the 12-gene signature in human tissues and cell lines.

RESULTS

Genomic instability is a defining feature of human cancers of epithelial origin (1, 2). We have previously established a biological relationship between the degree of genomic instability, poor-prognosis gene expression signatures, and clinical outcomes in patients with breast cancer (8). A set of 12 genes that was differentially expressed between genomically stable and unstable tumors predicted recurrence-free survival (including metastasis and relapse) and overall survival in multiple independent breast cancer cohorts. In the present study, we sought to investigate whether this genomic instability gene signature also predicts clinical outcomes in other cancer types of epithelial origin including colon cancer ($n=92$), non-small cell lung cancer ($n=637$), and ovarian cancer ($n=124$). The signature genes are listed in Table IV.

Genomic instability signature is an independent predictor of colon cancer recurrence

To construct a molecular classifier to predict colon cancer recurrence, 50 patients with stage II colon adenocarcinoma (9) were used as a training cohort. Nine genes within the 12-gene genomic instability signature were identified from the DNA microarray data. These signature genes were used to classify recurrence in each patient with the JRip algorithm. The performance of the classifier was evaluated in a 10-fold cross-validation on the training set (Tab. V). The genomic instability signature correctly predicted recurrence in 72% (36/50) of patients, with a sensitivity of 68% (17/25) and a specificity of 76% (19/25). The model identified in the training cohort was applied to predict recurrence in each patient in the validation cohort ($n=42$), combining patients retrieved from Barrier et al (10, 11). In the validation, the genomic

TABLE IV - THE 12-GENE GENOMIC INSTABILITY SIGNATURE

Gene symbol	Gene description	Clone ID	Chromosome location
NXF1	Nuclear RNA report factor 1	1722870	11q12-q13
cDNA DKFZp762M127	Homo sapiens mRNA	1822809	11
P28	Dynein, axonemal, light intermediate	1998792	1p35.1
KIAA0882	KIAA0882 protein	2190664	4q31.1
v-myb	Avian myeloblastosis viral oncogene	2555590	6q22-q23
CDKN2A	Cyclin-dependent kinase inhibitor	2740235	9p21
unknown	Human clone 23948 mRNA sequence	3123244	15q22.32
RERG	RAS-like, estrogen-regulated; growth inhibitor	644989	12p13.1
SCYA18	Small inducible cytokine subfamily	690231	17q11.2
STK15	Serine/threonine kinase 15	2007691	20q13.2-q13.3
HNF3A	Hepatocyte nuclear factor 3	171194	12p13.1
unknown	Incyte EST	88935	N/A

TABLE V - PREDICTION ACCURACY OF COLON CANCER RECURRENCE USING THE 12-GENE GENOMIC INSTABILITY SIGNATURE

Patients	Sensitivity (recurrence within 5 y)	Specificity (no recurrence within 5 y)	Overall accuracy	P value*
Training set (n=50) (9) All stage II	68% (17/25)	76% (19/25)	72% (36/50)	0.01
Validation set (n=42) (10;11)	73.9% (14/19)	65.2% (15/23)	69.1% (29/42)	0.04

*P<0.05 means that the overall prediction accuracy of the 12-gene genomic instability signature is significantly higher than that of random prediction (2-sided Z-tests).

instability signature correctly predicted recurrence in 69.1% (29/42) of patients, with a sensitivity of 73.9% (14/19) and a specificity of 65.2% (15/23). The cohorts from Barrier et al (9, 10) contained only stage II lymph node-negative colon adenocarcinomas. These results indicate that the 12-gene genomic instability signature provided independent prognostic information in addition to tumor stage. Once validated in larger, independent cohorts this signature could be potentially used to select lymph node-negative patients for receiving adjuvant chemotherapy.

Genomic instability signature predicts lung cancer survival

To explore the clinical relevance of the 12-gene genomic instability signature for the prognostication of patients with non-small cell lung cancer, the lung cancer cohort retrieved from Bild et al (12) was used as a training cohort. A Cox model of overall survival was constructed based on the 12-gene signature, with each gene variable as a covariate. A survival risk score was generated for

