Published in final edited form as:

Radiat Res. 2009 February ; 171(2): 141–154. doi:10.1667/RR1223.1.

A Radiation-Derived Gene Expression Signature Predicts Clinical Outcome for Breast Cancer Patients

Brian D. Piening^{a,c,1}, Pei Wang^{a,1}, Aravind Subramanian^b, and Amanda G. Paulovich^{a,2}

^a *Fred Hutchinson Cancer Research Center, Seattle Washington*

^b *Broad Institute of MIT and Harvard, Cambridge, Massachusetts*

^c *Molecular and Cellular Biology Program, University of Washington, Seattle, Washington*

Abstract

Activation of the DNA damage response pathway is a hallmark for early tumorigenesis, while loss of pathway activity is associated with disease progression. Thus we hypothesized that a gene expression signature associated with the DNA damage response may serve as a prognostic signature for outcome in cancer patients. We identified ionizing radiation-responsive transcripts in human lymphoblast cells derived from 12 individuals and used this signature to screen a panel of cancer data sets for the ability to predict long-term survival of cancer patients. We demonstrate that gene sets induced or repressed by ionizing radiation can predict clinical outcome in two independent breast cancer data sets, and we compare the radiation signature to previously described gene expression-based outcome predictors. While genes repressed in response to radiation likely represent the well-characterized proliferation signature predictive of breast cancer outcome, genes induced by radiation likely encode additional information representing other deregulated biological properties of tumors such as checkpoint or apoptotic responses.

INTRODUCTION

Widespread use of mammography has radically shifted the distribution of breast cancer stages at diagnosis by increasing the proportion of early-stage cancers (1,2). Although the majority of these apparently early-stage cancers will be cured by surgery and loco-regional radiation, a substantial portion (up to one-third) of node-negative cancers will eventually recur and metastasize (3), and this aggressive subset of tumors requires additional treatment with adjuvant chemotherapy at the time of diagnosis. Because chemotherapy is associated with substantial side effects (4–7), new markers are urgently needed to identify patients who are at low risk for progression so that the subset of patients who will not relapse can be spared the toxicity of adjuvant chemotherapy and/or endocrine therapy (8–10). Conversely, identification of women at high risk of recurrence is critical to provide appropriate treatment that may prevent or delay recurrence and death.

There has been a rapid increase in efforts to phenotype breast tumors molecularly on the basis of gene expression patterns to predict outcome (11). The majority of studies have focused on the identification of prognostic signatures by hierarchical clustering using gene expression data from tumor tissue specimens. This approach has been fruitful. Specifically for breast cancer, no less than eight different expression signatures have been proposed for predicting clinical outcome [reviewed in ref. (12)]. The first major signature identified, the “70 gene expression

²Address for correspondence: 1100 Fairview Ave N, PO Box 19024, Mail Stop LE-360, Seattle, WA 98109; e-mail: apaulovi@fhcrc.org.

¹These authors contributed equally to this work.

profile”, demonstrated significant capability for identifying tumors at low and high risk of distant recurrence in several studies (13–15). A second profile, the “21 gene” RT-PCR assay, developed for recurrence risk classification in lymph node-negative ER⁺ patients (16), was also shown to be a significant predictor of disease outcome (17,18). Both of these gene signatures are now being assessed in prospective clinical trials for their impact on outcome when used in treatment assignment (19,20). More recently, a gene expression grade index (GGI) has shown promise in discriminating prognosis in grade 2 tumors (21,22). Other signatures that are associated with prognosis have been described, each offering a slight variation in terms of either signature derivation, targeted subgroup, or outcome of interest (23–29).

Although hierarchical clustering of tumor array data has been fruitful, we appear to have reached the limit of prognostic information that can be gleaned from this approach. In a 2006 analysis, Fan *et al.* (30) showed that despite minimal overlap between cancer outcome signatures, each yielded similar prognostic capabilities (31). In a more recent 2007 opinion piece, Sotiriou *et al.* (12) suggest that despite low overlap, all signatures are essentially detecting the same property: the high potential for proliferation in poor-outcome tumors.

Because all signatures are converging on the proliferation cluster, we need to explore alternative approaches for identifying additional molecular properties of tumors that can, when combined with the proliferation gene signature, expand our prognostic capabilities. Chang *et al.* (32) identified features in the transcriptional response of normal fibroblasts to serum (the “wound-response signature”), which links wound healing and cancer progression in a variety of common epithelial tumors, including breast cancer. To avoid rediscovery of the well-characterized proliferation gene cluster, the authors specifically excluded all genes known to be involved in proliferation. In a subsequent study of a series of 295 early breast cancer patients, they showed that both overall survival and distant metastasis-free survival are markedly diminished in patients whose tumors express this signature compared to tumors that did not express the signature (33). Of note, the signature improved risk stratification independently of known clinicopathological risk factors and previously established prognostic signatures (33). This study represents a novel approach to identifying prediction signatures by leveraging known biology underlying tumorigenic progression; in this study we describe a similar approach, focused not on the wound response but rather on the cellular response to ionizing radiation.

Recently, it has been shown that the DNA damage response (DDR) pathway is constitutively activated in early-stage breast cancers but not in normal epithelium (34,35). This observation has led to a model for tumorigenesis wherein activation of oncogenes promotes deregulated cell division, leading to replication stress, the formation of DNA double-strand breaks (by replication fork breakage) (36), and consequent activation of the ATM/ATR-dependent DDR (37). Notably, this activation is associated with the induction of oncogene-induced senescence, which is dependent on the DDR sensor kinase ATM (38,39). As such, activation of the DDR may act as a first-line defense to halt tumor progression, and further mutations (*p53*, for example) may allow tumors to evade this mechanism and progress to acquire increasingly more malignant properties. Consistent with this hypothesis, both the proliferative propensities of tumors (12) and their *p53* status (31,40) are prognostic markers in breast cancer. Based on the observation that the DDR is constitutively activated in early-stage lesions but is inactivated in late-stage tumors (34), we hypothesized that the transcriptional correlate of an activated DDR might have prognostic capability in breast cancers.

In this study, we performed whole genome transcriptional analysis to identify radiation-responsive genes in human lymphoblast cell lines derived from 12 individuals. Our results extend the current literature by reporting a number of transcripts not previously implicated as

being radiation-responsive. Using this radiation-responsive gene set as bait for Gene Set Enrichment Analysis (41), we identified a number of significant connections to human cancer. We demonstrate that our radiation-responsive signature serves as an effective predictor of survival outcome in two independent breast cancer data sets. We compare the radiation-response signature to previously developed outcome signatures and show similar prognostic capability, despite varying degrees of gene/gene overlap. Like the wound-response signature, the radiation-response signature is an effective tumor outcome classifier derived from a biological hypothesis based on known tumor physiology (e.g. abnormal checkpoint function and chromosomal instability).

MATERIALS AND METHODS

Cell Lines and Growth Conditions

The human lymphoblast (LBL) cells used in this study were obtained either from the Coriell Cell Repository [<http://ccr.coriell.org/>] (GM10832, GM10833, GM10835, GM6990, GM7057, GM13113, GM10860, GM10834, GM10861)] or were kindly provided by Dr. Daniel Haber and Daphne Bell at the M.G.H. Charlestown Navy Yard (NML37, NML47, NML50).

Fresh aliquots of cells were thawed and grown in RPMI 1640 medium containing 15% FBS (Sigma, heat-inactivated), 100 U/ml penicillin and 100 µg/ml streptomycin in a humidified incubator at 37°C in 95% air/5% CO₂. Cells were fed by addition of fresh, prewarmed medium two or three times per week as required to maintain the density at $2-7 \times 10^5$ cells/ml until there were sufficient quantities of cells for the experiment (10–14 days). At this point, 2×10^7 cells were harvested by centrifugation and resuspended in 30 ml of fresh medium. Resuspended cells were then divided into three 10-ml aliquots in T-25 Falcon tissue culture flasks and returned to the incubator, where they were allowed to return to baseline after this manipulation; specifically, the cultures were placed on the same shelf within the incubator, and the incubator door was not opened until the cells were irradiated 36–40 h later.

Cell Irradiation

After the 36–40-h recovery period described above, all cultures were removed *en masse* from the incubator and transported on a cart to the Gammacell 40 source. One aliquot of each culture was mock-irradiated and a second aliquot was exposed to 5.0 Gy using a ¹³⁷Cs Gammacell source operating at a dose rate of 96.1 cGy/min. After exposure or mock exposure, all cultures were then returned to the incubator for 5 h, after which they were harvested by centrifugation. The cell cycle distribution (assessed by flow cytometry) of the lymphoblasts was not significantly altered during the relatively short response period after irradiation (data not shown).

RNA Isolation, Hybridization, Washing and Scanning of Affymetrix U133A and U133B Chips

Total RNA was extracted using a standard Trizol (Gibco BRL) extraction (42,43). RNA samples were quantified by absorbance at 260 nm, and purity was assessed by determining ratios of A260/A280. For first-strand cDNA synthesis, 1 ml 100 pmol/ml T7-(T)24 primer [Genosys; GGCCAGTGAATTGTGAATACGACTCACTATAGGGAGGCGG-(T)24] was hybridized to 15 mg total RNA (65°C for 10 min, then transferred to ice) and subsequently incubated with 400 U Superscript II reverse transcriptase (Gibco BRL) in 1× buffer for 1 h at 42°C. Second-strand cDNA synthesis was carried out using 40 U DNA polymerase I (Gibco BRL), 10 U DNA ligase (Gibco BRL), 2 U RNase H (Gibco BRL), and 0.23 mM dNTPs (Gibco BRL) in 1× buffer at 16°C for 2 h. The resultant double-stranded cDNA was cleaned up by phenol extraction (PLG Phase-Lock Gel tubes, 5'-3') and ethanol precipitation. *In vitro* transcription was carried out according to the manufacturer's recommendations (Am-bion T7 MegaScript Kit) in the presence of Bio-11-CTP and Bio-16-UTP (Enzo). Resultant cRNAs

were cleaned up using Qiagen's RNeasy Mini Kit and fragmented in 1× RNA Fragmentation Buffer (5× stock: 200 mM Tris-acetate, pH 8.1, 500 mM KOAc, 150 mM MgOAc) for 35 min at 95°C. Fragmented cRNAs were hybridized to Affymetrix U133a and U133b arrays in MES Hybridization Buffer (100 mM MES, 1 M [Na⁺], 20 mM EDTA, 0.01% Tween 20) at 45°C, 60 rpm for 16 h. Control Oligo B2 (5'-Bio-GTCAAGATGCTACCGTTCA-3') was included. After hybridization, the probe arrays were washed (100 mM MES, 0.1 M [Na⁺], 0.01% Tween 20), and the remaining biotin-labeled target was then stained with a streptavidin R-phycoerythrin (Molecular Probes). The signal was then amplified with a biotinylated anti-streptavidin antibody (goat; Vector Laboratories), followed by a second staining with the streptavidin-conjugated fluorescent stain prior to scanning; 1× stain buffer contained 100 mM MES, 1 M [Na⁺], 0.05% Tween 20, and 0.005% An-tifoam. Arrays were scanned using a Hewlett Packard (HP) scanner using MAS5 software (www.affymetrix.com), and the final data set was normalized using RMA (quantile norm, default settings) (44). Normalized data are provided in the Supplementary Information.

Microarray Data Analysis

To search for genes differentially expressed between the irradiated group and unirradiated group, the signal-to-noise ratio (SNR) was calculated for each of the 44,760 probe sets as follows:

$$SNR_i \equiv \frac{\mu_{0,i} - \mu_{1,i}}{\sigma_{0,i} + \sigma_{1,i}}, \quad (1)$$

where $\mu_{0,i}$ and $\sigma_{0,i}$ are the mean and standard deviation of the i th probe set of the irradiated experimental group and $\mu_{1,i}$ and $\sigma_{1,i}$ are the mean and standard deviation of the unirradiated grouping. (i ranges from 0 to 44759). The data set was permuted 5000 times to generate a null distribution from which the q -value statistic was derived to control the false discovery rate (FDR) (45,46).

Pathway Enrichment Analysis Using GSEA

Pathway analysis using GSEA (41) was done in two different ways. First, *a priori* defined gene sets were tested against the radiation response for discrimination between the irradiated and unirradiated states. Gene sets were compiled from gene ontologies (47). Ontologies were separated at top-level categories (Biological Process, Cellular Component and Molecular Function), and gene sets were created by iteration over all subcategories and placing all genes common to a subcategory in a gene set representing that subcategory. Gene sets containing <25 or >500 genes were excluded from the analysis. The Gene Sets used in this analysis are provided in a zip file as Supplementary Information. Other parameters used as input for the GSEA program were as follows: The algorithm was Signal2Noise, the normalization method was meandiv, permutations were done by phenotype, the scoring scheme was weighted, and 1000 permutations were used for each analysis.

Curation of Expression Data Sets from the Literature

The NKI expression data set and patients' clinical information (14) were downloaded from the Stanford University public repository (http://microarray-pubs.stanford.edu/wound_NKI/explore.html). There are 244 arrays in total with 24,136 clones on each array. Global normalization was performed [center mean to 0 and scale MAD (median absolute deviation) to 1]. A second expression data set described by Loi *et. al.* (21) was downloaded from the Gene Expression Omnibus database (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE6532>, "LUMINAL. Rdata"). There are 277 arrays (from patients who received adjuvant tamoxifen treatment) with 44,928

clones on each array. The data set representing irradiated mammary tumor epithelial cells was extracted from a larger data set published by Amundson *et al.* (48), and the raw measurements were quantile normalized (R package “DNAMR”).

Analysis of Expression Signatures in Curated Expression Data Sets

Statistical analysis was conducted using the open source program R (<http://lib.stat.cmu.edu/R/CRAN/>). For hierarchical clustering, each gene was standardized to mean = 0 and MAD = 1. Euclidean distance was used with complete linkage in the R function *hclust* to derive hierarchical clustering. For survival analysis, R package *survival* was used to plot Kaplan-Meier survival curves. Log-rank tests were performed for comparing Kaplan-Meier curves of two different groups. The RI (or RR) signature of a sample was defined as the projection of measurements of the RI (or RR) genes on the first principal component direction of the RI (or RR) gene set across all the samples. These RI (or RR) signatures were then used in various Cox proportional hazard models fitted with the R function *coxph*. Fisher’s exact test was performed using the R function *fisher.test*. To identify genes serving as the key drivers of the clustering, we used the two-sided *t* statistic to test whether each gene was expressed differently between the two clusters. We then accounted for multiple hypothesis testing by deriving the *q* values (R packages *testp* and *qvalue*) and controlling the FDR.