every patient, with a higher risk score representing a greater probability of treatment failure (i.e., death). Based on the histogram representing distribution of gene expression-defined risk scores in this cohort (Fig. 1A), a cutoff value of zero, the peak value in the histogram, was used to stratify patients into high-risk and low-risk groups. This cutoff value represents the linear additive expression levels of all the signature genes in lung cancer patients. This stratification separated patients into 2 groups with distinct overall survival (log-rank p=0.0005) in Kaplan-Meier analysis (Fig. 1B). This cutoff risk score was applied to 2 additional cohorts from Bhattacharjee et al (13) (n=84) and Shedden et al (14) (n=442). The 12-gene signature generated significant prognostic categorization in both lung adenocarcinoma cohorts from Bhattacharjee et al (13) (log-rank p=0.05; Fig 1C) and Shedden et al (14) (log-rank p=0.01; Fig. 1D) in Kaplan-Meier analyses. In all 3 studied non-small cell lung cancer cohorts, the low-risk groups had a 2-year postoperative survival above 80%, representing a significantly better prognosis compared with

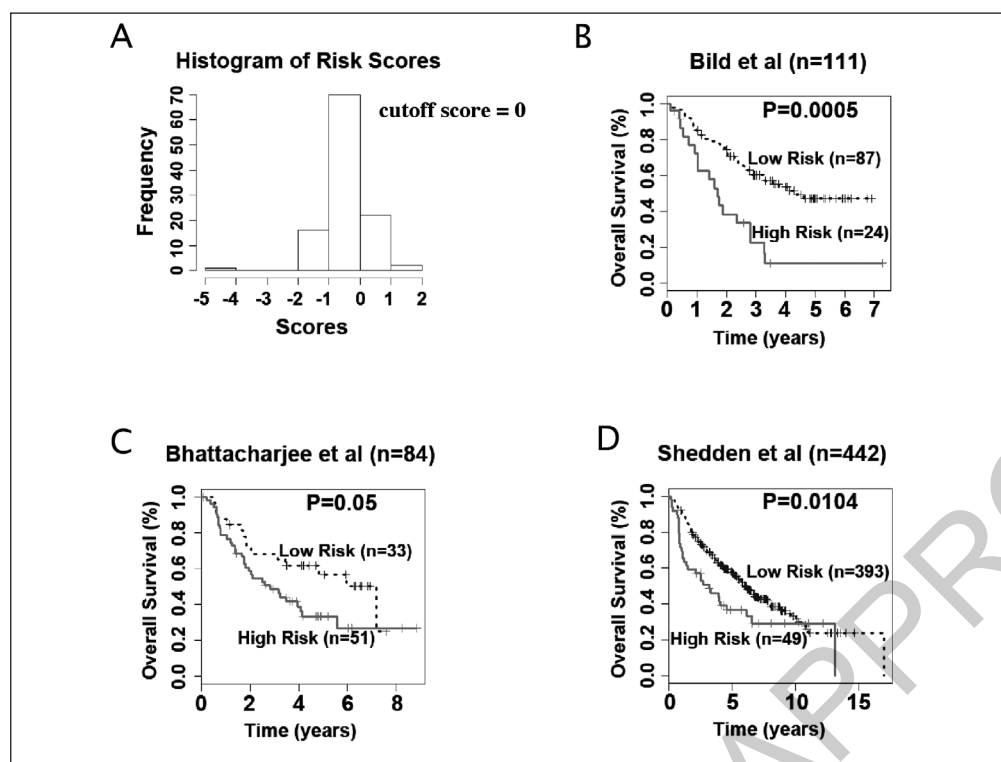


Fig 1 - The 12-gene genomic instability signature predicts overall survival in non-small cell lung cancer. (A) Histogram of gene expression-defined risk scores in the training cohort from Bild et al (12). The peak value with risk score zero in the histogram was defined as the cutoff in prognostic categorization. Gene expression-defined high-risk and low-risk patient groups had significantly different postoperative survival in cohorts from (B) Bild et al (12), (C) Bhattacharjee et al (13), and (D) Shedden et al (14) in Kaplan-Meier analyses.

the corresponding high-risk groups, for which the 2-year survival ranged from 38% to 58%. Furthermore, the 12-gene genomic instability signature had a significant hazard ratio ($HR \geq 1.55$, $p < 0.05$) in predicting poor prognosis in all 3 studied lung cancer cohorts (Tab. VI). The results show that the expression patterns of the 12-gene genomic instability signature could be used to predict postoperative survival in non-small cell lung cancer patients.