In the classification analysis, patients were first separated into two groups: the poor clinical outcome group (deceased within 10 years of follow-up) and the good clinical outcome group. Then, for each gene signature, the leave-one-out cross validation was performed as follows: (1) For a selected test sample, we treated the rest of the data as the training set. (2) We calculated the mean expression level for the training samples in the good- and bad-survival groups and computed the distance between the expression level of the test sample and the centroids of the good- and poor-survival groups:

$$d(X^*) = \sum_{i=1}^p \frac{(x_i^* - \bar{x}_{ik})^2}{(s_i + s_0)^2} - 2 \log \pi_k,$$

where $X^* = \{x_i^*\}_i$ is the expression of the test sample; p is the total number of genes in this gene list; $\{\bar{x}_{ik}\}_i$ is the centroid of class k ($= 1$ or 0); s_i is the standard deviation of x_i ; s_0 is set to be 0.01 here; π_k is the proportion of samples in class k . (3) We then classified the test sample into the class with the smaller distance. (4) Last, we repeated the process with a new test sample until we had iterated over all samples.

Based on the classification results of the eight gene signatures, we further divided the patients into subgroups based on both the true outcomes and the classification by the expression signatures. We then compared the distribution of other known clinical variables in group 1 (good survival and classified as “good”) to group 3 (poor survival but classified as “good”), and group 2 (poor survival but classified as “good”) to group 4 (poor survival and classified as “poor”). For continuous variables such as age, diameter and lymph node, the Wilcoxon rank test was used. For binary or category variables, such as tumor grade (1,2,3), vascular invasion (0, 1–3, >3), ER (+, –), mastectomy or conserving therapy, and chemo/hormonal or other therapy, Fisher’s exact test or χ^2 tests were used.

RESULTS

Cellular Response to Ionizing Radiation

To characterize the human transcriptional response to γ radiation comprehensively, we generated a microarray data set consisting of 12 different lymphoblast cell lines profiled 5 h after exposure to 5 Gy radiation; control samples were mock-irradiated. We identified 160 genes induced after irradiation (211 probe sets) and 59 genes repressed after irradiation (70 probe sets) at $FDR \leq 0.05$ (Supplementary Table S1a–b). This large study identified known radiation-responsive transcripts (e.g. *GADD45A*, *XPC*, *PPM1D*, *FAS*, etc.) as well as many genes not previously implicated in the radiation response (e.g. *ZNF79*, *C11orf24*, *CSNK1G1*, *KCNN3*, etc.). We next asked what biological functions are associated with the genes repressed or induced after irradiation by computing enrichment scores for gene ontology (GO)-based gene sets as described in the Materials and Methods. We found that while the radiation-repressed (RR) gene set is enriched in genes that function in mitosis, the radiation-induced (RI) genes function in a number of different cellular pathways, including apoptosis and cell cycle arrest (see Supplementary Table S2). These observations are consistent with known biology from which we expect a radiation response to include activation of cell cycle checkpoints (and likely decreased expression of mitosis-related genes) as well as induction of repair factors, negative cell cycle regulators, and genes involved in apoptosis (e.g. caspases).

Survey of Human Cancer Data Sets Suggests that the Radiation Signature is Prognostic in Breast Cancer

Next, we explored the prognostic capabilities of radiation-responsive genes in human cancer data sets. We initially used the GSEA algorithm (41) to screen the sets of the top 160 RI genes and 59 RR genes for correlation with a series of clinical traits in a panel of cancer microarray data sets taken from the literature. These traits included clinical outcome in multiple cancer types (14,49–54), estrogen receptor status in breast cancers (14,55), *BRCA1* status in breast cancers (13), and p53 status in multiple cancer cell lines (56). The results of this screen are presented in Supplementary Table S3. The breast cancer data sets yielded particularly interesting results, including correlation of a radiation-derived gene set with clinical outcomes as well as both estrogen receptor and *BRCA1/2* mutant status.

Prognostic Capabilities of the Radiation Signature in Breast Cancer

After these observations from the initial screen, we further assessed the prognostic capabilities of the radiation signature in breast cancer using a more thorough analysis. We initially focused on a clinical outcome data set consisting of 244 breast cancer patients from the Netherlands Cancer Institute (NKI) (14). After we applied stringent filters to minimize false positives in the radiation signature (q value ≤ 0.02 , $SNR \geq 1.2$), we identified 50 RI genes and 68 RR genes. We cross-mapped these radiation-induced and radiation-repressed signatures to the NKI data set (15 RI genes + 31 RR genes overlapped; see Supplementary Table S1c) and clustered tumor samples in the space of the genes repressed by radiation (the RR gene set) or the space of genes induced by radiation (the RI gene set). As shown in Fig. 1A, by applying hierarchical clustering, the RR gene set naturally separated the tumor samples into two distinct clusters based solely on the expression of RR genes. We next performed Kaplan-Meier survival analysis on the two clusters (Fig. 1B), and we observed that primary tumors from the left cluster (Fig. 1A) were significantly more likely to progress to death than tumors from the right cluster ($P = 2.13 \times 10^{-7}$, log rank test). To determine the probability that this separation could be due to random chance, we repeatedly sampled a gene set of the same size (31 genes) from the original expression data set and again clustered the data and performed the log rank test. In three out of the 500 iterations, the log rank test P values were smaller than our calculated P value; hence the possibility of observing this association by chance was 0.006. We repeated the analysis using the RI signature, which also separated the tumor samples into good- and poor-outcome

clusters (Fig. 1C and D). The log rank test for the difference in outcome between the clusters yielded a P value of 0.0006, which corresponds to an error rate of 0.034 based on random sampling (17 out of 500 iterations have the same or a smaller P value). We also tested for an association between tumor clusters and tumor grade and ER status, which are represented visually above the heat maps in Fig. 1A and C. Tumor grade was strongly correlated with the clustering results of both the RR and RI gene sets (P for RR = 2.58×10^{-12} ; P for RI = 0.00092). Estrogen receptor status was significantly correlated with the clustering result of the RR set ($P = 2.05 \times 10^{-8}$) but was not significantly correlated with the clustering result of the RI signature ($P = 0.074$).

To extend and validate these results, we tested the prognostic power of the radiation-repressed and induced gene signatures in a second breast cancer microarray data set derived from tumors from an independent patient population; these data were recently published by Loi *et al.* (21). As with the NKI data set, we cross-mapped the RI and RR genes to the data set of Loi *et al.* (41 RI + 51 RR genes overlapped; see Supplementary Table S1d) and performed clustering analysis in the space of either the RR or RI genes (Fig. 2A and C). As before, the RR signature yielded two distinct clusters (Fig. 2A), with the left cluster showing significantly higher mortality (Fig. 2B, $P = 0.009$) as well as an association with grade 3 tumors (44/47 patients, $P = 1.0 \times 10^{-9}$). As was observed in the NKI data set, the RI signature also separated the samples into two distinct clusters (Fig. 2C) that showed significant differences in clinical outcome (Fig. 2D). The log rank test for the difference in outcome between the clusters yielded a P value of 0.023, which corresponds to a random chance of 0.05 based on random sampling (27 out of 500 iterations have the same or a smaller P value). The clustering by RI genes was again correlated with tumor grade, with the left cluster enriched for poor-outcome tumors (Fig. 2C, $P = 0.0057$). Note that the sizes of the clusters differ between the two clinical data sets (Fig. 1A and C and Fig. 2A and C); this is most likely due to significant differences in the patient populations that make up the two data sets, such as the exclusive focus of the data set of Loi *et al.* on early-stage tumors. We conclude that the RR and RI signatures are prognostic for breast cancer, as demonstrated in two independent patient populations.