Genomic instability signature predicts ovarian cancer outcome

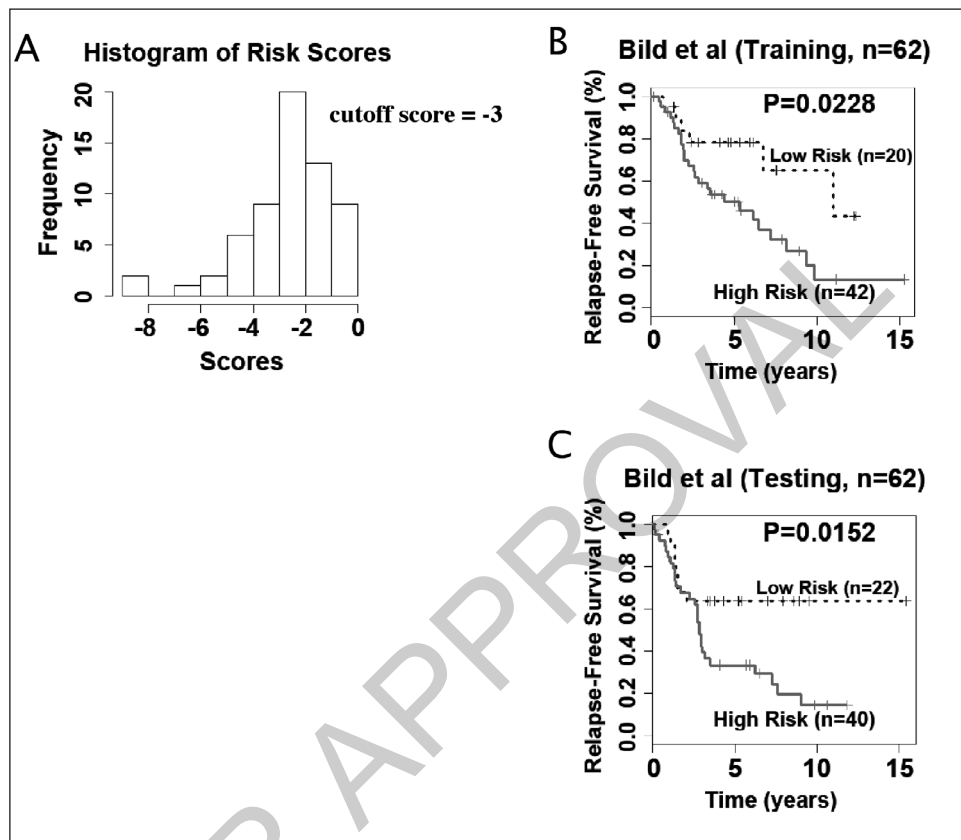
Ovarian cancer is a common malignancy in women. The prognosis of ovarian cancer patients is bleak due to a

TABLE VI - HAZARD RATIO OF PROGNOSTIC MODEL FOR LUNG CANCER AND OVARIAN CANCER PRESENTED IN FIGURE 1 AND FIGURE 2

Patient cohort	Hazard ratio (95% CI)
Lung cancer	
Bild et al (12) (n=111)	2.87 (1.69, 4.87)
Bhattacharjee et al (13) (n=84)	2.63 (1.61, 4.29)
Shedden et al (14) (n=442)	1.55 (1.23, 1.96)
Ovarian cancer	
Bild et al (12) (Training, n=62)	16.82 (5.05, 56.06)
Bild et al (12) (Testing, n=62)	2.08 (1.00, 4.29)

usually advanced disease stage at the time of diagnosis. In order to extend the potential usefulness of our genomic instability signature, we explored its value for predicting clinical outcome in patients with ovarian cancer. The ovarian cancer cohort (n=124) from Bild et al (12) was randomly split into a training set (n=62) and a testing set (n=62). A Cox model was built on the training set using the signature genes as covariates. A survival risk score was generated for each patient. A cutoff value was identified based on the histogram of the risk scores in the training set (Fig. 2A). Patients with a risk score greater than the cutoff were stratified into the high-risk group and the remaining patients into the low-risk group. The high- and low-risk groups had significantly (\log -rank $p=0.02$) different relapse-free survival in the training cohort in Kaplan-Meier analysis (Fig. 2B). This training model and stratification scheme were applied to the testing set and generated significant prognostic stratification (\log -rank $p=0.015$) in Kaplan-Meier analysis (Fig. 2C). The survival risk score estimated by the model also showed strong association with the ovarian cancer relapse-free survival ($HR=2.08$, 95% CI 1.00-4.29) in the test set (Table VI). These results demonstrate that the 12-gene genomic instability signature could identify more aggressive ovarian cancers that were more likely to develop recurrence after surgical resection and initial treatment. The high-risk patients defined with this genomic instability gene signature might benefit from second-line chemotherapy.

Fig. 2 - The 12-gene genomic instability signature predicts overall survival in ovarian cancer. (A) Histogram of gene expression-defined risk scores in the training cohort from Bild et al (12). The peak value with a risk score of -3 in the histogram was defined as the cutoff in prognostic categorization. Gene expression-defined high-risk and low-risk ovarian cancer groups had significantly different postoperative survival in both training (B) and testing (C) cohorts.



Functional pathway analysis

The 12-gene genomic instability signature was able to distinguish more aggressive tumors in multiple cancer types, indicating that the signature might be involved in important mechanisms of tumor genesis and progression. Functional pathway analysis was performed based on a curated database of molecular interactions reported in the literature using IPA. The results show that the signature genes interact with multiple prominent cancer signaling pathways, including the oncogenes *NFKB1A*, *MYC*, *BCL2*, *CDK1 (CDC2)*, *E2F1*, and *SOD2* as well as the tumor suppressor genes *TP53* and *CASP9* (Fig. 3). Specifically, *Aurora-A (AURKA)* is an important regulator of mitosis whose amplification is associated with human cancer. Perturbed expression of *Aurora-A* in mammalian cells induces carcinogenic phenotypes, including centrosome amplification, genomic instability, and transformation. It was found that *UBE2N* binds to *Aurora-A* (18) and inhibits it both in vitro and in vivo through the stimulation of ubiquitination (19). In MCF7 cells, human *Aurora-A* protein increases polyubiquitination of human *IKBA (NFKB1A)* protein (19). *p14^{ARF} (CDKN2A)*-induced G₂ arrest and subsequent apoptosis inhibit the growth of human tumoral cells lacking functional *TP53*, which correlated with inhibition of *CDC2 (CDK1)* activity (20). Studies showed that human *p14^{ARF} (CDKN2A)* protein

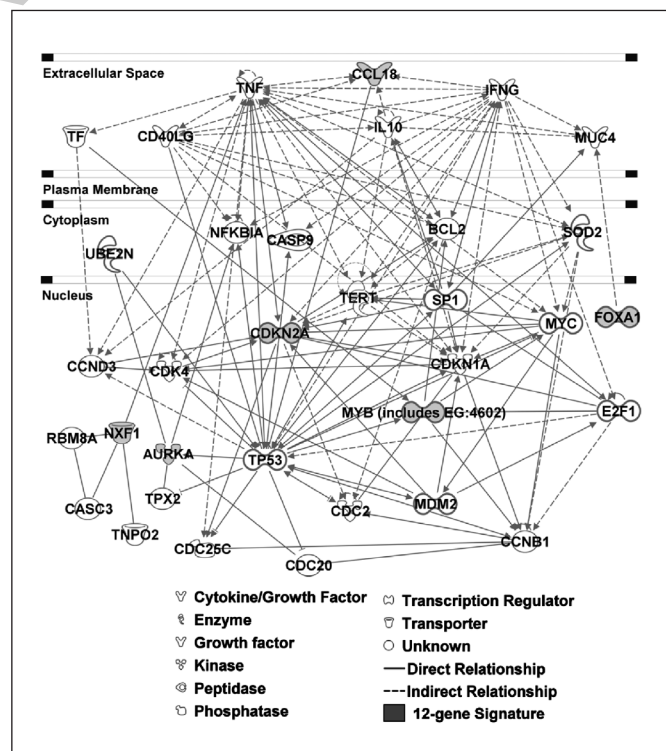


Fig. 3 - Functional pathway analysis of the 12-gene genomic instability signature using Ingenuity Pathway Analysis. The biological network showed genes interacting with the signature genes as reported in the literature.

activates the *Caspase 9 (CASP9)* and mitochondrial apoptosis pathway entirely independent of *TP53*, and these caspase-9-like activities were greatly enhanced in cells lacking functional *p21 (CDKN1A)* (21). Furthermore, overexpression of *SOD2* decreases accumulation of *CDKN2A* during confluent growth of fibroblasts (22). *BCL-2* is a direct target of *c-Myb*, and overexpression of *c-Myb* was accompanied by upregulation of *BCL-2* expression (23). The binding of *E2F* family members (*E2F1* through *E2F5*) is blocked by methylated *E2F* elements derived from *E2F1* and *CDC2* promoters (24). In contrast, the binding of *E2F1*, but not *E2F2* through *E2F5*, is significantly inhibited by methylated *E2F* elements derived from the *c-Myc* and *c-Myb* promoters (24). *CCL18* is one of the most abundant chemokines produced by immature dendritic cells, and may be part of an inhibitory pathway to limit specific immune responses at peripheral sites in maturing human dendritic cells (25). *TNF- α* (*TNF*) protein decreases expression of *CCL18* mRNA (25).