Cox Proportional Hazard Model

To fully understand the prognostic significance of the clusters (Fig. 1) identified by deregulation of the radiation response pathway, we next performed univariate and multivariate analyses with the Cox proportional hazards regression model. Parameters for which we had clinical annotation in the two data sets were included in the analyses. For the NKI data set, we found that in univariate analysis (Table 1, left panel) the RI signature, RR signature, wound response, and five clinical variables (age, lymph node status, tumor grade, vascular invasion, and ER status) are significant risk factors for poor clinical outcome ($P < 0.05$). Additionally, the combined RI and RR gene set yielded a significant hazard ratio, with a P value of 3×10^{-9} . Next we performed a multivariate analysis under two models with the RR and RI signatures either separately (Model 1) or combined (Model 2). In the first model, both the RR signature and the age of the patient yielded significant hazard ratios, but the RI signature did not. However, using the combined RR and RI signature yielded a significant P value in the multivariate model ($P = 0.021$) that was marginally lower than the RR signature alone (0.030) (Table 1, Model 2). Identical analyses were performed for the data set of Loi *et al.* (Table 2). Once again, both the RR and RI sets (separately or combined) are strong risk factors in the univariate analysis, with the combined (RR + RI) signature performing marginally better than the individual signatures. In the multivariate analysis with other clinical parameters (Table 2, Model 1), ER status was the predominant risk factor in the model using RR and RI as separate sets ($P = 0.015$). However, as with the NKI data set, when the RR and RI are combined into a single set (Table 2, Model 2), the radiation signature is highly significant ($P = 0.009$). These results strongly suggest that the combined radiation gene signature offers significant potential

in predicting outcome and that the RI gene set contributes additional information over the RR set alone.

Comparison of Outcome Signatures Suggests that the RI Signature Contains Novel Prognostic Information

We next compared the RR and RI signatures to previously published signatures in terms of gene overlap (Table 3 and Supplementary Table S4). Other signatures used in this comparison were the Amsterdam 70 gene (14), the wound response (32), the 76-gene metastasis signature (57), the recurrence score (16), the Genomic Grade Index (GGI) (22), and the PTEN signature (58). We found that while there is significant overlap between the RR gene set and the other outcome signatures (especially with the GGI signature; see below), there is little overlap between the RI gene set and the other outcome signatures.

The GGI signature, with which the RR signature shares the most overlap (34 out of the 97 genes in the GGI signature), is known to represent a transcriptional correlate of proliferation (12,21,30). The lesser overlap with some of the other signatures maybe due to design, because the authors of the Wound Response and PTEN signatures specifically excluded cell cycle-related genes from the final list. Indeed, matching the RR signature to the 193 genes removed from the Wound Response signature yielded an overlap of 28 genes. Hence our finding that the RR gene set encodes prognostic information is not surprising, given the multitude of studies that have converged on the now well-characterized prognostic proliferation module, as discussed below.

In stark contrast, the RI genes show no overlap with the proliferation cluster identified in the GGI signature; rather, the RI genes function in a number of different pathways, including apoptosis signaling and cell cycle arrest. Our finding that the RI gene set shows no overlap with the GGI signature and is *not* comprised of genes involved in proliferation suggests that the RI gene set contains novel prognostic information.

Relationship between the RI Signature and p53-Responsive Genes

p53 is a major factor controlling the transcriptional response to ionizing radiation (59). More than 20 studies have analyzed the prognostic or predictive value of TP53 mutation [reviewed in ref. (40)] or TP53 gene expression-based signatures in breast cancer (60); nearly all have shown that TP53 mutation is associated with poor prognosis (31). Hence we hypothesized that aberrant regulation of the RI genes (predictive of poor outcome) could be a result of altered p53 activity in the poor-outcome tumors. To test this hypothesis, we determined whether genes that were the key drivers of the outcome clustering were known transcriptional targets of p53. Genes that are drivers for the clustering will exhibit large expression differences between the two clusters, so we ranked genes by their differential expression between the good- and poor-outcome clusters in the data set of Loi *et al.* and asked whether the top-ranked genes are known to be regulated by p53. We found that 13 of the top 27 RI genes are *not* known p53-regulated genes (Table 4), suggesting either that these are novel p53 targets or that the RI genes are not entirely p53-regulated. Moreover, when we removed the non-p53 regulated genes from the RI signature and repeated the clustering, there was no significant discrimination between good- and poor-outcome tumors (data not shown), demonstrating that the genes not known to be p53-responsive are important for discriminating outcome. Furthermore, we found no overlap between our RI and RR gene sets and a p53-associated signature derived from breast tumors (60). We conclude from these results that the predictive power of the RI signature is not entirely dependent on known p53-responsive genes.

Cell Type Specificity of the Radiation Response

Our radiation response signature was derived from lymphoblast cell lines. Because the radiation response may vary for different cell types from the same individual, we sought to determine whether a signature derived from irradiated breast epithelial cell lines would prove to be a stronger predictor. To do this, we analyzed a subset of samples from a recently published study looking at gene expression across a panel of tumor cell lines derived from a variety of tissues (48). From this data set, we analyzed microarray data from the five breast cancer cell lines with or without treatment with 8 Gy radiation. Analysis of this data set yielded 66 induced and 10 repressed genes at an $FDR \leq 0.02$ (Supplementary Table S5). None of these genes overlapped with our lymphoblast-derived radiation gene sets. Unlike the radiation signature derived from the lymphoblast study, the gene set derived from the mammary epithelial tumor cell lines was not able to significantly distinguish good- from poor-outcome tumor samples (data not shown). (Of note, this breast tumor line data set represents a small number of tumor lines on a custom array platform of <7000 genes, so its power to derive a predictor may be poor, and its chances of overlapping with our whole genome array data are limited.) Because tumors harbor known defects in the DNA damage response, it is perhaps not surprising that a presumably abnormal radiation signature derived from tumor cell lines does not predict outcome in breast cancers. It is conceivable that a radiation signature derived from primary mammary epithelial cells could yield a stronger predictor than that derived from lymphoblasts, but no data set is currently available to test this hypothesis.

Agreement among Outcome Signatures

Next we compared the prognostic capability of the radiation response signature to previously published breast outcome signatures using the NKI data set (14). This comparison was done using the leave-one-out cross-validation procedure, which consisted of calculating the mean expression of each signature with respect to the good- and poor-outcome groups and comparing the distances of the left-out sample with the centroid of each group (see the Materials and Methods). The results for each signature are reported in Table 5. Despite differences in the sizes of the gene sets, the signatures showed similar error rates and sensitivity for classifying patients for outcome.

To examine whether a subset of tumors was commonly misclassified by all of the gene signatures, we plotted a heat map of classification status for each sample as determined by each of the eight prediction signatures (Fig. 3). Each signature was used to predict the classification of patients into good-outcome (survival >10 years) and poor-outcome (survival <10 years) groups. There was good concordance among the signatures (Fig. 3). Interestingly, there exists a subset of tumors that is misclassified by all or most of the gene signatures. For example, in the poor-outcome group, there is a subgroup of 10–20 patients that are commonly misclassified as “good” by most signatures. Conversely, in the good-outcome group, a similarly sized subset of samples has been misclassified as “poor” by most signatures. We next examined whether misclassification of tumors was associated with the other known clinical parameters for which annotation was available (age, diameter, lymph node metastasis, tumor grade, vascular invasion, ER status, mastectomy and chemotherapy; see Supplementary Tables S6 and S7). For poor-outcome patients that were misclassified as “good outcome” by the majority of expression signatures, high tumor grade ($P = 0.02$) was enriched. In contrast, no clinical variables were significantly enriched in tumors associated with good outcome but misclassified as poor outcome. We conclude that the gene expression signatures perform similarly in tumor classification and that there are subsets of both poor- and good-outcome tumors that are commonly misclassified by all of the signatures.