DISCUSSION

Genomic instability is a defining and ubiquitous feature of human cancers (1, 2). In many instances the degree of genomic instability is correlated with the clinical course, with highly unstable tumors showing a poorer prognosis (3, 26, 27). Also, the degree of genomic instability increases with cancer progression: preinvasive dysplastic lesions usually carry only few genomic imbalances, often in the form of gains and losses of entire chromosomes or chromosome arms (2, 28). These early changes can occur in an otherwise stable genome. During tumor progression, additional imbalances accumulate, including regional high-level copy number genomic amplification. Specific tumor entities can present with different degrees of genomic instability. For instance, HNPCC-associated colorectal carcinomas are karyotypically stable and generally have a better prognosis than sporadic tumors, which are usually highly aneuploid (29-31). A similar situation can be observed in breast carcinomas: here, patients with tumors that are diploid or close to diploid fare considerably better than patients with aneuploid carcinomas (3, 27). In a recent publication we explored whether that difference in the degree of genomic instability and its close correlation with prognosis are biologically related to the gene expression signatures of poor prognosis that were identified over the past few years in large cohorts of breast cancer patients. In order to test this hypothesis, we profiled 48 breast tumors that were defined as being genomically stable and unstable. This resulted in a 12-gene signature of genomic instability that clearly separated the 2 groups (8). In addition, the 12-gene signature predicted outcome in multiple independent datasets, and, conversely, the clinically used poor-prognosis signature

correctly predicted the degree of genomic instability in our dataset. This established a biological link between independently derived prognostic signatures (5-7). We were now interested in exploring whether this signature would also assist prognostication of disease outcome in cancer types other than breast carcinoma. In order to test this hypothesis, we applied the signature to published datasets from lung (12-14), colorectal (9-11), and ovarian carcinomas (12).

In each studied cancer type, a patient stratification scheme was developed based on the expression of the 12-gene genomic instability signature, and was validated on independent patient cohorts. Based on the clinical outcome provided in 3 colon cancer cohorts, a machine learning algorithm, JRip, was used in the model construction on the training set (n=50) with stage II colon carcinoma to predict patients' recurrence after surgery. The model accuracy was 72% on the training cohort in a 10-fold cross-validation. This prognostic model was applied to the combined testing set (n=42) and achieved an overall accuracy of 69.1% in the cross-cohort validation. These results are more accurate ($p < 0.05$) than random predictions. In the prognostic validation of non-small cell lung cancer, a prognostic model was built with the Cox model using the gene expression profiles as covariates. The cutoff point for prognostic categorization was based on a histogram of gene expression-defined risk scores in the training cohort (n=111). This stratification scheme was applied to 2 additional patient cohorts (n=526). In each cohort, the gene signature separated patients into different prognostic groups with remarkably different (\log -rank $p < 0.05$) clinical outcomes in Kaplan-Meier analyses. Similarly, the scheme was used in the prognostic validation of ovarian cancer. In both training and testing cohorts (n=124), the gene expression-defined model provided significant (\log -rank $p < 0.02$) postoperative prognostic stratification in Kaplan-Meier analyses. The 12-gene instability signature resulted in significantly higher hazard ratios in poor-prognosis groups over good-prognosis groups in all studied lung cancer and ovarian cancer cohorts.

Genome-wide association studies utilizing human tissue samples have enhanced the prognostic capacity of cancer outcomes. Four breast cancer signatures, including intrinsic subtypes (32), a poor-prognosis signature (MammaPrint[®]) (7), recurrence score (Oncotype DX[®]) (33), and wound response (34), represent largely the same prognostic space (35). Our 12-gene genomic instability signature predicted disease-free and overall survival in multiple breast cancer cohorts including early- and advanced-stage disease (8). Its accuracy proved to be comparable to that of Oncotype DX[®] and the signature is potentially more accurate than the other above-mentioned signatures in predicting survival (7). More importantly, the 12-gene signature showed prognostic ability beyond early-stage breast cancer, which constitutes the patient

group for MammaPrint® and Oncotype DX®. The 12-gene genomic instability signature quantified disease-free survival and overall survival in a broad patient population including patients with advanced tumor stage (T3/T4) and grade (G3), lymph node metastasis, and negative estrogen receptor status (8). The signature could extend the breast cancer prognostic space defined by MammaPrint® and Oncotype DX®, among other breast cancer signatures with potential clinical utility (12, 36-38).

This study demonstrates that the 12-gene genomic instability signature identified by us could predict clinical outcomes in multiple cancer types of epithelial origin. Thus, it provides further evidence linking genomic instability and associated gene expression with disease courses in human cancers. The IPA functional pathway analysis illustrates a biological network with interactions between the signature genes and numerous well-established cancer hallmarks, all of which were experimentally validated and reported in previous literature. These results further attest to the important roles of the genomic instability gene signature in tumor genesis and progression. While it is clear that the degree of genomic instability has diagnostic and prognostic implications, it is tempting to speculate that pursuing genomic instability therapeutically could provide entry points for a target that is unique to cancer cells.

ACKNOWLEDGMENTS

The authors express their gratitude to Mr Buddy Chen for editing the manuscript and illustrations.

Financial support: This research is supported by the following grants: T.R. by the National Cancer Institute (Intramural Research Program); N.L.G. by the National Library of Medicine (R01LM009500) and NCRR (P20 RR16440 and Supplement) from the NIH.

Conflict of interest statement: The Authors do not have conflict of interest in this work.

Address for correspondence:

Nancy L. Guo
Mary Babb Randolph Cancer Center/Community Medicine
West Virginia University
Morgantown, WV 26506-9300, USA
e-mail: lguo@hsc.wvu.edu

Thomas Ried
Section of Cancer Genomics National Institute of Health /
National Cancer Institute
50 South Drive Room 1408
Bethesda, MD 20892, USA
e-mail: riedt@mail.nih.gov

REFERENCES

1. Albertson DG, Collins C, McCormick F, Gray JW. Chromosome aberrations in solid tumors. *Nat Genet* 2003; 34: 369-76.
2. Ried T, Heselmeyer-Haddad K, Blegen H, Schrock E, Auer G. Genomic changes defining the genesis, progression, and malignancy potential in solid human tumors: a phenotype/genotype correlation. *Genes Chromosomes Cancer* 1999; 25: 195-204.
3. Auer G, Eriksson E, Azavedo E, Caspersson T, Wallgren A. Prognostic significance of nuclear DNA content in mammary adenocarcinomas in humans. *Cancer Res* 1984; 44: 394-6.
4. Kronenwett U, Huwendiek S, Ostring C, et al. Improved grading of breast adenocarcinomas based on genomic instability. *Cancer Res* 2004; 64: 904-9.
5. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001; 98: 10869-74.
6. Sotiriou C, Neo SY, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A* 2003; 100: 10393-8.
7. van de Vijver MJ, He YD, van 't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002; 347: 1999-2009.
8. Habermann JK, Doering J, Hautaniemi S, et al. The gene expression signature of genomic instability in breast cancer is an independent predictor of clinical outcome. *Int J Cancer* 2009;124: 1552-64.
9. Barrier A, Boelle PY, Roser F, et al. Stage II colon cancer prognosis prediction by tumor gene expression profiling. *J Clin Oncol* 2006; 24: 4685-91.
10. Barrier A, Roser F, Boelle PY, et al. Prognosis of stage II