DISCUSSION

Questions that arise from this study are why a transcriptional signature of the cellular radiation response shows prognostic capabilities in breast cancer, and how the radiation-derived predictor is different from other known predictors in breast cancer. There are interesting parallels between tumorigenesis and the radiation response (Fig. 4). To promote tumorigenesis, two key cellular processes become altered: Proliferation is deregulated (by the activation of oncogenes), and apoptosis and cell cycle checkpoint functions are disabled (by the inactivation of tumor suppressor genes). These same key cellular processes are represented in the transcriptional response of cells to ionizing radiation, since the RR gene set is enriched in mitosis and proliferation functions, and the RI gene set is enriched in cell cycle arrest and apoptosis, among others (Supplementary Table S2).

The majority of outcome signatures for breast cancer essentially detect the same property: the high potential for proliferation in poor-outcome tumors (12). Hence it is not surprising that the set of genes repressed by radiation in our study is able to predict breast cancer outcome, since this gene set is enriched in genes involved in mitosis/proliferation. We suggest that the reason most breast cancer outcome signatures to date focus on proliferation is that they were trained on gene expression array data sets from human tumors, in which the strongest transcriptional program driving this clustering is one associated with a singular biological property: proliferation. However, these results do not rule out the possibility that additional biological properties of breast cancers, whose transcriptional correlate is not as strong as that of the proliferation signature (and hence does not drive the hierarchical clustering of tumors), may also carry significant prognostic information.

Indeed, our finding that the set of genes induced by radiation (RI) is enriched in checkpoint/apoptosis functions and is distinct from the proliferation cluster indicates that the RI signature may contain novel predictive information, less related to high proliferation and likely representing additional tumor properties such as genomic instability. In addition to genes involved in apoptosis and cell cycle arrest, a number of functional pathways not previously implicated in the radiation response were among the top radiation-induced pathways. Foremost among these were the categories corresponding to glycosaminoglycan, aminoglycan and proteoglycan metabolism (Supplementary Table S2). Perhaps relatedly, proteoglycans such as heparan sulfate have been shown to be linked to tumor progression (61).

Also consistent with our observation that the RI gene set carries prognostic information, others have shown that human precancerous (and early-stage cancerous) lesions show constitutive activation of the DNA damage response (DDR) pathway, and this activation disappears as tumors progress (Fig. 4A) (34,35), suggesting that the DNA damage checkpoint acts as an antiproliferative barrier for tumorigenesis (38,39). Hence one can imagine that a lack of this checkpoint would be associated with tumor progression and poor clinical outcome.

Because p53 is one of the main transcriptional activators in the normal radiation response, and poor-outcome tumors are commonly p53-negative, it can be hypothesized that the RI signature's predictive power is at least partially derived from p53 status. As predicted, p53 regulated genes make up a significant component of the RI signature. However, as shown in Table 4, many of the genes that drive the clustering into good- and poor-outcome groups are not known to be regulated by p53 but are required for accurate classification of the tumors. This suggests either that many of these genes are actually p53-dependent targets that have not yet been discovered as such or that the RI signature's predictive capabilities are not entirely p53-dependent, representing a novel connection between the radiation response and tumor outcome.

Despite an overwhelming body of evidence indicating the potential clinical prognostic utility of the DDR pathway in human breast cancer, the lack of a clinically tractable robust assay to assess the activity of this pathway has prevented its translation into the clinical care of breast cancer patients (62). Because the DDR pathway contains p53-dependent and p53-independent responses, p53 status alone does not completely assess the activity or functionality of the signal transduction cascade (63). Hence it is likely that the (thus far) nearly exclusive focus of clinical testing on p53 has failed to fully use this complex cellular pathway as a clinical marker. The RI signature, which encompasses multiple facets of the DDR pathway, serves as an initial foray into a more robust DDR signature for prognostication, and further characterization of a number of the key drivers for this clustering may facilitate development of a multianalyte panel allowing more accurate prognostication for breast cancer patients.

Of note, our radiation-responsive transcriptional study was performed in lymphoblast cell lines, yet the signature serves as an excellent predictor of survival outcome in mammary *epithelial* tumors. This is not the first predictive signature to be derived from an entirely different cell type. The wound response signature (32) is another efficient predictor of survival outcome in breast cancer, but it was in fact derived from serum-stimulated fibroblast cells. The transcriptional radiation response of lymphoid cells has been shown previously to be applicable to other tissues as well, in particular for predicting radiotoxicity of normal tissues in prostate cancer patients treated with radiation therapy (64). It is unknown how the radiation response differs among different normal cell types.

Like the wound response signature, the radiation signature offers additional support for the development of gene expression signatures based on known pathways whose mutations contribute to tumorigenesis. Because tumor progression may proceed through a series of mutations affecting multiple pathways, one can imagine the utility of a panel of gene expression signatures for tumor diagnosis, each based on the functionality of a specific anti-tumor barrier. Such a tactic has root in the idea of the “Vogelgram”, in which tumorigenesis can be understood as a multistep genetic progression (65). An ordered panel of pathway-specific changes in gene expression may offer a complete picture of the genomic landscape for specific tumors and could determine where each tumor sits in such a progression model. Such data would be important for both prognosis and deciding on a plan of treatment (66).

In a 2006 analysis, Fan *et al.* showed that despite minimal overlap among cancer outcome signatures, each yielded similar prognostic capabilities (30). The fact that all expression signatures identify a number of poor-outcome tumors that are not accurately classified by known clinical parameters is particularly encouraging. However, it is equally interesting that there appears to be a subset of tumors that are commonly misclassified by all of the expression signatures (Fig. 4). This observation suggests that targeted studies will be required to further refine molecular signatures to encapsulate these tumors that aren’t easily classifiable by any available means.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors would like to thank Drs. Eric Lander and Todd Golub for mentorship and support during the generation of the radiation microarray data set. Additionally, we thank Ken Ross and Richard Ivey for discussions, Jeffrey Patterson for use of analysis tools, Chenwei Lin for bioinformatics assistance, Christine Ladd and Lei Yan for technical assistance, and Andrea Hightower for assistance in preparing tables. AGP is supported by NIH U19 AI 067770-02 and NIH R01 CA 129604. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. BDP is supported through the Amgen Interdisciplinary Fellowship Fund.

References

1. Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, Mandelblatt JS, Yakovlev AY, Habbema JD, Feuer EJ. Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med* 2005;353:1784–1792. [PubMed: 16251534]
2. Fracheboud J, Otto SJ, van Dijk JA, Broeders MJ, Verbeek AL, de Koning HJ. Decreased rates of advanced breast cancer due to mammography screening in The Netherlands. *Br J Cancer* 2004;91:861–867. [PubMed: 15292936]
3. Cole BF, Gelber RD, Gelber S, Coates AS, Goldhirsch A. Polychemotherapy for early breast cancer: an overview of the randomised clinical trials with quality-adjusted survival analysis. *Lancet* 2001;358:277–286. [PubMed: 11498214]
4. Eifel P, Axelson JA, Costa J, Crowley J, Curran WJ Jr, Deshler A, Fulton S, Hendricks CB, Kemeny M, Roter D. National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, November 1–3, 2000. *J Natl Cancer Inst* 2001;93:979–989. [PubMed: 11438563]
5. Shapiro CL, Recht A. Side effects of adjuvant treatment of breast cancer. *N Engl J Med* 2001;344:1997–2008. [PubMed: 11430330]
6. Monnier A. Clinical management of adverse events in adjuvant therapy for hormone-responsive early breast cancer. *Ann Oncol* 2007;18(Suppl 8):viii36–viii44. [PubMed: 17890213]
7. Perez EA. Safety profiles of tamoxifen and the aromatase inhibitors in adjuvant therapy of hormone-responsive early breast cancer. *Ann Oncol* 2007;18(Suppl 8):viii26–viii35. [PubMed: 17890211]
8. Fisher B, Dignam J, Tan-Chiu E, Anderson S, Fisher ER, Wittliff JL, Wolmark N. Prognosis and treatment of patients with breast tumors of one centimeter or less and negative axillary lymph nodes. *J Natl Cancer Inst* 2001;93:112–120. [PubMed: 11208880]
9. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. *Nat Genet* 2003;33:49–54. [PubMed: 12469122]
10. van't Veer LJ, Paik S, Hayes DF. Gene expression profiling of breast cancer: a new tumor marker. *J Clin Oncol* 2005;23:1631–1635. [PubMed: 15755970]
11. Marchionni L, Wilson RF, Wolff AC, Marinopoulos S, Parmigiani G, Bass EB, Goodman SN. Systematic review: gene expression profiling assays in early-stage breast cancer. *Ann Intern Med* 2008;148:358–369. [PubMed: 18252678]
12. Sotiriou C, Piccart MJ. Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? *Nat Rev Cancer* 2007;7:545–553. [PubMed: 17585334]
13. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530–536. [PubMed: 11823860]
14. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Bernards R. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;347:1999–2009. [PubMed: 12490681]
15. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, d'Assignies MS, Bergh J, Lidereau R, Piccart MJ. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* 2006;98:1183–1192. [PubMed: 16954471]
16. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Wolmark N. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351:2817–2826. [PubMed: 15591335]
17. Habel LA, Shak S, Jacobs MK, Capra A, Alexander C, Pho M, Baker J, Walker M, Watson D, Quesenberry CP. A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Res* 2006;8:R25. [PubMed: 16737553]
18. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, Cronin M, Baehner FL, Watson D, Wolmark N. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 2006;24:3726–3734. [PubMed: 16720680]
19. Cardoso F, van't Veer LJ, Rutgers E, Loi S, Mook S, Piccart-Gebhart MJ. Clinical application of the 70-gene profile: the MIN-DACT trial. *J Clin Oncol* 2008;26:729–735. [PubMed: 18258980]

20. Sparano JA, Paik S. Development of the 21-gene assay and its application in clinical practice and clinical trials. *J Clin Oncol* 2008;26:721–728. [PubMed: 18258979]
21. Loi S, Haibe-Kains B, Desmedt C, Lallemand F, Tutt AM, Gillet C, Ellis P, Harris A, Bergh J, Sotiriou C. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* 2007;25:1239–1246. [PubMed: 17401012]
22. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, Nordgren H, Farmer P, Praz V, Delorenzi M. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 2006;98:262–272. [PubMed: 16478745]
23. Acharya CR, Hsu DS, Anders CK, Anguiano A, Salter KH, Walters KS, Redman RC, Tuchman SA, Moylan CA, Potti A. Gene expression signatures, clinicopathological features, and individualized therapy in breast cancer. *J Am Med Assoc* 2008;299:1574–1587.
24. Chang JC, Wooten EC, Tsimelzon A, Hilsenbeck SG, Gutierrez MC, Tham YL, Kalidas M, Elledge R, Mohsin S, O'Connell P. Patterns of resistance and incomplete response to docetaxel by gene expression profiling in breast cancer patients. *J Clin Oncol* 2005;23:1169–1177. [PubMed: 15718313]
25. Klijn JG, de Vries EG. The value of gene expression profiles using microarrays for the individualisation of adjuvant therapy after surgery for breast cancer. *Ned Tijdschr Geneesk* 2005;149:618–622. [PubMed: 15813427]
26. Ma X, Yang Q, Wilson KT, Kundu N, Meltzer SJ, Fulton AM. Promoter methylation regulates cyclooxygenase expression in breast cancer. *Breast Cancer Res* 2004;6:R316–R321. [PubMed: 15217498]
27. Pawitan Y, Bjohle J, Amler L, Borg AL, Eghazi S, Hall P, Han X, Holmberg L, Huang F, Bergh J. Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res* 2005;7:R953–R634. [PubMed: 16280042]
28. Ring BZ, Seitz RS, Beck R, Shasteen WJ, Tarr SM, Cheang MC, Yoder BJ, Budd GT, Nielsen TO, Ross DT. Novel prognostic immunohistochemical biomarker panel for estrogen receptor-positive breast cancer. *J Clin Oncol* 2006;24:3039–3047. [PubMed: 16809728]
29. Wang Y, Rannala B. In silico analysis of disease-association mapping strategies using the coalescent process and incorporating ascertainment and selection. *Am J Hum Genet* 2005;76:1066. [PubMed: 15818531]
30. Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB, van't Veer LJ, Perou CM. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006;355:560–569. [PubMed: 16899776]
31. Olivier M, Langerod A, Carrieri P, Bergh J, Klaar S, Eyfjord J, Theillet C, Rodriguez C, Lidereau R, Børresen-Dale AK. The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer. *Clin Cancer Res* 2006;12:1157–1167. [PubMed: 16489069]
32. Chang HY, Sneddon JB, Alizadeh AA, Sood R, West RB, Montgomery K, Chi JT, van de Rijn M, Botstein D, Brown PO. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol* 2004;2:E7. [PubMed: 14737219]
33. Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshirani R, Sorlie T, Dai H, He YD, van't Veer LJ, de Vijver MJ. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc Natl Acad Sci USA* 2005;102:3738–3743. [PubMed: 15701700]
34. Bartkova J, Horejsi Z, Koed K, Kramer A, Tort F, Zieger K, Guldberg P, Sehested M, Nesland JM, Bartek J. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* 2005;434:864–870. [PubMed: 15829956]
35. Gorgoulis VG, Vassiliou LV, Karakaidos P, Zacharatos P, Kotsinas A, Liloglou T, Venere M, Ditullio RA Jr, Kastrinakis NG, Halazonetis T. Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* 2005;434:907–913. [PubMed: 15829965]
36. Osborn AJ, Elledge SJ, Zou L. Checking on the fork: the DNA-replication stress-response pathway. *Trends Cell Biol* 2002;12:509–516. [PubMed: 12446112]
37. Halazonetis TD, Gorgoulis VG, Bartek J. An oncogene-induced DNA damage model for cancer development. *Science* 2008;319:1352–1355. [PubMed: 18323444]