- colon cancer by non-neoplastic mucosa gene expression profiling. *Oncogene* 2006; 26: 2642-8.
11. Barrier A, Lemoine A, Boelle PY, et al. Colon cancer prognosis prediction by gene expression profiling. *Oncogene* 2005; 24: 6155-64.
 12. Bild AH, Yao G, Chang JT, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 2006; 439: 353-7.
 13. Bhattacharjee A, Richards WG, Staunton J, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A* 2001; 98: 13790-5.
 14. Shedden K, Taylor JM, Enkemann SA, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* 2008; 14: 822-7.
 15. Cohen WW. Fast effective rule induction. In: *Proceedings of the Twelfth International Conference on Machine Learning*. Lake Tahoe, CA: Morgan Kaufmann Publishers 1995; 115-23.
 16. Witten IH, Frank E. *Data mining: practical machine learning tools and techniques* (2nd edition). San Francisco, CA: Morgan Kaufmann Publishers, 2005.
 17. Everitt B, Hothorn T. *A handbook of statistical analyses using R*. Boca Raton, FL: Chapman & Hall/CRC, 2006.
 18. Ewart-Toland A, Briassouli P, de Koning JP, et al. Identification of Stk6/STK15 as a candidate low-penetrance tumor-susceptibility gene in mouse and human. *Nat Genet* 2003; 34: 403-12.
 19. Briassouli P, Chan F, Linardopoulos S. The N-terminal domain of the Aurora-A Phe-31 variant encodes an E3 ubiquitin ligase and mediates ubiquitination of I κ B α . *Hum Mol Genet* 2006; 15: 3343-50.
 20. Eymin B, Leduc C, Coll JL, Brambilla E, Gazzeri S. p14ARF induces G2 arrest and apoptosis independently of p53 leading to regression of tumours established in nude mice. *Oncogene* 2003; 22: 1822-35.
 21. Hemmati PG, Normand G, Verdoodt B, et al. Loss of p21 disrupts p14 ARF-induced G1 cell cycle arrest but augments p14 ARF-induced apoptosis in human carcinoma cells. *Oncogene* 2005; 24: 4114-28.
 22. Sarsour EH, Agarwal M, Pandita TK, Oberley LW, Goswami PC. Manganese superoxide dismutase protects the proliferative capacity of confluent normal human fibroblasts. *J Biol Chem* 2005; 280: 18033-41.
 23. Salomoni P, Perrotti D, Martinez R, Franceschi C, Calabretta B. Resistance to apoptosis in CTLL-2 cells constitutively expressing c-Myb is associated with induction of BCL-2 expression and Myb-dependent regulation of bcl-2 promoter activity. *Proc Natl Acad Sci U S A* 1997; 94: 3296-301.
 24. Campanero MR, Armstrong MI, Flemington EK. CpG methylation as a mechanism for the regulation of E2F activity. *Proc Natl Acad Sci U S A* 2000 Jun 6; 97(12):6481-6.
 25. Vulcano M, Struyf S, Scapini P, et al. Unique regulation of CCL18 production by maturing dendritic cells. *J Immunol* 2003; 170: 3843-9.
 26. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* 2006; 38: 1043-8.
 27. Kronenwett U, Ploner A, Zetterberg A, Bergh J, Hall P, Auer G, Pawitan Y. Genomic instability and prognosis in breast carcinomas. *Cancer Epidemiol Biomarkers Prev* 2006; 15: 1630-5.
 28. Ried T. Homage to Theodor Boveri (1862-1915): Boveri's theory of cancer as a disease of the chromosomes, and the landscape of genomic imbalances in human carcinomas. *Environ Mol Mutagen* 2009; 50: 593-601.
 29. Eshleman JR, Casey G, Kochera ME, et al. Chromosome number and structure both are markedly stable in RER colorectal cancers and are not destabilized by mutation of p53. *Oncogene* 1998; 17: 719-25.
 30. Ghadimi BM, Sackett DL, Difilippantonio MJ, et al. Centrosome amplification and instability occurs exclusively in aneuploid, but not in diploid colorectal cancer cell lines, and correlates with numerical chromosomal aberrations. *Genes Chromosomes Cancer* 2000; 27: 183-90.
 31. Schlegel J, Stumm G, Scherthan H, et al. Comparative genomic in situ hybridization of colon carcinomas with replication error. *Cancer Res* 1995; 55: 6002-5.
 32. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003; 100: 8418-23.
 33. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004; 351: 2817-26.
 34. Chang HY, Nuyten DS, Sneddon JB, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci U S A* 2005; 102: 3738-43.
 35. Massague J. Sorting out breast-cancer gene signatures. *N Engl J Med* 2007; 356: 294-7.
 36. Chi JT, Wang Z, Nuyten DS, et al. Gene expression programs in response to hypoxia: cell type specificity and prognostic significance in human cancers. *PLoS Med* 2006; 3: e47.
 37. Liu R, Wang X, Chen GY, et al. The prognostic role of a gene signature from tumorigenic breast-cancer cells. *N Engl J Med* 2007; 356: 217-26.
 38. Minn AJ, Gupta GP, Siegel PM, et al. Genes that mediate breast cancer metastasis to lung. *Nature* 2005; 436: 518-24.