38. Di Micco R, Fumagalli M, Cicalese A, Piccinin S, Gasparini P, Luise C, Schurra C, Garre M, Nuciforo PG, d'Adda di Fagagna F. Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature* 2006;444:638–642. [PubMed: 17136094]
39. Bartkova J, Rezaei N, Lontos M, Karakaidos P, Kletsas D, Issaeva N, Vassiliou LV, Kolettas E, Niforou K, Gorgoulis VG. Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature* 2006;444:633–637. [PubMed: 17136093]
40. Petitjean A, Achatz MI, Borresen-Dale AL, Hainaut P, Olivier M. TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. *Oncogene* 2007;26:2157–2165. [PubMed: 17401424]
41. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005;102:15545–15550. [PubMed: 16199517]
42. Chomczynski P, Sacchi N. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nat Protoc* 2006;1:581–585. [PubMed: 17406285]
43. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 1987;162:156–159. [PubMed: 2440339]
44. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;4:249–264. [PubMed: 12925520]
45. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 2001;98:5116–5121. [PubMed: 11309499]
46. Storey J. The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann Stat* 2003;31:2013–2035.
47. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25:25–29. [PubMed: 10802651]
48. Amundson SA, Do KT, Vinikoor LC, Lee RA, Koch-Paiz CA, Ahn J, Reimers M, Chen Y, Scudiero DA, Fornace AJ Jr. Integrating global gene expression and radiation survival parameters across the 60 cell lines of the National Cancer Institute Anti-cancer Drug Screen. *Cancer Res* 2008;68:415–424. [PubMed: 18199535]
49. Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, Lin L, Chen G, Gharib TG, Hanash S. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 2002;8:816–824. [PubMed: 12118244]
50. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Meyerson M. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci USA* 2001;98:13790–13795. [PubMed: 11707567]
51. Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, McLaughlin ME, Kim JY, Goumnerova LC, Black PM, Golub TR. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* 2002;415:436–442. [PubMed: 11807556]
52. Nutt CL, Mani DR, Betensky RA, Tamayo P, Cairncross JG, Ladd C, Pohl U, Hartmann C, McLaughlin ME, Louis DN. Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Res* 2003;63:1602–1607. [PubMed: 12670911]
53. Monti S, Savage KJ, Kutok JL, Feuerhake F, Kurtin P, Mihm M, Wu B, Pasqualucci L, Neuberg D, Shipp MA. Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood* 2005;105:1851–1861. [PubMed: 15550490]
54. Singh D, Febbo PG, Ross K, Jackson DG, Manola J, Ladd C, Tamayo P, Renshaw AA, D'Amico AV, Sellers WR. Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* 2002;1:203–209. [PubMed: 12086878]

55. West M, Blanchette C, Dressman H, Huang E, Ishida S, Spang R, Zuzan H, Olson JA Jr, Marks JR, Nevins JR. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci USA* 2001;98:11462–11467. [PubMed: 11562467]
56. Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, Ladd C, Reich M, Latulippe E, Golub TR. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci USA* 2001;98:15149–15154. [PubMed: 11742071]
57. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Foekens JA. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 2005;365:671–679. [PubMed: 15721472]
58. Saal LH, Johansson P, Holm K, Gruvberger-Saal SK, She QB, Maurer M, Koujak S, Ferrando AA, Malmstrom P, Parsons R. Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. *Proc Natl Acad Sci USA* 2007;104:7564–7569. [PubMed: 17452630]
59. Jen KY, Cheung VG. Identification of novel p53 target genes in ionizing radiation response. *Cancer Res* 2005;65:7666–7673. [PubMed: 16140933]
60. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, Pawitan Y, Hall P, Klaar S, Bergh J. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci USA* 2005;102:13550–13555. [PubMed: 16141321]
61. Sanderson RD, Yang Y, Suva LJ, Kelly T. Heparan sulfate proteoglycans and heparanase—partners in osteolytic tumor growth and metastasis. *Matrix Biol* 2004;23:341–352. [PubMed: 15533755]
62. Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, Somerfield MR, Hayes DF, Bast RC Jr. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol* 2007;25:5287–5312. [PubMed: 17954709]
63. Harper JW, Elledge SJ. The DNA damage response: ten years after. *Mol Cell* 2007;28:739–745. [PubMed: 18082599]
64. Svensson JP, Stalpers LJ, Esveldt-van Lange RE, Franken NA, Haveman J, Klein B, Turesson I, Vrieling H, Giphart-Gassler M. Analysis of gene expression using gene sets discriminates cancer patients with and without late radiation toxicity. *PLoS Med* 2006;3:e422. [PubMed: 17076557]
65. Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. *Cell* 1996;87:159–170. [PubMed: 8861899]
66. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, Joshi MB, Harpole D, Lancaster JM, Nevins JR. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 2006;439:353–357. [PubMed: 16273092]

FIG. 1.

Radiation-repressed (RR) or radiation-induced (RI) gene signatures cluster breast cancer patients from the Netherland Cancer Institute (14) according to clinical outcome. Panel A shows clustering of tumors in the space of the RR genes. The bars above the heat map indicate clinical variables of tumor grade (black, poorly differentiated; gray, intermediate; white, well differentiated), and ER status (black, positive; white, negative). Panel B shows Kaplan-Meier survival analysis for the two clusters separated in panel A. Panels C and D show clustering analysis and survival analysis in the space of the RI genes. See Supplementary Table S1c for the genes interrogated in this analysis.

FIG. 2. Radiation-repressed (RR) or radiation-induced (RI) gene signatures cluster breast cancer patients from the study of Loi *et al.* (21) according to clinical outcome. Panel A shows clustering of tumors in the space of the RR genes. Panel B shows Kaplan-Meier survival analysis for the two clusters separated in panel A. Panels C and D show clustering analysis and survival analysis in the space of the RI genes. The bars above the heat map indicate clinically assessed tumor grade (black, poorly differentiated; gray, intermediate; white, well differentiated). See Supplementary Table S1d for the genes interrogated in this analysis.

FIG. 3.

Heat map showing classification of each patient in the NKI data set (14) based on cross-validation for all the samples using the panel of gene signatures. Each column represents one patient, with yellow indicating classification in the good clinical outcome group and blue representing classification in the poor clinical outcome group. The survival time is indicated by the top bars, with red indicating a death event. Tumor grade (grade 1, white; grade 2, gray; grade 3, black) and ER status (ER^- , white; ER^+ , black) are also represented by the middle bars.

A. Current model for tumorigenesis.

B. Parallels between DNA damage response and tumorigenesis.

FIG. 4.

Parallels between the molecular stages of tumorigenesis and the DNA damage response. Panel A: Evidence suggests that common forms of cancer may develop through a multistep process that may begin with a (dominant) mutation that results in oncogene activation and consequent aberrant proliferation. Aberrant proliferation results in replication stress, leading to DNA double-strand breaks (similar to ionizing radiation) and activation of the DNA damage response (34,37,38). Ultimately, mutations inactivate the DNA damage response, unleashing clonal evolution and tumor progression, leading to poor survival. It is well established that a proliferation cluster of genes predicts outcome in breast cancer. Panel B: Exposure to ionizing radiation activates the DNA damage response pathway, resulting in both the induction of one set and repression of a second set of genes. Pathway analysis reveals that the repressed genes function in mitosis (and overlap with the proliferation cluster predicting breast cancer outcomes), whereas the induced genes function in a variety of biochemical pathways including checkpoint/cell cycle arrest and apoptosis and do not show significant overlap with the proliferation cluster.

Univariate and Multivariate Analysis of Risk Factors for Death in a Group of Breast Cancer Patients from the Netherlands Cancer Institute (14)

	Multivariate analysis					
	Univariate analysis			Model 1		
	Hazard ratio	P value		Hazard ratio	P value	Hazard ratio
RI signature	1.1	0.04		1.071	0.170	—
RR signature	0.896	5.61×10^{-9}		0.940	0.030	—
Rad signature (RI + RR)	0.898	2.9×10^{-9}		—	—	0.021
Age	0.593	0.013		0.652	0.050	0.033
Diameter	1.19	0.13		1.033	0.810	0.910
Lymph node (0, 1, 2)	1.12	0.012		1.004	0.970	0.960
Tumor grade (1, 2, 3)	1.98	1.9×10^{-5}		1.160	0.460	0.590
VascularInvasion.13	0.87	0.55		1.741	0.160	0.140
VascularInvasion.4	1.90	0.02		2.825	0.140	0.130
ER status (1 = pos, 0 = neg)	0.334	2.7×10^{-6}		0.583	0.069	0.150
Mastectomy or conserving therapy	1.12	0.62		0.983	0.950	0.820
Chemo- or hormonal therapy	0.931	0.75		0.574	0.120	0.130
Wound response	3.67	9.3×10^{-8}		1.382	0.12	0.230

Notes. The left two columns list hazard ratios and P values from a univariate Cox analysis of survival outcome calculated independently for each parameter. The subsequent columns show two variations on a multivariate Cox analysis, using RR and RI sets as independent variables with other clinical properties (Model 1), and using the combined radiation set with the same clinical variables (Model 2). Significant findings are highlighted in boldface.

Univariate and Multivariate Analysis of Risk Factors for Death in a Group of Breast Cancer Patients from Loi *et al.* (21)

TABLE 2

	Multivariate analysis					
	Univariate analysis		Model 1		Model 2	
	Hazard ratio	P value	Hazard ratio	P value	Hazard ratio	P value
RI signature	1.09	0.005	1.065	0.130	—	—
RR signature	0.946	3×10^{-5}	0.970	0.160	—	—
Rad signature (RI + RR)	0.948	2.6×10^{-5}	—	—	0.954	0.009
Age	1	1	0.99	0.470	0.992	0.540
Tumor grade (1, 2, 3)	1.56	0.015	1.043	0.850	1.009	0.970
Tumor size (0–8.2)	1.34	0.00012	1.218	0.071	1.182	0.110
ER status (1 = pos, 0 = neg)	0.217	0.0097	0.22	0.015	0.232	0.019
Lymph node (0, 1)	1.21	0.39	1.110	0.690	1.139	0.620

Notes: The left two columns list hazard ratios and P values from a univariate Cox analysis of survival outcome calculated independently for each parameter. The subsequent columns show two variations on a multivariate Cox analysis, using RR and RI sets as independent variables with other clinical properties (Model 1), and using the combined radiation set with the same clinical variables (Model 2). Significant findings are highlighted in boldface.

TABLE 3

Overlap of Radiation-Induced (RI) and Radiation-Repressed (RR) Gene Signatures with Other Breast Cancer Outcome Signatures

	Total genes	Overlap with RR	Overlap with RI
Amsterdam 70-gene (14)	70	1	1
76-gene (57)	75	0	0
Wound response (32)	512	1	2
GGI (22)	97	34	0
PTEN (58)	246	11	4
RS (16)	21	3	0

TABLE 4

p53 Dependence of Radiation-Induced Genes

RI gene	P value ^a	p53 regulated?	Reference
BLOC1S2	0	not reported	
BRMS1L	0	not reported	
BTBD10	0	not reported	
C12orf5	0	yes	<i>Cancer Res.</i> 65 , 7666–7673 (2005)
C8orf38	0	not reported	
DRAM	0	yes	<i>Cell</i> 26 , 121–134 (2006)
FBXO22	0	not reported	
FTSJ2	0	not reported	
LRDD	0	yes	<i>Genome Biol.</i> 7 , R25 (2006)
NOTCH1	0	yes	<i>Mol. Cell. Biol.</i> 27 , 3732–3742 (2007)
PPM1D	0	yes	<i>Genomics</i> 64 , 298–306 (2000)
TRIAP1	0	yes	<i>Cancer Res.</i> 65 , 1197–1206 (2005)
APTX	1.00×10^{-4}	not reported	
SESN2	2.00×10^{-4}	yes	<i>Oncogene</i> 21 , 6017–6031 (2002)
XPC	5.00×10^{-4}	yes	<i>Genome Biol.</i> 7 , R25 (2006)
RBM14	0.001	not reported	
ZNF337	0.001	not reported	
CSNK1G1	0.0012	yes	<i>J. Cell Biol.</i> 173 , 533–544 (2006)
MAD1L1	0.0021	yes	<i>J. Biol. Chem.</i> 278 , 37439–37450 (2003)
TRIM22	0.0029	yes	<i>Genome Biol.</i> 7 , R25 (2006)
CDKN1A	0.0043	yes	<i>Cancer Res.</i> 65 , 7666–7673 (2005)
TMEM142C	0.0084	not reported	
PHLDA3	0.0128	yes	<i>Cancer Res.</i> 65 , 7666–7673 (2005)
C1orf183	0.0276	not reported	
VPS33B	0.0281	not reported	
CPM	0.0361	not reported	
PRKAB1	0.0365	yes	<i>Cancer Res.</i> 65 , 7666–7673 (2005)
GLS2	0.0599	yes	<i>Genome Biol.</i> 7 , R25 (2006)
ZMAT3	0.0602	yes	<i>Genome Biol.</i> 7 , R25 (2006)
FDXR	0.0793	yes	<i>Oncogene</i> 21 , 6017–6031 (2002)
ZNF79	0.0912	not reported	
TM7SF3	0.1251	not reported	
POLH	0.1358	yes	<i>Oncogene</i> 21 , 6017–6031 (2002)
DDB2	0.1791	yes	<i>Genome Biol.</i> 7 , R25 (2006)
MRPL49	0.286	yes	<i>J. Invest. Dermatol.</i> 126 , 2490–506 (2006)
VWCE	0.4255	not reported	
ANKRA2	0.5336	not reported	
GADD45A	0.6555	yes	<i>Genome Biol.</i> 7 , R25 (2006)
TMEM168	0.7914	not reported	
SESN1	0.8738	yes	<i>Genome Biol.</i> 7 , R25 (2006)
ASTN2	0.8815	yes	<i>Cell</i> 124 , 207–219 (2006)

RI gene	<i>P</i> value ^a	p53 regulated?	Reference
---------	-----------------------------	----------------	-----------

^aFor the 41 RI genes overlapping with the breast tumor data set of Loi *et al.*, we list the *P* value for differential expression between the good- and poor-outcome clusters, along with information on whether gene expression is known to be driven by p53. Genes with significant *P* values (indicated by boldface) showing differential expression between the two clusters are hypothesized to play a role in driving the clustering.

TABLE 5

Error Rate and Sensitivity from Leave-One-Out Cross-Validation Analysis of a Panel of Breast Cancer Outcome Signatures for Tumor Classification Using the NKI Data Set

Gene set	No. of genes in signature	Overlap between signature and genes on the NKI arrays	Error rate	Sensitivity
Radiation response	68(RR) + 50(RI)	34(RR) + 15(RI)	30.2%	68.4%
Wound response	512	269	28.7%	75.4%
70-gene	70	47	34.9%	63.2%
RS	20	13	28.7%	73.7%
GGI	94	60	28.7%	73.7%
76-gene	76	38	29.5%	68.4%
PTEN	246	104	28.7%	68.4%

Notes. Error rate = percentage of patients being classified in the wrong group. Sensitivity = percentage of patients in the bad survival group being classified correctly.